

ENGINEERED NITROALDOLASES

[0001] The present application claims priority to US Prov. Pat. Appln. Ser. No. 63/350,128, filed June 8, 2022, which is hereby incorporated by reference in its entirety for all purposes.

TECHNICAL FIELD

[0002] The present invention provides engineered nitroaldolase polypeptides useful for the production of β -nitro alcohols and tertiary amino alcohols, as well as compositions and methods utilizing these engineered polypeptides.

REFERENCE TO SEQUENCE LISTING, TABLE OR COMPUTER PROGRAM

[0003] The official copy of the Sequence Listing is submitted concurrently with the specification as an XML file, with a file name of "CX2-230WO1 ST26.xml", a creation date of June 1, 2023, and a size of 2,633,728 bytes. The Sequence Listing filed is part of the specification and is incorporated in its entirety by reference herein.

BACKGROUND

[0004] Chiral tertiary amino alcohols are important building blocks in the pharmaceutical industry. In particular, chiral tertiary amino alcohols are intermediates for compounds used in the treatment of cystic fibrosis and chronic obstructive pulmonary disease (COPD). However, traditional chemical synthesis methods for chiral tertiary amino alcohols can have drawbacks, such as requiring long synthetic routes, utilizing expensive substrates, providing poor yields and selectivity, and generating toxic by-products.

[0005] Therefore, more cost effective and environmentally sensitive biocatalytic synthesis routes are of interest to efficiently produce these compounds. However, only a few biocatalytic routes are known to produce chiral tertiary amino alcohol compounds.

[0006] One approach may involve generation of chiral β -amino alcohols via a β -hydroxy- α -amino acid intermediate using a two-enzyme cascade (Dückers et al. Appl Microbiol Biotechnol, 2010, 88:409–424). However, in all reported literature examples, this route is limited to producing chiral secondary alcohols and not the desired chiral tertiary alcohol.

[0007] Another potential route is the enzyme-catalyzed enantioselective addition of nitromethane to trifluoroacetone to form a chiral tertiary nitroaldol product, followed by chemical reduction to the chiral amino alcohol. This is not a known natural reaction. However, there is some literature precedent of hydroxynitrile lyases and other enzymes catalyzing this reaction.

[0008] These include an (*R*)-selective nitroaldolase reaction catalyzed by hydroxynitrile lyases from *Acidobacterium capsulatum* and *Granulicella tundricula* (M. Bekerle-Bogner, M. Gruber-Khadjawi, H. Wiltsche, R. Wiedner, H. Schwab, K. Steiner, ChemCatChem 2016, 8, 2214). Henry (nitroaldol) reactions have also been reported using an hydroxynitrile lyase from *Hevea brasiliensis* (Gruber-Khadjawi, M., Purkarthofer, T., Skranc, W. and Griengl, H. (2007) Adv. Synth. Catal., 349: 1445-1450) and a

transglutaminase from *Streptoreverticillium griseovorticillatum* (Tang R., Guan Z., He Y., Zhu W., J of Molec Catal B: Enzymatic, 63: 1–2, 2010, 62-67). Other examples of similar enzyme-catalyzed reactions are known in the literature.

[0009] However, no current reports demonstrate activity of a hydroxynitrile lyase/nitroaldolase on a ketone substrate, and no current reports demonstrate highly (*S*)-selective activity of a hydroxynitrile lyase as a nitroaldolase. A nitroaldolase with (*S*)-selective activity on ketones would enable the synthesis of a range useful chiral amino alcohol compounds.

SUMMARY

[0010] The present invention provides novel biocatalysts and associated methods of use for the synthesis β -nitro alcohols and tertiary amino alcohols. The nitroaldolase biocatalysts of the present disclosure are engineered variants of a polypeptide (SEQ ID NO: 2) encoded by a homolog gene from *Baliospermum montanum* (SEQ ID NO:1). These engineered polypeptides are capable of catalyzing the conversion of trifluoroacetone and nitromethane to a β -nitro alcohol, which is a precursor to a tertiary amino alcohol product.

[0011] The present invention provides an engineered polypeptide comprising an amino acid sequence with at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NO: 2, comprising at least one substitution or one substitution set at one or more positions selected from 7/9, 7/9/33/68/72/158, 7/9/33/72/175/197/230, 7/9/33/72/230, 7/9/68/72/197, 7/9/158/175, 7/9/175, 7/9/197, 7/33, 7/68/72/175, 7/72, 7/72/97/175/230, 9, 9/11/33/72/197/230, 9/33/68/158/197, 9/33/72/230, 9/33/97/230, 9/68/72, 9/97/175/202/218, 9/175, 9/202, 11, 13, 14/33/68/72/97, 18, 18/33/68/97/158/175/239, 18/33/72/97/158/175, 18/33/72/97/158/239, 18/33/72/97/175/202, 18/33/72/175/202/237/239, 18/33/72/197/237, 18/68/72, 18/68/72/158/175/237/239, 18/68/72/175/202/239, 18/68/72/239, 18/68/175/223/237/239, 18/68/239, 18/72/97/175/237/239, 18/72/175/197, 18/72/237, 18/72/239, 18/97/175/237/239, 18/97/223/237/239, 18/175, 18/197/239, 18/239, 20/124, 33/68/72, 33/68/72/175/230, 33/68/223/239, 33/72, 33/72/97/197, 33/72/175, 33/175, 39, 44/97, 52, 68/72, 68/72/97, 68/72/97/197, 68/72/197, 68/72/197/239, 68/97/175, 68/97/175/197/223, 72/97/158/223/237, 72/175/197, 97/158/175/197, 97/175, 103, 121, 122, 124, 125, 128, 133, 148, 175, 197, and 197/202, wherein the positions are numbered with reference to SEQ ID NO: 2. In some additional embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from 7V/9V, 7V/9V/33V/68L/72E/158Y, 7V/9V/33V/72E/175P/197V/230I, 7V/9V/33V/72E/230I, 7V/9V/68L/72E/197V, 7V/9V/158Y/175P, 7V/9V/175P, 7V/9V/197V, 7V/33V, 7V/68L/72E/175P, 7V/72E, 7V/72E/97I/175P/230I, 9V, 9V/11G/33V/72E/197V/230I, 9V/33V/68L/158Y/197V, 9V/33V/72E/230I, 9V/33V/97I/230I, 9V/68L/72E, 9V/97I/175P/202I/218M, 9V/175P, 9V/202I, 11S, 13G, 14Y/33V/68L/72E/97I, 18C/33V/68L/97I/158Y/175P/239F, 18C/33V/72E/97I/158Y/175P, 18C/33V/72E/97I/158Y/239F, 18C/33V/72E/97I/175P/202I, 18C/33V/72E/175P/202I/237P/239F, 18C/33V/72E/197V/237P, 18C/68L/72E,

18C/68L/72E/158Y/175P/237P/239F, 18C/68L/72E/175P/202I/239F, 18C/68L/72E/239F, 18C/68L/175P/223P/237P/239F, 18C/68L/239F, 18C/72E/97I/175P/237P/239F, 18C/72E/175P/197V, 18C/72E/237P, 18C/72E/239F, 18C/97I/175P/237P/239F, 18C/97I/223P/237P/239F, 18C/175P, 18C/197V/239F, 18C/239F, 18I, 20H/124C, 33V/68L/72E, 33V/68L/72E/175P/230I, 33V/68L/223P/239F, 33V/72E, 33V/72E/97I/197V, 33V/72E/175P, 33V/175P, 39A, 39G, 44N/97I, 52S, 68L/72E, 68L/72E/97I, 68L/72E/97I/197V, 68L/72E/197V, 68L/72E/197V/239F, 68L/97I/175P, 68L/97I/175P/197V/223P, 72E/97I/158Y/223P/237P, 72E/175P/197V, 97I/158Y/175P/197V, 97I/175P, 103M, 103S, 121Y, 122C, 122Q, 124C, 124E, 124T, 124Y, 125R, 128H, 128Y, 133A, 133C, 148M, 175P, 197V, and 197V/202I, wherein the positions are numbered with reference to SEQ ID NO: 2. In some embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from I7V/I9V, I7V/I9V/A33V/I68L/K72E/F158Y, I7V/I9V/A33V/K72E/K175P/I197V/V230I, I7V/I9V/A33V/K72E/V230I, I7V/I9V/I68L/K72E/I197V, I7V/I9V/F158Y/K175P, I7V/I9V/K175P, I7V/I9V/I197V, I7V/A33V, I7V/I68L/K72E/K175P, I7V/K72E, I7V/K72E/V97I/K175P/V230I, I9V, I9V/T11G/A33V/K72E/I197V/V230I, I9V/A33V/I68L/F158Y/I197V, I9V/A33V/K72E/V230I, I9V/A33V/V97I/V230I, I9V/I68L/K72E, I9V/V97I/K175P/V202I/Q218M, I9V/K175P, I9V/V202I, T11S, C13G, H14Y/A33V/I68L/K72E/V97I, L18C/A33V/I68L/V97I/F158Y/K175P/I239F, L18C/A33V/K72E/V97I/F158Y/K175P, L18C/A33V/K72E/V97I/F158Y/I239F, L18C/A33V/K72E/V97I/K175P/V202I, L18C/A33V/K72E/K175P/V202I/I237P/I239F, L18C/A33V/K72E/I197V/I237P, L18C/I68L/K72E, L18C/I68L/K72E/F158Y/K175P/I237P/I239F, L18C/I68L/K72E/K175P/V202I/I239F, L18C/I68L/K72E/I239F, L18C/I68L/K175P/K223P/I237P/I239F, L18C/I68L/I239F, L18C/K72E/V97I/K175P/I237P/I239F, L18C/K72E/K175P/I197V, L18C/K72E/I237P, L18C/K72E/I239F, L18C/V97I/K175P/I237P/I239F, L18C/V97I/K223P/I237P/I239F, L18C/K175P, L18C/I197V/I239F, L18C/I239F, L18I, Y20H/V124C, A33V/I68L/K72E, A33V/I68L/K72E/K175P/V230I, A33V/I68L/K223P/I239F, A33V/K72E, A33V/K72E/V97I/I197V, A33V/K72E/K175P, A33V/K175P, V39A, V39G, D44N/V97I, G52S, I68L/K72E, I68L/K72E/V97I, I68L/K72E/V97I/I197V, I68L/K72E/I197V, I68L/K72E/I197V/I239F, I68L/V97I/K175P, I68L/V97I/K175P/I197V/K223P, K72E/V97I/F158Y/K223P/I237P, K72E/K175P/I197V, V97I/F158Y/K175P/I197V, V97I/K175P, H103M, H103S, F121Y, S122C, S122Q, V124C, V124E, V124T, V124Y, F125R, W128H, W128Y, F133A, F133C, L148M, K175P, I197V, and I197V/V202I, wherein the positions are numbered with reference to SEQ ID NO: 2. In some further embodiments, the engineered polypeptide comprises an amino acid sequence with at least 80% sequence identity to any even-numbered sequence set forth in SEQ ID NO: 4 to SEQ ID NO: 176.

