

CLAIMS

What is claimed is:

1. An engineered polypeptide having proline hydroxylase activity comprising 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 sequence identity to a reference sequence of SEQ ID NO: 4, 116, 162, 322, 412, 492, 562, 598, and/or 630 and one or more residue differences as compared to SEQ ID NO: 4, 116, 162, 322, 412, 492, 562, 598, and/or 630.

2. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 4 and one or more residue differences as compared to SEQ ID NO: 4 at residue positions selected from: 194, 123, 21, 28, 58/247, 65, 80, 85, 95, 98, 117, 120, 159, 185, 199, 200, 233, 237, 243, 250, 268, 281, 282, 287, 289, 307, 324, 326, 327, 330, 338, 343, 346, and 348.

3. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 4 and one or more residue differences as compared to SEQ ID NO: 4 at residue positions selected from: 21, 28, 45, 65, 95, 112, 117, 139, 177, 185, 199, 233, 243, 250, 281, 282, 287, 289, 307, 324, 326, 327, 335, 338, 343, and 346.

4. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 4 and one or more residue differences as compared to SEQ ID NO: 4 at residue positions selected from: 48/66/189/194, 48/66/194, and 66/82/85/135/189/194/267.

5. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 4 and one or more residue differences as compared to SEQ ID NO: 4 at residue positions selected from: 20/56/76/168/169/296, 20/56/232/294, 20/119/294/296, 56/76/119/124/147/232, 56/76/294, 76/168/232/294, 76/294/296, 76/296, 147, and 232.

6. The engineered polypeptide of Claim 1, comprising 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

sequence identity to reference sequence SEQ ID NO: 116 and one or more residue differences as compared to SEQ ID NO: 116 at residue positions selected from: 123, 189, 195, 233, and 296.

7. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 116 and one or more residue differences as compared to SEQ ID NO: 116 at residue positions selected from: 20/21/56, 20/21/56/76/95/232/294/307/335, 20/21/56/76/147/225/232/233/281/294/296/307/335, 20/21/56/95/147/281/294/307, 20/21/56/281/307, 20/21/76/232/243, 20/21/95/232/307, 20/21/95/281/294/296, 20/21/147/189/233/243/281/307, 20/56, 20/56/76/95/281/307, 20/56/76/147/294/296/307, 20/56/95/147/294, 20/56/281, 20/76, 20/76/95/281/294/296, 20/76/95/281/296/307, 20/76/233/294/307, 20/76/243/281/294, 21/76/147/233/294/307, 21/76/147/243/296/307/335, 21/95/185/189/232/281/296, 21/95/233/243/281/296, 21/95/294/296/307/335, 21/95/307, 21/281/307, 29/76/281, 56/76/95/232/243/281, 56/76/147/281/307, 56/76/243/294, 56/76/281/294, 56/76/296, 56/76/307, 56/95/147/307/335/348, 56/95/232/233/281/294/307, 56/95/243/281, 56/147/281, 56/232/243/281, 56/232/281, 56/232/281/294/296, 56/233/281/294/296, 56/281/307, 76/95/232/243/281/307, 76/95/243/281/307/335, 76/95/294/307, 76/147, 76/147/233/243/294, 76/147/233/281/294/307, 76/147/243/294/296/307/335, 76/147/281/307, 76/189/296, 76/232/233/243/294/296/307, 76/281, 76/281/294, 76/294/296, 95/120, 95/147/335, 95/232/243/281/294/307, 95/232/281/294/296, 95/281/294/296, 95/335, 147, 147/225/232/243/281/296/307/335, 147/233/243/281/307, 147/233/281/307/335, 147/243/281, 147/307, 232/233/281/294/296/307, 232/281, 232/284/307, 233/243/281/296/307/335, 233/281/296/307, 243/281/294/296, 281, 281/294, 281/307, 307, and 335.

8. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 116 and one or more residue differences as compared to SEQ ID NO: 116 at residue positions selected from: 21/76/147/243/296/307/335, 56/76/147/281/307, and 95/147/335.

9. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 162 and one or more residue differences as compared to SEQ ID NO: 162 at residue positions selected from: 2/85/123/237, 28/115/117/120/123/268/270/343/346/348, 45/123/326, 65/117/120/123/343/346, 85/123/281/282, 114/115/117/120/123/268/271/313/326/343/346, 123/139/233/237/281/282/289/324/326, and 123/199/200/247/250/338.

10. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 322 and one or more residue differences as compared to SEQ ID NO: 322 at residue positions selected from: 26, 54, 61, 129, 132, 149, 156, 175, 189, 201, 209, 228, 236, 248, 262, 272, 277, 291, and 345.

11. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 322 and one or more residue differences as compared to SEQ ID NO: 322 at residue positions selected from: 25, 43, 54, 58, 61, 79, 129, 132, 143, 156, 163, 175, 179, 201, 209, 236, 248, 278, 291, 345, and 347.

12. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 322 and one or more residue differences as compared to SEQ ID NO: 322 at residue positions selected from: 85/117/120/135/208/270/324/343/346, 85/117/120/135/208/281/282/289, 85/117/120/270/281/289, 85/117/135/139/208, and 117/120/208/270/324/343/346.

13. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 412 and one or more residue differences as compared to SEQ ID NO: 412 at residue positions selected from: 47, 48, 56/118, 85, 95, 95/289, 113, 118, 118/247, 154, 162, 162/204, 164, 164/198/271, 168, 169, 187, 195, 243, 271, 275, 281, 314, 330, and 342.

14. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 412 and one or more residue differences as compared to SEQ ID NO: 412 at residue positions selected from: 25/129/163/236/262/345/347, 120/156/175/179/201, 129/189/236/262/277/278, 129/236/262, 156/175/179/228, and 162.

15. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 492 and one or more residue differences as compared to SEQ ID NO: 492 at residue positions selected from: 15, 17, 28, 29, 65, 135, 167, 177, 199, 208, 228, 235, 287, 294, 307, and 343.

16. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 492 and one or more residue differences as compared to SEQ ID NO: 492 at residue positions selected from: 85/187/281/347, 85/187/347, 118/120/162/175/179/330, 118/120/162/175/330, 162/175/179/330, 175/228/330, 195/347, and 278/314/347.

17. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 562 and one or more residue differences as compared to SEQ ID NO: 562 at residue positions selected from: 15, 40, 43, 44, 59, 79, 82, 149, 164, 179, 345, and 347.

18. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 562 and one or more residue differences as compared to SEQ ID NO: 562 at residue positions selected from: 29/85/177/208/228/347, 29/85/208/228/343/347, 29/177/195/228/343, 29/208/228/278/294/347, 56/195/278, 85/187/205/208/278, 113/177/187/195/208/278/294/343/347, and 177/205/208/228.

19. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 598 and one or more residue differences as compared to SEQ ID NO: 598 at residue positions selected from: 47, 162, 209, 219, 227, and 342.

20. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 598 and one or more residue differences as compared to SEQ ID NO: 598 at residue positions selected from: 17/44/179/195/250/313/345, 17/44/199/313, 43/44/195/199, 44/149/164/171/187, 44/179/195/199, 44/179/195/199/345, 79/163/164/171/187/201/286/288, 82/163/164, 82/163/164/171/187/201/203/208/286/288/320, 149/164/171/288, and 187/286.

21. The engineered polypeptide of Claim 1, comprising 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 sequence identity to reference sequence SEQ ID NO: 630 and one or more residue differences as compared to SEQ ID NO: 630 at residue positions selected from: 82/164/171/203/208,

135/163/164/201/203/208, 162, 162/219/236, 162/219/313/338, 162/236/342, 162/313/342, and 164/171/201/203/282.

22. The engineered polypeptide of Claim 1, wherein said polypeptide has at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to at least one of the even-numbered sequences in SEQ ID NOs: 6-658.

23. The engineered polypeptide of Claim 1, wherein said engineered polypeptide is capable of converting L-proline to *trans*-3-hydroxyproline.

24. The engineered polypeptide of Claim 23, wherein said engineered polypeptide is capable of converting L-proline to *trans*-3-hydroxyproline with at least 1.2 fold, 1.5 fold, 2 fold, 3 fold, 4 fold, 5 fold, 10 fold, 20 fold, 30 fold, 40 fold, 50 fold, 60 fold, 70 fold, 80 fold, 90 fold, 100 fold or more the activity of the naturally occurring enzyme.

25. The engineered polypeptide of Claim 23, wherein said engineered polypeptide is capable of converting L-proline to *trans*-3-hydroxyproline with greater than 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99 % or more diastomeric excess of *trans*-3-hydroxyproline.

26. A polynucleotide encoding the engineered polypeptide of any of Claims 1-25.

27. A polynucleotide encoding the engineered polypeptide of Claim 22.

28. A polynucleotide encoding the engineered polypeptide of Claim 24.

29. The polynucleotide of any of Claims 26-28, wherein said polynucleotide comprises a nucleic acid sequence optimized for expression in *E. coli*.

30. An expression vector comprising the polynucleotide of any of Claims 26-29, optionally further comprising at least one control sequence.

31. The expression vector of Claim 30, wherein said vector comprises the engineered polypeptide of SEQ ID NO: 4, 116, 162, 322, 412, 492, 562, 598, and/or 630.

32. A host cell comprising the polynucleotide of any of Claims 26-29.

33. A host cell comprising the expression vector of Claim 30 or 31.

34. The host cell of Claim 32 or 33, wherein the host cell is *E. coli*.
35. A method of preparing an engineered polypeptide, comprising culturing the host cell of any of Claims 32-34, under conditions suitable for expression of the polypeptide.
36. The method of Claim 35, further comprising a step of isolating the engineered polypeptide.