[0012] The present invention also provides an engineered polypeptide comprising an amino acid sequence with at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NO: 2, comprising at least one substitution or one substitution set selected from: 7/11/33/72/97, 7/11/33/97, 7/11/72, 7/11/158, 7/11/237, 7/33/72/97/158/175/197/237/239, 9/11/33, 9/11/33/72, 9/11/72, 9/33/68/97/237/239,

9/72/97/158/237/239, 9/218/239, 10, 11, 12, 13, 14, 18, 33/68/72/97/237/239, 33/72/158/239, 39, 40, 41, 48, 49, 51, 54, 54/126, 57, 68/175/239, 82, 84, 85, 86, 103, 104, 105, 107, 115, 117, 119, 119/120, 120, 122, 123, 124, 127, 130, 131, 133, 134, 137/145, 144, 145, 146, 147, 148, 149, 150, 152, 157, 158, 159, 169, 173, 174, 175, 180, 181, 183, 203, 204, 208, 209, 210, 215, 218, 236, and 237, wherein the positions are numbered with reference to SEQ ID NO: 2. In some embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from 7V/11G/33V/72E/97I, 7V/11G/33V/97I, 7V/11G/72E, 7V/11G/158Y, 7V/11G/237P, 7V/33V/72E/97I/158Y/175P/197V/237P/239F, 9V/11G/33V, 9V/11G/33V/72E, 9V/11G/72E, 9V/33V/68L/97I/237P/239F, 9V/72E/97I/158Y/237P/239F, 9V/218M/239F, 10G, 10P, 10R, 11R, 11S, 12K, 13L, 14R, 18S, 18T, 33V/68L/72E/97I/237P/239F, 33V/72E/158Y/239F, 39R, 39W, 40L, 40S, 40T, 40V, 41C, 48W, 49Q, 49R, 51A, 51Q, 51S, 51T, 54N/126I, 54Y, 57N, 68L/175P/239F, 82T, 84L, 85A, 85L, 85M, 85Q, 86W, 103P, 104A, 104G, 104S, 105D, 107G, 107W, 115C, 115K, 115L, 115V, 117C, 117L, 119L, 119N/120F, 120C, 120L, 120Y, 122G, 122T, 123A, 123P, 123T, 124P, 127A, 127L, 127M, 130M, 130V, 131L, 133W, 134D, 134F, 134L, 134R, 134V, 134W, 137I/145V, 144L, 145F, 145L, 145P, 146D, 146G, 146N, 147R, 148P, 148V, 149W, 150I, 150L, 150V, 152P, 157V, 158I, 158L, 158M, 158R, 158W, 158Y, 159G, 159H, 159L, 169G, 169I, 169S, 173E, 173Y, 174C, 174G, 175D, 180G, 180V, 181C, 183K, 203Q, 204E, 204R, 204V, 208A, 208C, 208E, 208G, 208L, 208T, 208V, 209C, 210V, 215L, 218T, 236G, 236L, 237L, 237S, and 237W, wherein the positions are numbered with reference to SEQ ID NO: 2. In some further embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from I7V/T11G/A33V/K72E/V97I, I7V/T11G/A33V/V97I, I7V/T11G/K72E, I7V/T11G/F158Y, I7V/T11G/I237P, I7V/A33V/K72E/V97I/F158Y/K175P/I197V/I237P/I239F, I9V/T11G/A33V, I9V/T11G/A33V/K72E, I9V/T11G/K72E, I9V/A33V/I68L/V97I/I237P/I239F, I9V/K72E/V97I/F158Y/I237P/I239F, I9V/Q218M/I239F, H10G, H10P, H10R, T11R, T11S, I12K, C13L, H14R, L18S, L18T, A33V/I68L/K72E/V97I/I237P/I239F, A33V/K72E/F158Y/I239F, V39R, V39W, A40L, A40S, A40T, A40V, S41C, L48W, E49Q, E49R, I51A, I51Q, I51S, I51T, W54N/T126I, W54Y, Y57N, I68L/K175P/I239F, G82T, I84L, N85A, N85L, N85M, N85Q, I86W, H103P, N104A, N104G, N104S, A105D, M107G, M107W, A115C, A115K, A115L, A115V, V117C, V117L, K119L, K119N/K120F, K120C, K120L, K120Y, S122G, S122T, E123A, E123P, E123T, V124P, D127A, D127L, D127M, D130M, D130V, S131L, F133W, S134D, S134F, S134L, S134R, S134V, S134W, T137I/A145V, T144L, A145F, A145L, A145P, V146D, V146G, V146N, E147R, L148P, L148V, G149W, D150I, D150L, D150V, T152P, I157V, F158I, F158L, F158M, F158R, F158W, F158Y, S159G, S159H, S159L, A169G, A169I, A169S, V173E, V173Y, R174C, R174G, K175D, E180G, E180V, Q181C, L183K, Y203Q, G204E, G204R, G204V, Q208A, Q208C, Q208E, Q208G, Q208L, Q208T, Q208V, I209C, F210V, Q215L, Q218T, K236G, K236L, I237L, I237S, and I237W, wherein the positions are numbered with reference to SEQ ID NO: 2. In some further embodiments, the engineered polypeptide comprises an amino acid sequence with at least 80% sequence identity to any even-numbered sequence set forth in SEQ ID NO: 178 to SEQ ID NO: 462.

[0013] The present invention further provides an engineered polypeptide comprising an amino acid sequence with at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NO: 24, comprising at least at least one substitution or one substitution set selected from 9/39/122/157, 18, 18/33/39/230, 18/33/175, 18/39/44/118, 18/175, 20, 33/39, 33/39/44/118/121, 33/39/44/121, 33/39/44/152, 33/39/44/175, 33/39/44/197, 33/39/44/210/230, 33/39/118/121/175, 33/39/121/175, 33/39/197/210, 33/39/210, 33/39/210/230, 33/118/121, 33/118/121/175, 39, 39/44/103, 39/44/118, 39/44/152/230, 39/44/175/197/230, 39/103, 39/103/125, 39/103/125/127/146, 39/103/125/127/146/150, 39/103/125/146, 39/103/127/150, 39/103/150, 39/118/121, 39/197, 39/197/210, 43, 44/103, 81, 103, 103/125/127, 103/125/146, 103/127, 103/150, 118/121, 118/121/175, 118/121/197, 121, 152, 152/197, 154, 159, 163, 175, 210, 232, and 238, wherein the positions are numbered with reference to SEQ ID NO: 24. In some additional embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from 9V/39W/122T/157V, 18I, 18I/33V/39W/230I, 18I/33V/175P, 18I/39W/44N/118L, 18I/175P, 20H, 33V/39W, 33V/39W/44N/118L/121Y, 33V/39W/44N/121Y, 33V/39W/44N/152F, 33V/39W/44N/175P, 33V/39W/44N/197V, 33V/39W/44N/210M/230I, 33V/39W/118L/121Y/175P, 33V/39W/121Y/175P, 33V/39W/197V/210M, 33V/39W/210M, 33V/39W/210M/230I, 33V/118L/121W/175P, 33V/118L/121Y, 39W, 39W/44N/118L, 39W/44N/152F/230I, 39W/44N/175P/197V/230I, 39W/44Y/103S, 39W/103M, 39W/103M/125R/146L, 39W/103M/150P, 39W/103Q, 39W/103S, 39W/103S/125P/127G/146L/150P, 39W/103S/125R, 39W/103S/125R/127G/146L, 39W/103S/127G/150P, 39W/118L/121Y, 39W/197V, 39W/197V/210M, 43M, 43Q, 43R, 43S, 43T, 44N/103S, 81F, 103L, 103M, 103M/125P/146L, 103M/127G, 103Q, 103S, 103S/125P/127G, 103S/150P, 103V, 118L/121Y, 118L/121Y/175P, 118L/121Y/197V, 121Y, 152A/197V, 152F, 154I, 159Q, 163P, 175P, 210M, 232G, and 238M, wherein the positions are numbered with reference to SEQ ID NO: 24. In some further embodiments, the engineered polypeptide comprises at least at least one substitution or one substitution set selected from I9V/V39W/S122T/I157V, L18I, L18I/A33V/V39W/V230I, L18I/A33V/K175P, L18I/V39W/D44N/Y118L, L18I/K175P, Y20H, A33V/V39W, A33V/V39W/D44N/Y118L/F121Y, A33V/V39W/D44N/F121Y, A33V/V39W/D44N/T152F, A33V/V39W/D44N/K175P, A33V/V39W/D44N/I197V, A33V/V39W/D44N/F210M/V230I, A33V/V39W/Y118L/F121Y/K175P, A33V/V39W/F121Y/K175P, A33V/V39W/I197V/F210M, A33V/V39W/F210M, A33V/V39W/F210M/V230I, A33V/Y118L/F121W/K175P, A33V/Y118L/F121Y, V39W, V39W/D44N/Y118L, V39W/D44N/T152F/V230I, V39W/D44N/K175P/I197V/V230I, V39W/D44Y/H103S, V39W/H103M, V39W/H103M/F125R/V146L, V39W/H103M/D150P, V39W/H103Q, V39W/H103S, V39W/H103S/F125P/D127G/V146L/D150P, V39W/H103S/F125R, V39W/H103S/F125R/D127G/V146L, V39W/H103S/D127G/D150P, V39W/Y118L/F121Y, V39W/I197V, V39W/I197V/F210M, I43M, I43Q, I43R, I43S, I43T, D44N/H103S, G81F, H103L, H103M, H103M/F125P/V146L, H103M/D127G, H103Q, H103S, H103S/F125P/D127G, H103S/D150P, H103V, Y118L/F121Y, Y118L/F121Y/K175P, Y118L/F121Y/I197V, F121Y, T152A/I197V, T152F,

A154I, S159Q, I163P, K175P, F210M, S232G, and Q238M, wherein the positions are numbered with reference to SEQ ID NO: 24. In some additional embodiments, the engineered polypeptide comprises an amino acid sequence with at least 80% sequence identity to any even-numbered sequence set forth in SEQ ID NO: 464 to SEQ ID NO: 596.

[0014] The present invention further provides an engineered polypeptide comprising an amino acid sequence with at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NO: 24, comprising at least at least one substitution or one substitution set selected from 7, 9/13, 9/13/39, 9/13/39/68/72/239, 9/13/39/68/239, 9/13/39/72/122, 9/13/39/97/239, 9/13/68/72/97/239, 9/13/68/97/134/239, 9/13/68/97/239, 9/13/72, 9/13/134, 11, 12/42/125/127, 12/103, 12/103/150, 13, 13/39, 13/39/68, 13/39/68/72/97, 13/39/72/97/122/134/223, 13/39/97/239, 13/39/239, 13/68, 13/68/72/97/122/239, 16, 18/33/39/44/118/121/175/197/210, 18/33/39/44/121, 18/33/39/112/118/121/197, 18/33/39/121, 18/33/39/121/175, 18/33/39/121/175/197, 18/33/39/121/175/210, 18/33/39/121/210, 18/33/118/121, 18/33/121, 18/33/121/197, 18/33/121/210/230, 18/39/44/118/121/230, 18/39/44/121/175/210, 18/39/118/121, 18/39/118/121/210, 18/39/121/175, 18/39/121/175/210, 18/39/121/197/210, 18/118/121, 18/118/121/175, 18/118/121/210, 18/118/121/230, 18/121, 18/121/175, 18/121/210, 18/121/230, 20, 21, 33/39/44/118/121, 33/39/44/118/121/210/230, 33/39/44/121, 33/39/44/121/152, 33/39/44/121/175/210, 33/39/44/121/230, 33/39/73/121/197, 33/39/118/121/175, 33/39/118/121/175/230, 33/39/121, 33/39/121/175, 33/39/121/210/230, 33/118/121, 33/118/121/175, 33/118/121/197, 33/121, 33/121/230, 37/91, 39/44/118/121/175/210, 39/68/72/239, 39/118/121, 39/118/121/210, 39/121, 39/121/210/230, 43, 44, 76, 79, 101, 118/121, 118/121/175, 118/121/197, 121, 121/175/230, 121/210, 148, 153, 159, 161, 172, 200, 207, 233, 245, and 248, wherein the positions are numbered with reference to SEQ ID NO: 24. In some additional embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from 7E, 9V/13G, 9V/13G/39W, 9V/13G/39W/68L/72E/239F, 9V/13G/39W/68L/239F, 9V/13G/39W/72E/122T, 9V/13G/39W/97I/239F, 9V/13G/68L/72E/97I/239F, 9V/13G/68L/97I/134D/239F, 9V/13G/68L/97I/239F, 9V/13G/72E, 9V/13G/134D, 11N, 12A/42W/125R/127G, 12A/103S, 12A/103S/150P, 13F, 13G/39W, 13G/39W/68L, 13G/39W/68L/72E/97I, 13G/39W/72E/97I/122T/134D/223P, 13G/39W/97I/239F, 13G/39W/239F, 13G/68L, 13G/68L/72E/97I/122T/239F, 16M, 16V, 18C/33V/39W/112R/118L/121W/197V, 18C/33V/39W/121W/175P/210M, 18C/33V/39W/121W/210M, 18C/33V/39W/121Y/175P, 18C/33V/118L/121W, 18C/33V/118L/121Y, 18C/33V/121Y, 18C/33V/121Y/197V, 18C/33V/121Y/210M/230I, 18C/39W/44N/121Y/175P/210M, 18C/39W/118L/121W, 18C/39W/121W/175P/210M, 18C/39W/121W/197V/210M, 18C/118L/121W, 18C/118L/121Y, 18C/121V, 18C/121V/175P, 18C/121W, 18C/121W/175P, 18C/121W/210M, 18C/121Y, 18C/121Y/230I, 18I/33V/39W/44N/118L/121W/175P/197V/210M, 18I/33V/39W/44N/121T, 18I/33V/39W/121T/175P, 18I/33V/39W/121V/175P/197V, 18I/33V/39W/121W, 18I/33V/118L/121W, 18I/39W/44N/118L/121W/230I, 18I/39W/44N/121V/175P/210M, 18I/39W/118L/121W/210M,

18I/39W/121V/175P, 18I/118L/121T, 18I/118L/121W, 18I/118L/121W/175P, 18I/118L/121W/210M, 18I/118L/121Y, 18I/118L/121Y/230I, 18I/121T/175P, 18I/121W, 18I/121Y/210M, 20G, 20K, 20R, 20T, 21V, 33V/39G/44N/121W/152A, 33V/39W/44N/118L/121W/210M/230I, 33V/39W/44N/118L/121Y, 33V/39W/44N/121V, 33V/39W/44N/121Y, 33V/39W/44N/121Y/175P/210M, 33V/39W/44N/121Y/230I, 33V/39W/73N/121Y/197V, 33V/39W/118L/121W/175P, 33V/39W/118L/121Y/175P, 33V/39W/118L/121Y/175P/230I, 33V/39W/121V, 33V/39W/121W, 33V/39W/121W/210M/230I, 33V/39W/121Y/175P, 33V/118L/121W/175P, 33V/118L/121W/197V, 33V/118L/121Y, 33V/121W, 33V/121Y/230I, 37L/91G, 39G/68L/72E/239F, 39W/44N/118L/121W/175P/210M, 39W/118L/121W/210M, 39W/118L/121Y, 39W/118L/121Y/210M, 39W/121W, 39W/121W/210M/230I, 43E, 43Q, 43S, 43W, 44V, 76P, 79F, 101L, 118L/121W, 118L/121Y, 118L/121Y/175P, 118L/121Y/197V, 121V, 121W, 121Y, 121Y/175P/230I, 121Y/210M, 148F, 153W, 159Q, 161M, 172A, 172E, 172G, 172Q, 172R, 172V, 200G, 207N, 233R, 245D, 245G, 248G, and 248S, wherein the positions are numbered with reference to SEQ ID NO: 24. In some further embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from I7E, I9V/C13G, I9V/C13G/V39W, I9V/C13G/V39W/I68L/K72E/I239F, I9V/C13G/V39W/I68L/I239F, I9V/C13G/V39W/K72E/S122T, I9V/C13G/V39W/V97I/I239F, I9V/C13G/I68L/K72E/V97I/I239F, I9V/C13G/I68L/V97I/S134D/I239F, I9V/C13G/I68L/V97I/I239F, I9V/C13G/K72E, I9V/C13G/S134D, S11N, I12A/G42W/F125R/D127G, I12A/H103S, I12A/H103S/D150P, C13F, C13G/V39W, C13G/V39W/I68L, C13G/V39W/I68L/K72E/V97I, C13G/V39W/K72E/V97I/S122T/S134D/K223P, C13G/V39W/V97I/I239F, C13G/V39W/I239F, C13G/I68L, C13G/I68L/K72E/V97I/S122T/I239F, A16M, A16V, L18C/A33V/V39W/H112R/Y118L/F121W/I197V, L18C/A33V/V39W/F121W/K175P/F210M, L18C/A33V/V39W/F121W/F210M, L18C/A33V/V39W/F121Y/K175P, L18C/A33V/Y118L/F121W, L18C/A33V/Y118L/F121Y, L18C/A33V/F121Y, L18C/A33V/F121Y/I197V, L18C/A33V/F121Y/F210M/V230I, L18C/V39W/D44N/F121Y/K175P/F210M, L18C/V39W/Y118L/F121W, L18C/V39W/F121W/K175P/F210M, L18C/V39W/F121W/I197V/F210M, L18C/Y118L/F121W, L18C/Y118L/F121Y, L18C/F121V, L18C/F121V/K175P, L18C/F121W, L18C/F121W/K175P, L18C/F121W/F210M, L18C/F121Y, L18C/F121Y/V230I, L18I/A33V/V39W/D44N/Y118L/F121W/K175P/I197V/F210M, L18I/A33V/V39W/D44N/F121T, L18I/A33V/V39W/F121T/K175P, L18I/A33V/V39W/F121V/K175P/I197V, L18I/A33V/V39W/F121W, L18I/A33V/Y118L/F121W, L18I/V39W/D44N/Y118L/F121W/V230I, L18I/V39W/D44N/F121V/K175P/F210M, L18I/V39W/Y118L/F121W/F210M, L18I/V39W/F121V/K175P, L18I/Y118L/F121T, L18I/Y118L/F121W, L18I/Y118L/F121W/K175P, L18I/Y118L/F121W/F210M, L18I/Y118L/F121Y, L18I/Y118L/F121Y/V230I, L18I/F121T/K175P, L18I/F121W, L18I/F121Y/F210M, Y20G, Y20K, Y20R, Y20T, K21V, A33V/V39G/D44N/F121W/T152A, A33V/V39W/D44N/Y118L/F121W/F210M/V230I, A33V/V39W/D44N/Y118L/F121Y, A33V/V39W/D44N/F121V, A33V/V39W/D44N/F121Y,

A33V/V39W/D44N/F121Y/K175P/F210M, A33V/V39W/D44N/F121Y/V230I, A33V/V39W/K73N/F121Y/I197V, A33V/V39W/Y118L/F121W/K175P, A33V/V39W/Y118L/F121Y/K175P, A33V/V39W/Y118L/F121Y/K175P/V230I, A33V/V39W/F121V, A33V/V39W/F121W, A33V/V39W/F121W/F210M/V230I, A33V/V39W/F121Y/K175P, A33V/Y118L/F121W/K175P, A33V/Y118L/F121W/I197V, A33V/Y118L/F121Y, A33V/F121W, A33V/F121Y/V230I, D37L/E91G, V39G/I68L/K72E/I239F, V39W/D44N/Y118L/F121W/K175P/F210M, V39W/Y118L/F121W/F210M, V39W/Y118L/F121Y, V39W/Y118L/F121Y/F210M, V39W/F121W, V39W/F121W/F210M/V230I, I43E, I43Q, I43S, I43W, D44V, L76P, E79F, V101L, Y118L/F121W, Y118L/F121Y, Y118L/F121Y/K175P, Y118L/F121Y/I197V, F121V, F121W, F121Y, F121Y/K175P/V230I, F121Y/F210M, L148F, L153W, S159Q, S161M, L172A, L172E, L172G, L172Q, L172R, L172V, V200G, D207N, A233R, L245D, L245G, I248G, and I248S, wherein the positions are numbered with reference to SEQ ID NO: 24. In some additional embodiments, the engineered polypeptide comprises an amino acid sequence with at least 80% sequence identity to any even-numbered sequence set forth in SEQ ID NO: 598 to SEQ ID NO: 842.

[0015] The present invention further provides an engineered polypeptide comprising an amino acid sequence with at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NO: 496, comprising at least at least one substitution or one substitution set selected from 43/103, 43/103/172/238, 43/103/238/241, 43/238, 81/232, 103, 103/159/238, 103/194/238, 103/238, 122, 232, 238, and 238/241, wherein the positions are numbered with reference to SEQ ID NO: 496. In some additional embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from 43E/103S/238M/241R, 43M/103L, 43Q/103S/238M/241R, 43Q/238M, 43S/103S/172R/238M, 43S/238M, 81F/232G, 103L, 103S, 103S/159Q/238M, 103S/194L/238M, 103S/238M, 103V, 122C, 122I, 122K, 122L, 122M, 122Q, 122R, 122T, 122V, 232G, 238M, and 238M/241R, wherein the positions are numbered with reference to SEQ ID NO: 496. In some further embodiments, the engineered polypeptide comprises at least at least one substitution or one substitution set selected from I43E/H103S/Q238M/K241R, I43M/H103L, I43Q/H103S/Q238M/K241R, I43Q/Q238M, I43S/H103S/L172R/Q238M, I43S/Q238M, G81F/S232G, H103L, H103S, H103S/S159Q/Q238M, H103S/Y194L/Q238M, H103S/Q238M, H103V, S122C, S122I, S122K, S122L, S122M, S122Q, S122R, S122T, S122V, S232G, Q238M, and Q238M/K241R, wherein the positions are numbered with reference to SEQ ID NO: 496. In some additional embodiments, the engineered polypeptide comprises an amino acid sequence with at least 80% sequence identity to any even-numbered sequence set forth in SEQ ID NO: 844 to SEQ ID NO: 892.

[0016] The present invention further provides an engineered polypeptide comprising an amino acid sequence with at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NO: 496, comprising at least at least one substitution or one substitution set selected from 12, 43, 43/47/103, 43/47/103/232, 43/103, 43/103/121/172/241, 43/103/172, 43/103/172/238, 43/103/238, 43/103/238/241, 43/121, 43/194/238, 43/238, 44/103/121/238, 44/103/238,

44/103/238/241, 44/121/238/241, 47/103, 63, 103, 103/121, 103/121/172/238, 103/159/238, 103/172/238, 103/194/238, 103/238, 122, 131, 146, 178, 180, and 221, wherein the positions are numbered with reference to SEQ ID NO: 496. In some additional embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from 12V, 43E/103S/121W/172R/241R, 43E/103S/238M, 43E/103S/238M/241R, 43M/47G/103L, 43M/47G/103V, 43M/103L, 43M/103S, 43M/103V, 43Q/47G/103S/232G, 43Q/103L/172R, 43Q/103S, 43Q/103S/238M/241R, 43Q/121W, 43Q/238M, 43S, 43S/103L, 43S/103S, 43S/103S/172R, 43S/103S/172R/238M, 43S/103S/238M/241R, 43S/194L/238M, 43S/238M, 44V/103S/121W/238M, 44V/103S/238M, 44V/103S/238M/241R, 44V/121W/238M/241R, 47G/103L, 63P, 103L, 103S, 103S/121W, 103S/121W/172R/238M, 103S/159Q/238M, 103S/172R/238M, 103S/194L/238M, 103S/238M, 103V, 122H, 122I, 122K, 122L, 122M, 122N, 122R, 122T, 122V, 131Y, 146I, 178A, 180L, 180V, and 221G, wherein the positions are numbered with reference to SEQ ID NO: 496. In some further embodiments, the engineered polypeptide comprises at least at least one substitution or one substitution set selected from I12V, I43E/H103S/Y121W/L172R/K241R, I43E/H103S/Q238M, I43E/H103S/Q238M/K241R, I43M/Q47G/H103L, I43M/Q47G/H103V, I43M/H103L, I43M/H103S, I43M/H103V, I43Q/Q47G/H103S/S232G, I43Q/H103L/L172R, I43Q/H103S, I43Q/H103S/Q238M/K241R, I43Q/Y121W, I43Q/Q238M, I43S, I43S/H103L, I43S/H103S, I43S/H103S/L172R, I43S/H103S/L172R/Q238M, I43S/H103S/Q238M/K241R, I43S/Y194L/Q238M, I43S/Q238M, N44V/H103S/Y121W/Q238M, N44V/H103S/Q238M, N44V/H103S/Q238M/K241R, N44V/Y121W/Q238M/K241R, Q47G/H103L, T63P, H103L, H103S, H103S/Y121W, H103S/Y121W/L172R/Q238M, H103S/S159Q/Q238M, H103S/L172R/Q238M, H103S/Y194L/Q238M, H103S/Q238M, H103V, S122H, S122I, S122K, S122L, S122M, S122N, S122R, S122T, S122V, S131Y, V146I, F178A, E180L, E180V, and N221G, wherein the positions are numbered with reference to SEQ ID NO: 496. In some additional embodiments, the engineered polypeptide comprises an amino acid sequence with at least 80% sequence identity to any even-numbered sequence set forth in SEQ ID NO: 894 to SEQ ID NO: 980.

[0017] The present invention further provides an engineered polypeptide comprising an amino acid sequence with at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NO: 860, comprising at least at least one substitution or one substitution set selected from 12, 12/43/63/103/146/180/232/241, 12/43/63/103/146/241, 12/43/81/103/241, 12/43/81/146/232/241, 12/43/81/180/241, 12/43/81/232/241, 12/43/81/241, 12/43/103, 12/43/103/146/180/232/241, 12/43/103/146/180/241, 12/43/103/180/241, 12/43/103/241, 12/43/146/180/232/241, 12/43/180/241, 12/63/81/103/241, 12/81, 12/81/180/241, 12/103/146/241, 12/103/180/241, 12/146, 12/146/180/241, 12/180/232/241, 12/241, 18, 43/81/232/241, 43/103/146/180, 43/103/180/241, 43/103/241, 43/180/232, 43/232/241, 63/103/180/232, 78, 81, 81/103/146, 81/146/180/241, 81/146/241, 81/180/241, 84, 103, 103/146, 103/146/180, 103/146/180/241, 103/180, 103/180/241, 103/261, 144, 145, 146/180/232, 146/180/241, 146/241, 150, 166, 172, 180/232, 232/241,

and 241, wherein the positions are numbered with reference to SEQ ID NO: 860. In some additional embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from 12V, 12V/43M/63P/103L/146I/180L/232G/241R, 12V/43M/81F/103L/241R, 12V/43M/81F/146I/232G/241R, 12V/43M/81F/180V/241R, 12V/43M/81F/232G/241R, 12V/43M/81F/241R, 12V/43M/103L/180L/241R, 12V/43M/103L/241R, 12V/43M/103V/146I/180V/241R, 12V/43M/103V/241R, 12V/43M/146I/180L/232G/241R, 12V/43M/180L/241R, 12V/43M/180V/241R, 12V/43Q/63P/103L/146I/241R, 12V/43Q/103L/146I/180V/232G/241R, 12V/43Q/103L/146I/180V/241R, 12V/43Q/103L/180V/241R, 12V/43Q/103V, 12V/63P/81F/103L/241R, 12V/81F, 12V/81F/180V/241R, 12V/103L/146I/241R, 12V/103L/180L/241R, 12V/103V/180V/241R, 12V/146I, 12V/146I/180L/241R, 12V/180L/232G/241R, 12V/180V/232G/241R, 12V/241R, 18V, 43M/81F/232G/241R, 43M/103L/146I/180V, 43M/103L/180V/241R, 43M/103L/241R, 43M/103V/146I/180V, 43M/180L/232G, 43M/232G/241R, 43Q/103V/146I/180L, 63P/103L/180L/232G, 78A, 81F/103L/146I, 81F/146I/180L/241R, 81F/146I/241R, 81F/180L/241R, 81Q, 84Y, 103A, 103L, 103L/146I, 103L/146I/180V, 103L/180L, 103L/180V/241R, 103N, 103T, 103T/261V, 103V/146I/180L/241R, 144V, 145V, 146I/180L/241R, 146I/180V/232G, 146I/241R, 150H, 166L, 172L, 172V, 180L/232G, 232G/241R, and 241R, wherein the positions are numbered with reference to SEQ ID NO: 860. In some further embodiments, the engineered polypeptide comprises at least at least one substitution or one substitution set selected from I12V, I12V/S43M/T63P/S103L/V146I/E180L/S232G/K241R, I12V/S43M/G81F/S103L/K241R, I12V/S43M/G81F/V146I/S232G/K241R, I12V/S43M/G81F/E180V/K241R, I12V/S43M/G81F/S232G/K241R, I12V/S43M/G81F/K241R, I12V/S43M/S103L/E180L/K241R, I12V/S43M/S103L/K241R, I12V/S43M/S103V/V146I/E180V/K241R, I12V/S43M/S103V/K241R, I12V/S43M/V146I/E180L/S232G/K241R, I12V/S43M/E180L/K241R, I12V/S43M/E180V/K241R, I12V/S43Q/T63P/S103L/V146I/K241R, I12V/S43Q/S103L/V146I/E180V/S232G/K241R, I12V/S43Q/S103L/V146I/E180V/K241R, I12V/S43Q/S103L/E180V/K241R, I12V/S43Q/S103V, I12V/T63P/G81F/S103L/K241R, I12V/G81F, I12V/G81F/E180V/K241R, I12V/S103L/V146I/K241R, I12V/S103L/E180L/K241R, I12V/S103V/E180V/K241R, I12V/V146I, I12V/V146I/E180L/K241R, I12V/E180L/S232G/K241R, I12V/E180V/S232G/K241R, I12V/K241R, L18V, S43M/G81F/S232G/K241R, S43M/S103L/V146I/E180V, S43M/S103L/E180V/K241R, S43M/S103L/K241R, S43M/S103V/V146I/E180V, S43M/E180L/S232G, S43M/S232G/K241R, S43Q/S103V/V146I/E180L, T63P/S103L/E180L/S232G, G78A, G81F/S103L/V146I, G81F/V146I/E180L/K241R, G81F/V146I/K241R, G81F/E180L/K241R, G81Q, I84Y, S103A, S103L, S103L/V146I, S103L/V146I/E180V, S103L/E180L, S103L/E180V/K241R, S103N, S103T, S103T/A261V, S103V/V146I/E180L/K241R, T144V, A145V, V146I/E180L/K241R, V146I/E180V/S232G, V146I/K241R, D150H, V166L, R172L, R172V, E180L/S232G, S232G/K241R, and K241R, wherein the positions are numbered with reference to SEQ ID NO: 860. In some additional

embodiments, the engineered polypeptide comprises an amino acid sequence with at least 80% sequence identity to any even-numbered sequence set forth in SEQ ID NO: 982 to SEQ ID NO: 1118.

[0018] The present invention further provides an engineered polypeptide comprising an amino acid sequence with at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NO: 860, comprising at least at least one substitution or one substitution set selected from 9, 12, 12/43/63/103/146/180/232/241, 12/43/63/103/146/241, 12/43/103/146/180/232/241, 12/43/103/146/180/241, 12/43/103/146/232/241, 12/43/103/146/241, 12/43/103/180/241, 12/43/146/180/232/241, 12/43/146/241, 12/43/180/232, 12/43/180/241, 12/63/103/146/232/241, 12/103/146/241, 12/103/180/241, 12/146, 12/146/180/241, 12/146/232/241, 12/146/241, 12/180, 12/180/232/241, 12/180/241, 12/241, 18, 43/63/103/180/232/241, 43/103/146/180, 43/103/180, 43/103/180/232/241, 43/103/180/241, 43/180/232, 63/103/180/232, 103, 103/146/180, 103/146/180/241, 103/180, 103/180/241, 103/241, 118, 119, 122, 144, 145, 146/180/241, 154, 160, 172, 180/232, 180/232/241, 180/241, 208, 211, and 212, wherein the positions are numbered with reference to SEQ ID NO: 860. In some additional embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from 9V, 12V, 12V/43M/63P/103L/146I/180L/232G/241R, 12V/43M/103L/146I/241R, 12V/43M/103L/180L/241R, 12V/43M/103V/146I/180V/241R, 12V/43M/103V/146I/232G/241R, 12V/43M/146I/180L/232G/241R, 12V/43M/180L/232G, 12V/43M/180L/241R, 12V/43Q/63P/103V/146I/241R, 12V/43Q/103L/146I/180V/232G/241R, 12V/43Q/103L/146I/180V/241R, 12V/43Q/103L/146I/241R, 12V/43Q/103L/180V/241R, 12V/43Q/146I/241R, 12V/63P/103L/146I/232G/241R, 12V/103L/180L/241R, 12V/103L/180V/241R, 12V/103V/146I/241R, 12V/103V/180L/241R, 12V/103V/180V/241R, 12V/146I, 12V/146I/180L/241R, 12V/146I/232G/241R, 12V/146I/241R, 12V/180L, 12V/180L/232G/241R, 12V/180L/241R, 12V/180V, 12V/180V/232G/241R, 12V/180V/241R, 12V/241R, 18V, 43M/63P/103L/180V/232G/241R, 43M/103L/146I/180V, 43M/103L/180V, 43M/103L/180V/241R, 43M/103V/146I/180V, 43M/180L/232G, 43Q/103L/180V/232G/241R, 43Q/103V/146I/180L, 43Q/103V/180L/241R, 43Q/103V/180V/232G/241R, 63P/103L/180L/232G, 103A, 103L, 103L/146I/180V, 103L/146I/180V/241R, 103L/180L, 103L/180V/241R, 103L/241R, 103N, 103T, 103V/146I/180L/241R, 103V/180V/241R, 118L, 118V, 119H, 122H, 144H, 144V, 145V, 146I/180L/241R, 154R, 160L, 172C, 172L, 172V, 180L/232G, 180L/241R, 180V/232G, 180V/232G/241R, 208F, 211F, and 212P, wherein the positions are numbered with reference to SEQ ID NO: 860. In some further embodiments, the engineered polypeptide comprises at least at least one substitution or one substitution set selected from I9V, I12V, I12V/S43M/T63P/S103L/V146I/E180L/S232G/K241R, I12V/S43M/S103L/V146I/K241R, I12V/S43M/S103L/E180L/K241R, I12V/S43M/S103V/V146I/E180V/K241R, I12V/S43M/S103V/V146I/S232G/K241R, I12V/S43M/V146I/E180L/S232G/K241R, I12V/S43M/E180L/S232G, I12V/S43M/E180L/K241R, I12V/S43Q/T63P/S103V/V146I/K241R, I12V/S43Q/S103L/V146I/E180V/S232G/K241R, I12V/S43Q/S103L/V146I/E180V/K241R, I12V/S43Q/S103L/V146I/K241R, I12V/S43Q/S103L/E180V/K241R, I12V/S43Q/V146I/K241R,

I12V/T63P/S103L/V146I/S232G/K241R, I12V/S103L/E180L/K241R, I12V/S103L/E180V/K241R, I12V/S103V/V146I/K241R, I12V/S103V/E180L/K241R, I12V/S103V/E180V/K241R, I12V/V146I, I12V/V146I/E180L/K241R, I12V/V146I/S232G/K241R, I12V/V146I/K241R, I12V/E180L, I12V/E180L/S232G/K241R, I12V/E180L/K241R, I12V/E180V, I12V/E180V/S232G/K241R, I12V/E180V/K241R, I12V/K241R, L18V, S43M/T63P/S103L/E180V/S232G/K241R, S43M/S103L/V146I/E180V, S43M/S103L/E180V, S43M/S103L/E180V/K241R, S43M/S103V/V146I/E180V, S43M/E180L/S232G, S43Q/S103L/E180V/S232G/K241R, S43Q/S103V/V146I/E180L, S43Q/S103V/E180L/K241R, S43Q/S103V/E180V/S232G/K241R, T63P/S103L/E180L/S232G, S103A, S103L, S103L/V146I/E180V, S103L/V146I/E180V/K241R, S103L/E180L, S103L/E180V/K241R, S103L/K241R, S103N, S103T, S103V/V146I/E180L/K241R, S103V/E180V/K241R, Y118L, Y118V, K119H, S122H, T144H, T144V, A145V, V146I/E180L/K241R, A154R, N160L, R172C, R172L, R172V, E180L/S232G, E180L/K241R, E180V/S232G, E180V/S232G/K241R, Q208F, S211F, and R212P, wherein the positions are numbered with reference to SEQ ID NO: 860. In some additional embodiments, the engineered polypeptide comprises an amino acid sequence with at least 80% sequence identity to any even-numbered sequence set forth in SEQ ID NO: 1120 to SEQ ID NO: 1212.

[0019] The present invention further provides an engineered polypeptide comprising an amino acid sequence with at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NO: 1120, comprising at least at least one substitution or one substitution set selected from 3, 7, 7/158, 7/158/175, 7/158/197/202, 27/43/115/172, 27/43/172, 29, 32, 43/110/115/172, 44, 46, 48/172, 55, 59, 67, 71, 72, 73, 94/103/118/146, 103, 103/144/146/172/180, 103/146/154/172/180/232, 103/154, 118, 118/144/215, 118/172/180, 137, 142, 146, 146/154, 154, 158, 172/190, 184, 198, 205, 216, 223, 226, 229, 247, 261, and 263, wherein the positions are numbered with reference to SEQ ID NO: 1120. In some additional embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from 3D, 7V, 7V/158Y, 7V/158Y/175P, 7V/158Y/197V/202I, 27E/43I/115S/172L, 27E/43I/172L, 29G, 29R, 32G, 32R, 43I/110T/115S/172L, 44S, 46H, 46V, 48I/172L, 55A, 55G, 55R, 55S, 59G, 59T, 67N, 71S, 72G, 73R, 94Q/103L/118V/146I, 103L, 103L/144V/146I/172L/180V, 103L/146I/154R/172L/180V/232G, 103L/154R, 118V, 118V/144V/215H, 118V/172L/180V, 137M, 137R, 142V, 146I, 146I/154R, 154R, 158Y, 172L/190S, 184S, 198L, 205G, 205L, 216M, 216V, 223A, 223C, 223E, 223L, 223M, 223S, 226L, 226V, 229L, 229V, 229W, 247V, 261G, 261H, 263E, 263G, 263L, and 263S, wherein the positions are numbered with reference to SEQ ID NO: 1120. In some further embodiments, the engineered polypeptide comprises at least at least one substitution or one substitution set selected from S3D, I7V, I7V/F158Y, I7V/F158Y/K175P, I7V/F158Y/I197V/V202I, Q27E/S43I/A115S/R172L, Q27E/S43I/R172L, A29G, A29R, N32G, N32R, S43I/I110T/A115S/R172L, N44S, R46H, R46V, L48I/R172L, E55A, E55G, E55R, E55S, E59G, E59T, S67N, G71S, K72G, K73R, P94Q/V103L/Y118V/V146I, V103L, V103L/T144V/V146I/R172L/L180V, V103L/V146I/A154R/R172L/L180V/S232G, V103L/A154R,

Y118V, Y118V/T144V/Q215H, Y118V/R172L/L180V, T137M, T137R, T142V, V146I, V146I/A154R, A154R, F158Y, R172L/T190S, D184S, R198L, E205G, E205L, L216M, L216V, K223A, K223C, K223E, K223L, K223M, K223S, K226L, K226V, C229L, C229V, C229W, Q247V, A261G, A261H, A263E, A263G, A263L, and A263S, wherein the positions are numbered with reference to SEQ ID NO: 1120. In some additional embodiments, the engineered polypeptide comprises an amino acid sequence with at least 80% sequence identity to any even-numbered sequence set forth in SEQ ID NO: 1214 to SEQ ID NO: 1344.

[0020] The present invention further provides an engineered polypeptide comprising an amino acid sequence with at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NO: 1120, comprising at least at least one substitution or one substitution set selected from 73, 80, 103/146/209, 103/172/209, 118/119/172/209, 135, 144/209/232, 172/180/209, 181, and 223, wherein the positions are numbered with reference to SEQ ID NO: 1120. In some additional embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from 73R, 80A, 80G, 103L/146I/209C, 103L/172C/209C, 118V/119H/172C/209C, 135V, 144V/209C/232G, 172C/180V/209C, 181S, and 223E, wherein the positions are numbered with reference to SEQ ID NO: 1120. In some further embodiments, the engineered polypeptide comprises at least at least one substitution or one substitution set selected from K73R, S80A, S80G, V103L/V146I/I209C, V103L/R172C/I209C, Y118V/K119H/R172C/I209C, N135V, T144V/I209C/S232G, R172C/L180V/I209C, Q181S, and K223E, wherein the positions are numbered with reference to SEQ ID NO: 1120. In some additional embodiments, the engineered polypeptide comprises an amino acid sequence with at least 80% sequence identity to any even-numbered sequence set forth in SEQ ID NO: 1346 to SEQ ID NO: 1366.

[0021] The present invention further provides an engineered polypeptide comprising an amino acid sequence with at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NO: 1348, comprising at least at least one substitution or one substitution set selected from 2, 7/46, 39, 46, 46/55, 46/55/72, 46/55/92, 46/55/118, 46/55/226/257/263, 46/55/263, 46/70, 46/70/92/118/261/263, 46/70/175/263, 46/72, 46/118, 55/257/261, 55/261/263, 55/263, 105, 118, 118/226/261/263, and 263, wherein the positions are numbered with reference to SEQ ID NO: 1348. In some additional embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from 2G, 7V/46H, 39F, 39G, 39M, 39Q, 39S, 39T, 39V, 46H, 46H/55R, 46H/55R/72C, 46H/55R/92A, 46H/55R/118V, 46H/55R/226V/257Q/263L, 46H/55R/263L, 46H/70N, 46H/70N/92A/118V/261G/263L, 46H/70N/175P/263G, 46H/72C, 46H/118V, 55R/257Q/261G, 55R/261G/263I, 55R/263G, 105G, 118V, 118V/226V/261G/263I, and 263G, wherein the positions are numbered with reference to SEQ ID NO: 1348. In some further embodiments, the engineered polypeptide comprises at least at least one substitution or one substitution set selected from V2G, I7V/R46H, W39F, W39G, W39M, W39Q, W39S, W39T, W39V, R46H, R46H/E55R, R46H/E55R/K72C, R46H/E55R/K92A, R46H/E55R/Y118V, R46H/E55R/K226V/S257Q/A263L, R46H/E55R/A263L,

R46H/E70N, R46H/E70N/K92A/Y118V/A261G/A263L, R46H/E70N/K175P/A263G, R46H/K72C, R46H/Y118V, E55R/S257Q/A261G, E55R/A261G/A263I, E55R/A263G, A105G, Y118V, Y118V/K226V/A261G/A263I, and A263G, wherein the positions are numbered with reference to SEQ ID NO: 1348. In some additional embodiments, the engineered polypeptide comprises an amino acid sequence with at least 80% sequence identity to any even-numbered sequence set forth in SEQ ID NO: 1368 to SEQ ID NO: 1422.

[0022] The present invention further provides an engineered polypeptide comprising an amino acid sequence with at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NO: 1348, comprising at least at least one substitution or one substitution set selected from 2, 7/46, 7/46/175, 22/46/55/118, 27/46/72, 27/46/146/158/202, 27/46/175, 27/118, 29/70/118, 39, 46, 46/55, 46/55/59/261/263, 46/55/72, 46/55/92, 46/55/118, 46/55/118/263, 46/55/146, 46/55/146/202, 46/55/226/257/263, 46/55/263, 46/70, 46/70/92/118/261/263, 46/70/175/263, 46/72, 46/92, 46/92/175, 46/118, 46/146, 46/146/257/263, 55/70/118, 55/257/261, 55/257/263, 55/261/263, 55/263, 70, 92/118, 105, 118, 118/226/261/263, 146/257/263, 146/261/263, 192/257/261, 202, 257/261/263, 261/263, and 263, wherein the positions are numbered with reference to SEQ ID NO: 1348. In some additional embodiments, the engineered polypeptide comprises at least one substitution or one substitution set selected from 2G, 7V/46H, 7V/46H/175P, 22V/46H/55R/118V, 27G/46H/72C, 27G/46H/146I/158Y/202I, 27G/46H/175P, 27G/118V, 29V/70N/118V, 39M, 39Q, 39S, 39T, 39V, 46H, 46H/55R, 46H/55R/59T/261G/263L, 46H/55R/72C, 46H/55R/92A, 46H/55R/118V, 46H/55R/118V/263S, 46H/55R/146I, 46H/55R/146I/202I, 46H/55R/226V/257Q/263L, 46H/55R/263L, 46H/70N, 46H/70N/92A/118V/261G/263L, 46H/70N/175P/263G, 46H/72C, 46H/92A, 46H/92A/175P, 46H/118V, 46H/146I, 46H/146I/257Q/263I, 55R/70N/118V, 55R/257Q/261G, 55R/257Q/263L, 55R/261G/263I, 55R/263G, 70N, 92A/118V, 105G, 118V, 118V/226V/261G/263I, 146I/257Q/263S, 146I/261G/263L, 192V/257Q/261G, 202I, 257Q/261G/263G, 261G/263I, and 263G, wherein the positions are numbered with reference to SEQ ID NO: 1348. In some further embodiments, the engineered polypeptide comprises at least at least one substitution or one substitution set selected from V2G, I7V/R46H, I7V/R46H/K175P, L22V/R46H/E55R/Y118V, Q27G/R46H/K72C, Q27G/R46H/V146I/F158Y/V202I, Q27G/R46H/K175P, Q27G/Y118V, A29V/E70N/Y118V, W39M, W39Q, W39S, W39T, W39V, R46H, R46H/E55R, R46H/E55R/E59T/A261G/A263L, R46H/E55R/K72C, R46H/E55R/K92A, R46H/E55R/Y118V, R46H/E55R/Y118V/A263S, R46H/E55R/V146I, R46H/E55R/V146I/V202I, R46H/E55R/K226V/S257Q/A263L, R46H/E55R/A263L, R46H/E70N, R46H/E70N/K92A/Y118V/A261G/A263L, R46H/E70N/K175P/A263G, R46H/K72C, R46H/K92A, R46H/K92A/K175P, R46H/Y118V, R46H/V146I, R46H/V146I/S257Q/A263I, E55R/E70N/Y118V, E55R/S257Q/A261G, E55R/S257Q/A263L, E55R/A261G/A263I, E55R/A263G, E70N, K92A/Y118V, A105G, Y118V, Y118V/K226V/A261G/A263I, V146I/S257Q/A263S, V146I/A261G/A263L, E192V/S257Q/A261G, V202I, S257Q/A261G/A263G, A261G/A263I, and A263G, wherein the positions are numbered with reference to SEQ ID NO: 1348. In some additional

embodiments, the engineered polypeptide comprises an amino acid sequence with at least 80% sequence identity to any even-numbered sequence set forth in SEQ ID NO: 1424 to SEQ ID NO: 1484.

[0023] The present invention also provides an engineered polynucleotide encoding at least one engineered polypeptide described in the above paragraphs. In some embodiments, the engineered polynucleotide comprises the odd-numbered sequences set forth in SEQ ID NO: 3 to SEQ ID NO: 1483.

[0024] The present invention further provides vectors comprising at least one engineered polynucleotide described above. In some embodiments, the vectors further comprise at least one control sequence.

[0025] The present invention also provides host cells comprising the vectors provided herein. In some embodiments, the host cell produces at least one engineered polypeptide provided herein.

[0026] The present invention further provides methods of producing an engineered nitroaldolase polypeptide, comprising the steps of culturing the host cell provided herein under conditions such that the engineered polynucleotide is expressed and the engineered polypeptide is produced. In some embodiments, the methods further comprise the step of recovering the engineered polypeptide.

DESCRIPTION OF THE INVENTION

[0027] 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.

[0028] 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 invention as a whole.

[0029] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the present invention. The section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described. 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.

[0030] As used in this specification and the appended claims, the singular forms “a”, “an” and “the” include plural referents unless the context clearly indicates otherwise. Thus, for example, reference to “a polypeptide” includes more than one polypeptide. Similarly, “comprise,” “comprises,” “comprising” “include,” “includes,” and “including” are interchangeable and not intended to be limiting.

[0031] It is to be understood that where descriptions of various embodiments use the term “comprising,” those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language “consisting essentially of” or “consisting of.” It is to be further understood that where descriptions of various embodiments use the term “optional” or “optionally” the subsequently described event or circumstance may or may not occur, and that the description includes instances where the event or circumstance occurs and instances in which it does not. It is to be understood that both the foregoing general description, and the following detailed description are exemplary and explanatory only and are not restrictive of this disclosure. The section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described.

Abbreviations:

[0032] The abbreviations used for the genetically encoded amino acids are conventional and are as follows:

Amino Acid	Three-Letter	One-Letter Abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T

Amino Acid	Three-Letter	One-Letter Abbreviation
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

[0033] When the three-letter abbreviations are used, unless specifically preceded by an “L” or a “D” or clear from the context in which the abbreviation is used, the amino acid may be in either the L- or D-configuration about α -carbon (C_α). For example, whereas “Ala” designates alanine without specifying the configuration about the α -carbon, “D-Ala” and “L-Ala” designate D-alanine and L-alanine, respectively.

[0034] When the one-letter abbreviations are used, upper case letters designate amino acids in the L-configuration about the α -carbon and lower case letters designate amino acids in the D-configuration about the α -carbon. For example, “A” designates L-alanine and “a” designates D-alanine. When polypeptide sequences are presented as a string of one-letter or three-letter abbreviations (or mixtures thereof), the sequences are presented in the amino (N) to carboxy (C) direction in accordance with common convention.

[0035] The abbreviations used for the genetically encoding nucleosides are conventional and are as follows: adenosine (A); guanosine (G); cytidine (C); thymidine (T); and uridine (U). Unless specifically delineated, the abbreviated nucleotides may be either ribonucleosides or 2'-deoxyribonucleosides. The nucleosides may be specified as being either ribonucleosides or 2'-deoxyribonucleosides on an individual basis or on an aggregate basis. When nucleic acid sequences are presented as a string of one-letter abbreviations, the sequences are presented in the 5' to 3' direction in accordance with common convention, and the phosphates are not indicated.

Definitions:

[0036] In reference to the present invention, the technical and scientific terms used in the descriptions herein will have the meanings commonly understood by one of ordinary skill in the art, unless specifically defined otherwise. Accordingly, the following terms are intended to have the following meanings.

[0037] “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.

[0038] “ATCC” refers to the American Type Culture Collection whose biorepository collection includes genes and strains.

[0039] “NCBI” refers to National Center for Biological Information and the sequence databases provided therein.

[0040] “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, phosphorylation, lipidation, myristilation, ubiquitination, etc.). Included within this definition are D- and L-amino acids, and mixtures of D- and L-amino acids, as well as polymers comprising D- and L-amino acids, and mixtures of D- and L-amino acids.

[0041] “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.

[0042] As used herein, “polynucleotide” and “nucleic acid” refer to two or more nucleosides that are covalently linked together. The polynucleotide may be wholly comprised of ribonucleotides (i.e., RNA), wholly comprised of 2’ deoxyribonucleotides (i.e., DNA), or comprised of mixtures of ribo- and 2’ deoxyribonucleotides. While the nucleosides will typically be linked together via standard phosphodiester linkages, the polynucleotides may include one or more non-standard linkages. The polynucleotide may be single-stranded or double-stranded, or may include both single-stranded regions and double-stranded regions. Moreover, while a polynucleotide will typically be composed of the naturally occurring encoding nucleobases (i.e., adenine, guanine, uracil, thymine and cytosine), it may include one or more modified and/or synthetic nucleobases, such as, for example, inosine, xanthine, hypoxanthine, etc. In some embodiments, such modified or synthetic nucleobases are nucleobases encoding amino acid sequences.

[0043] “Nitroaldolase”, “hydroxynitrile lyase” or “oxynitrilase” refers to an enzyme categorized under EC 4.1.2.47 having native activity in the interconversion of cyanohydrins to cyanide and corresponding aldehydes and ketones. These enzymes are also known to have activity as nitroaldolases, in the conversion of a nitroalkane and an aldehyde to a β -nitro alcohol. While the nitroaldolase enzymes of the present invention are derived from *Baliospermum montanum*, the present invention is not thus limited, and nitroaldolase enzymes may be derived from any suitable organism or created synthetically. Nitroaldolases, as used herein, include naturally occurring (wild-type) nitroaldolases, as well as non-naturally occurring engineered polypeptides generated by human manipulation.

[0044] “Coding sequence” refers to that portion of a nucleic acid (e.g., a gene) that encodes an amino acid sequence of a protein.

[0045] “Naturally-occurring” or “wild-type” refers to the form found in nature. For example, a naturally occurring or 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.

[0046] As used herein, “recombinant,” “engineered,” and “non-naturally occurring” when used with reference to a cell, nucleic acid, or polypeptide, refer 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. In some embodiments, the cell, nucleic acid or polypeptide is identical a naturally occurring cell, nucleic acid or polypeptide, but is produced or derived from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant

cells expressing genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise expressed at a different level.

[0047] “Percentage of sequence identity” and “percentage homology” are used interchangeably herein to refer to comparisons among polynucleotides or 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., *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 (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.

[0048] “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, 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 polypeptide 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. For instance, a “reference sequence based on SEQ ID NO:4 having at the residue corresponding to X14 a valine” or X14V refers to a reference sequence in which the corresponding residue at X14 in SEQ ID NO:4, which is a tyrosine, has been changed to valine.

[0049] “Comparison window” refers to a conceptual segment of at least about 20 contiguous nucleotide positions or amino acids 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.

[0050] As used herein, “substantial identity” refers to a polynucleotide or polypeptide sequence that has at least 80 percent sequence identity, at least 85 percent identity, at least between 89 to 95 percent sequence identity, or more usually, at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 residue positions, frequently over a window of at least 30-50 residues, wherein the percentage of sequence identity is calculated by comparing the reference sequence to a sequence that includes deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. In some specific embodiments applied to polypeptides, the

term “substantial identity” means that two polypeptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80 percent sequence identity, preferably at least 89 percent sequence identity, at least 95 percent sequence identity or more (e.g., 99 percent sequence identity). In some embodiments, residue positions that are not identical in sequences being compared differ by conservative amino acid substitutions.

[0051] “Corresponding to,” “reference to,” and “relative to” when used in the context of the numbering of a given amino acid or polynucleotide sequence refer 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 nitroaldolase, 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.

[0052] “Amino acid difference” or “residue difference” refers to a change in the amino acid residue at a position of a polypeptide sequence relative to the amino acid residue at a corresponding position in a reference sequence. The positions of amino acid differences generally are referred to herein as “X_n,” where n refers to the corresponding position in the reference sequence upon which the residue difference is based. For example, a “residue difference at position X₂₅ as compared to SEQ ID NO: 2” refers to a change of the amino acid residue at the polypeptide position corresponding to position 25 of SEQ ID NO: 2. Thus, if the reference polypeptide of SEQ ID NO: 2 has a valine at position 25, then a “residue difference at position X₂₅ as compared to SEQ ID NO: 2” an amino acid substitution of any residue other than valine at the position of the polypeptide corresponding to position 25 of SEQ ID NO: 2. In most instances herein, the specific amino acid residue difference at a position is indicated as “X_nY” where “X_n” specified the corresponding position as described above, and “Y” is the single letter identifier of the amino acid found in the engineered polypeptide (i.e., the different residue than in the reference polypeptide). In some embodiments, more than one amino acid can appear in a specified residue position (i.e., the alternative amino acids can be listed in the form X_nY/Z, where Y and Z represent alternate amino acid residues). In some instances (e.g., in Tables 2-1, 2-2, 3-1, 3-2, 4-1, 4-2, 5-1, 5-2, 6-1, 6-2, 7-1, and 7-2) the present invention also provides specific amino acid differences denoted by the conventional notation “A_nB”, where A is the single letter identifier of the residue in the reference sequence, “n” is the number of the residue position in the reference sequence, and B is the single letter identifier of the residue substitution in the sequence of the engineered polypeptide. Furthermore, in some instances, a polypeptide of the present invention 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 changes are made relative to the

reference sequence. In some additional embodiments, the present invention provides engineered polypeptide sequences comprising both conservative and non-conservative amino acid substitutions. As used herein, “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 is substituted with another aliphatic amino acid (e.g., alanine, valine, leucine, and isoleucine); an amino acid with a hydroxyl side chain is substituted with another amino acid with a hydroxyl side chain (e.g., serine and threonine); an amino acid 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 basic 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. Exemplary conservative substitutions are provided in Table 1 below.

Table 1. Conservative Amino Acid Substitution Examples	
Residue	Possible Conservative Substitutions
A, L, V, I	Other aliphatic (A, L, V, I) Other non-polar (A, L, V, I, G, M)
G, M	Other non-polar (A, L, V, I, G, M)
D, E	Other acidic (D, E)
K, R	Other basic (K, R)
N, Q, S, T	Other polar
H, Y, W, F	Other aromatic (H, Y, W, F)
C, P	None

[0053] “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.

[0054] “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 nitroaldolase 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.

[0055] “Insertion” refers to modification to the polypeptide by addition of one or more amino acids from the reference polypeptide. In some embodiments, the improved engineered nitroaldolase enzymes comprise insertions of one or more amino acids to the naturally occurring nitroaldolase polypeptide as well as insertions of one or more amino acids to other improved nitroaldolase polypeptides. 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.

[0056] “Fragment” as used herein refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion, but where the remaining amino acid sequence is identical to the corresponding positions in the sequence. Fragments can be at least 14 amino acids long, at least 20 amino acids long, at least 50 amino acids long or longer, and up to 70%, 80%, 90%, 95%, 98%, and 99% of the full-length nitroaldolase polypeptide, for example the polypeptide of SEQ ID NO: 4 or a nitroaldolase polypeptide provided in the even-numbered sequences of SEQ ID NOs: 4-1484.

[0057] “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 engineered nitroaldolase enzymes 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 engineered nitroaldolase enzyme can be an isolated polypeptide.

[0058] “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. Generally, a substantially pure nitroaldolase composition will comprise 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 engineered nitroaldolase polypeptide is a substantially pure polypeptide composition.

[0059] “Stereoselective” refers to a preference for formation of one stereoisomer over another in a chemical or enzymatic reaction. Stereoselectivity can be partial, where the formation of one stereoisomer is favored over the other, or it may be complete where only one stereoisomer is formed. When the stereoisomers are enantiomers, the stereoselectivity is referred to as enantioselectivity, the fraction (typically reported as a percentage) of one enantiomer in the sum of both. It is commonly alternatively reported in the art (typically as a percentage) as the enantiomeric excess (e.e.) calculated therefrom according to the formula $[\text{major enantiomer} - \text{minor enantiomer}] / [\text{major enantiomer} + \text{minor enantiomer}]$. Where the stereoisomers are diastereoisomers, the stereoselectivity is referred to as diastereoselectivity, the fraction (typically reported as a percentage) of one diastereomer in a mixture of two diastereomers, commonly alternatively reported as the diastereomeric excess (d.e.). Enantiomeric excess and diastereomeric excess are types of stereomeric excess.

[0060] “Highly stereoselective” refers to a chemical or enzymatic reaction that is capable of converting a substrate or substrates (e.g., substrate compounds (2) and (3)), to the corresponding amine product (e.g., Compound (1)), with at least about 85% stereomeric excess.

[0061] As used herein, “improved enzyme property” refers to at least one improved property of an enzyme. In some embodiments, the present invention provides engineered nitroaldolase polypeptides that exhibit an improvement in any enzyme property as compared to a reference nitroaldolase polypeptide and/or a nitroaldolase polypeptide, and/or another engineered nitroaldolase polypeptide. For the engineered nitroaldolase polypeptides described herein, the comparison is generally made to the parent enzyme from which the nitroaldolase is derived, although in some embodiments, the reference enzyme can be another improved engineered nitroaldolase. Thus, the level of “improvement” can be determined and compared between various nitroaldolase polypeptides, including wild-type, as well as engineered nitroaldolases. Improved properties include, but are not limited, to such properties as enzymatic activity (which can be expressed in terms of percent conversion of the substrate), thermo stability, solvent stability, pH activity profile, cofactor requirements, refractoriness to inhibitors (e.g., substrate or product inhibition), stereospecificity, and/or stereoselectivity (including enantioselectivity).

[0062] “Increased enzymatic activity” refers to an improved property of the engineered nitroaldolase 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 nitroaldolase) as compared to the reference nitroaldolase 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 1.2 times the

enzymatic activity of the corresponding parent enzyme, to as much as 2 times, 5 times, 10 times, 20 times, 25 times, 50 times or more enzymatic activity than the naturally occurring or another engineered nitroaldolase from which the nitroaldolase polypeptides were derived. Nitroaldolase activity can be measured by any one of standard assays, such as by monitoring changes in properties of substrates, cofactors, or products. In some embodiments, the amount of products generated can be measured by Gas Chromatography (GC). Comparisons of enzyme activities are made using a defined preparation of enzyme, a defined assay under a set condition, and one or more defined substrates, as further described in detail herein. Generally, when lysates are compared, the numbers of cells and the amount of protein assayed are determined as well as use of identical expression systems and identical host cells to minimize variations in amount of enzyme produced by the host cells and present in the lysates.

[0063] “Conversion” refers to the enzymatic conversion of the 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 nitroaldolase polypeptide can be expressed as “percent conversion” of the substrate to the product.

[0064] “Thermostable” refers to a nitroaldolase polypeptide that maintains similar activity (more than 60% to 80% for example) after exposure to elevated temperatures (*e.g.*, 40-80°C) for a period of time (*e.g.*, 0.5-24 hrs) compared to the parent enzyme exposed to the same elevated temperature.

[0065] “Solvent stable” refers to a nitroaldolase polypeptide that maintains similar activity (more than *e.g.*, 60% to 80%) after exposure to varying concentrations (*e.g.*, 5-99%) of solvent (ethanol, isopropyl alcohol, dimethylsulfoxide (DMSO), tetrahydrofuran, 2-methyltetrahydrofuran, acetone, toluene, butyl acetate, methyl tert-butyl ether, etc.) for a period of time (*e.g.*, 0.5-24 hrs) compared to the parent enzyme exposed to the same concentration of the same solvent.

[0066] “Thermo- and solvent stable” refers to a nitroaldolase polypeptide that is both thermostable and solvent stable.

[0067] The term “stringent hybridization conditions” is used herein to refer to conditions under which nucleic acid hybrids are stable. As known to those of skill in the art, the stability of hybrids is reflected in the melting temperature (T_m) of the hybrids. In general, the stability of a hybrid is a function of ion strength, temperature, G/C content, and the presence of chaotropic agents. The T_m values for polynucleotides can be calculated using known methods for predicting melting temperatures (See *e.g.*, Baldino et al., *Meth. Enzymol.*, 168:761-777 [1989]; Bolton et al., *Proc. Natl. Acad. Sci. USA* 48:1390 [1962]; Bresslauer et al., *Proc. Natl. Acad. Sci. USA* 83:8893-8897 [1986]; Freier et al., *Proc. Natl. Acad. Sci. USA* 83:9373-9377 [1986]; Kierzek et al., *Biochem.*, 25:7840-7846 [1986]; Rychlik et al., 1990, *Nucl. Acids Res.*, 18:6409-6412 [1990] (erratum, *Nucl. Acids Res.*, 19:698 [1991]); Sambrook et al., *supra*); Suggs et al., 1981, in *Developmental Biology Using Purified Genes*, Brown et al. [eds.], pp. 683-693, Academic Press, Cambridge, MA [1981]; and Wetmur, *Crit. Rev. Biochem. Mol. Biol.*, 26:227-259 [1991]). In some embodiments, the polynucleotide encodes the polypeptide disclosed herein and

hybridizes under defined conditions, such as moderately stringent or highly stringent conditions, to the complement of a sequence encoding an engineered nitroaldolase enzyme of the present invention.

[0068] “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 0.018 M NaCl at 65°C (*i.e.*, if a hybrid is not stable in 0.018 M 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.

[0069] “Heterologous” polynucleotide refers to any polynucleotide that is introduced into a host cell by laboratory techniques, and includes polynucleotides that are removed from a host cell, subjected to laboratory manipulation, and then reintroduced into a host cell.

[0070] “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 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 nitroaldolase enzymes may be codon optimized for optimal production from the host organism selected for expression.

[0071] As used herein, “preferred, optimal, high codon usage bias codons” refers interchangeably to codons that are used at higher frequency in the protein coding regions than other codons that code for the same amino acid. The preferred codons may be determined in relation to codon usage in a single gene, a set of genes of common function or origin, highly expressed genes, the codon frequency in the aggregate

protein coding regions of the whole organism, codon frequency in the aggregate protein coding regions of related organisms, or combinations thereof. Codons whose frequency increases with the level of gene expression are typically optimal codons for expression. A variety of methods are known for determining the codon frequency (e.g., codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariate analysis, for example, using cluster analysis or correspondence analysis, and the effective number of codons used in a gene (See e.g., GCG CodonPreference, Genetics Computer Group Wisconsin Package; CodonW, Peden, University of Nottingham; McInerney, *Bioinform.*, 14:372-73 [1998]; Stenico et al., *Nucl. Acids Res.*, 22:2437-46 [1994]; Wright, *Gene* 87:23-29 [1990]). Codon usage tables are available for many different organisms (See e.g., Wada et al., *Nucl. Acids Res.*, 20:2111-2118 [1992]; Nakamura et al., *Nucl. Acids Res.*, 28:292 [2000]; Duret, et al., *supra*; Henaut and Danchin, in *Escherichia coli and Salmonella*, Neidhardt, et al. (eds.), ASM Press, Washington D.C., p. 2047-2066 [1996]). The data source for obtaining codon usage may rely on any available nucleotide sequence capable of coding for a protein. These data sets include nucleic acid sequences actually known to encode expressed proteins (e.g., complete protein coding sequences-CDS), expressed sequence tags (ESTs), or predicted coding regions of genomic sequences (See e.g., Mount, *Bioinformatics: Sequence and Genome Analysis*, Chapter 8, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. [2001]; Uberbacher, *Meth. Enzymol.*, 266:259-281 [1996]; and Tiwari et al., *Comput. Appl. Biosci.*, 13:263-270 [1997]).

[0072] “Control sequence” is defined herein to include all components, which are necessary or advantageous for the expression of a polynucleotide and/or polypeptide of the present invention. 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, signal peptide 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.

[0073] “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.

[0074] “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.

[0075] “Suitable reaction conditions” refer to those conditions in the biocatalytic reaction solution (*e.g.*, ranges of enzyme loading, substrate loading, cofactor loading, temperature, pH, buffers, co-solvents, etc.) under which the nitroaldolase polypeptide of the present invention is capable of converting one or more substrate compounds to a product compound (*e.g.*, conversion of compound (1) and compound (2) to compound (3), as shown in Scheme 1). Exemplary “suitable reaction conditions” are provided in the present invention and illustrated by the Examples.

[0076] “Composition” refers to a mixture or combination of one or more substances, wherein each substance or component of the composition retains its individual properties. As used herein, a biocatalytic composition refers to a combination of one or more substances useful for biocatalysis.

[0077] “Loading”, such as in “compound loading” or “enzyme loading” or “cofactor loading” refers to the concentration or amount of a component in a reaction mixture at the start of the reaction.

[0078] “Substrate” in the context of a biocatalyst mediated process refers to the compound or molecule acted on by the biocatalyst. For example, compound (1) and compound (2) are substrates for a nitroaldolase.

[0079] “Product” in the context of a biocatalyst mediated process refers to the compound or molecule resulting from the action of the biocatalyst. For example, compound (3) is a product of the nitroaldolase-mediated conversion of compound (1) and compound (2).

[0080] “Alkyl” refers to saturated hydrocarbon groups of from 1 to 18 carbon atoms inclusively, either straight chained or branched, more preferably from 1 to 8 carbon atoms inclusively, and most preferably 1 to 6 carbon atoms inclusively. An alkyl with a specified number of carbon atoms is denoted in parenthesis (*e.g.*, (C₁-C₆)alkyl refers to an alkyl of 1 to 6 carbon atoms).

[0081] “Alkenyl” refers to hydrocarbon groups of from 2 to 12 carbon atoms inclusively, either straight or branched containing at least one double bond but optionally containing more than one double bond.

[0082] “Alkynyl” refers to hydrocarbon groups of from 2 to 12 carbon atoms inclusively, either straight or branched containing at least one triple bond but optionally containing more than one triple bond, and additionally optionally containing one or more double bonded moieties.

[0083] “Alkylene” refers to a straight or branched chain divalent hydrocarbon radical having from 1 to 18 carbon atoms inclusively, more preferably from 1 to 8 carbon atoms inclusively, and most preferably 1 to 6 carbon atoms inclusively, optionally substituted with one or more suitable substituents. Exemplary “alkylenes” include, but are not limited to, methylene, ethylene, propylene, butylene, and the like.

[0084] “Alkenylene” refers to a straight or branched chain divalent hydrocarbon radical having 2 to 12 carbon atoms inclusively and one or more carbon-carbon double bonds, more preferably from 2 to 8 carbon atoms inclusively, and most preferably 2 to 6 carbon atoms inclusively, optionally substituted with one or more suitable substituents.

[0085] “Heteroalkyl,” “heteroalkenyl,” and “heteroalkynyl,” refer respectively, to alkyl, alkenyl and alkynyl as defined herein in which one or more of the carbon atoms are each independently replaced with the same or different heteroatoms or heteroatomic groups. Heteroatoms and/or heteroatomic groups

which can replace the carbon atoms include, but are not limited to, -O-, -S-, -S-O-, -NR^γ-, -PH-, -S(O)-, -S(O)₂-, -S(O)NR^γ-, -S(O)₂NR^γ-, and the like, including combinations thereof, where each R^γ is independently selected from hydrogen, alkyl, cycloalkyl, heterocycloalkyl, aryl, and heteroaryl.

[0086] “Aryl” refers to an unsaturated aromatic carbocyclic group of from 6 to 12 carbon atoms inclusively having a single ring (e.g., phenyl) or multiple condensed rings (e.g., naphthyl or anthryl). Exemplary aryls include phenyl, pyridyl, naphthyl and the like.

[0087] “Arylalkyl” refers to an alkyl substituted with an aryl (i.e., aryl-alkyl- groups), preferably having from 1 to 6 carbon atoms inclusively in the alkyl moiety and from 6 to 12 carbon atoms inclusively in the aryl moiety. Such arylalkyl groups are exemplified by benzyl, phenethyl and the like.

[0088] “Aryloxy” refers to -OR^λ groups, where R^λ is an aryl group, which can be optionally substituted.

[0089] “Cycloalkyl” refers to cyclic alkyl groups of from 3 to 12 carbon atoms inclusively having a single cyclic ring or multiple condensed rings which can be optionally substituted with from 1 to 3 alkyl groups. Exemplary cycloalkyl groups include, but are not limited to, single ring structures such as cyclopropyl, cyclobutyl, cyclopentyl, cyclooctyl, 1-methylcyclopropyl, 2-methylcyclopentyl, 2-methylcyclooctyl, and the like, or multiple ring structures, including bridged ring systems, such as adamantyl, and the like.

[0090] “Cycloalkylalkyl” refers to an alkyl substituted with a cycloalkyl (i.e., cycloalkyl-alkyl- groups), preferably having from 1 to 6 carbon atoms inclusively in the alkyl moiety and from 3 to 12 carbon atoms inclusively in the cycloalkyl moiety. Such cycloalkylalkyl groups are exemplified by cyclopropylmethyl, cyclohexylethyl and the like.

[0091] “Amino” refers to the group -NH₂. Substituted amino refers to the group -NHRⁿ, NRⁿRⁿ, and NRⁿRⁿRⁿ, where each Rⁿ is independently selected from substituted or unsubstituted alkyl, cycloalkyl, cycloheteroalkyl, alkoxy, aryl, heteroaryl, heteroarylalkyl, acyl, alkoxycarbonyl, sulfanyl, sulfinyl, sulfonyl, and the like. Typical amino groups include, but are limited to, dimethylamino, diethylamino, trimethylammonium, triethylammonium, methylsulfonylamino, furanyl-oxy-sulfamino, and the like.

[0092] “Aminoalkyl” refers to an alkyl group in which one or more of the hydrogen atoms are replaced with one or more amino groups, including substituted amino groups.

[0093] “Aminocarbonyl” refers to -C(O)NH₂. Substituted aminocarbonyl refers to -C(O)NRⁿRⁿ, where the amino group NRⁿRⁿ is as defined herein.

[0094] “Oxy” refers to a divalent group -O-, which may have various substituents to form different oxy groups, including ethers and esters.

[0095] “Alkoxy” or “alkyloxy” are used interchangeably herein to refer to the group -OR^ξ, wherein R^ξ is an alkyl group, including optionally substituted alkyl groups.

[0096] “Carboxy” refers to -COOH.

[0097] “Carbonyl” refers to -C(O)-, which may have a variety of substituents to form different carbonyl groups including acids, acid halides, aldehydes, amides, esters, and ketones.

[0098] “Carboxyalkyl” refers to an alkyl in which one or more of the hydrogen atoms are replaced with one or more carboxy groups.

[0099] “Aminocarbonylalkyl” refers to an alkyl substituted with an aminocarbonyl group, as defined herein.

[0100] “Halogen” or “halo” refers to fluoro, chloro, bromo and iodo.

[0101] “Haloalkyl” refers to an alkyl group in which one or more of the hydrogen atoms are replaced with a halogen. Thus, the term “haloalkyl” is meant to include monohaloalkyls, dihaloalkyls, trihaloalkyls, etc. up to perhaloalkyls. For example, the expression “(C₁ - C₂) haloalkyl” includes 1-fluoromethyl, difluoromethyl, trifluoromethyl, 1-fluoroethyl, 1,1-difluoroethyl, 1,2-difluoroethyl, 1,1,1 trifluoroethyl, perfluoroethyl, etc.

[0102] “Hydroxy” refers to -OH.

[0103] “Hydroxyalkyl” refers to an alkyl group in which one or more of the hydrogen atoms are replaced with one or more hydroxy groups.

[0104] “Thiol” or “sulfanyl” refers to -SH. Substituted thiol or sulfanyl refers to -S-Rⁿ, where Rⁿ is an alkyl, aryl or other suitable substituent.

[0105] “Alkylthio” refers to -SR^ξ, where R^ξ is an alkyl, which can be optionally substituted. Typical alkylthio group include, but are not limited to, methylthio, ethylthio, n-propylthio, and the like.

[0106] “Alkylthioalkyl” refers to an alkyl substituted with an alkylthio group, -SR^ξ, where R^ξ is an alkyl, which can be optionally substituted.

[0107] “Sulfonyl” refers to -SO₂-. Substituted sulfonyl refers to -SO₂-Rⁿ, where Rⁿ is an alkyl, aryl or other suitable substituent.

[0108] “Alkylsulfonyl” refers to -SO₂-R^ξ, where R^ξ is an alkyl, which can be optionally substituted. Typical alkylsulfonyl groups include, but are not limited to, methylsulfonyl, ethylsulfonyl, n-propylsulfonyl, and the like.

[0109] “Alkylsulfonylalkyl” refers to an alkyl substituted with an alkylsulfonyl group, -SO₂-R^ξ, where R^ξ is an alkyl, which can be optionally substituted.

[0110] “Heteroaryl” refers to an aromatic heterocyclic group of from 1 to 10 carbon atoms inclusively and 1 to 4 heteroatoms inclusively selected from oxygen, nitrogen and sulfur within the ring. Such heteroaryl groups can have a single ring (e.g., pyridyl or furyl) or multiple condensed rings (e.g., indoliziny or benzothienyl).

[0111] “Heteroarylalkyl” refers to an alkyl substituted with a heteroaryl (i.e., heteroaryl-alkyl- groups), preferably having from 1 to 6 carbon atoms inclusively in the alkyl moiety and from 5 to 12 ring atoms inclusively in the heteroaryl moiety. Such heteroarylalkyl groups are exemplified by pyridylmethyl and the like.

[0112] “Heterocycle”, “heterocyclic” and interchangeably “heterocycloalkyl” refer to a saturated or unsaturated group having a single ring or multiple condensed rings, from 2 to 10 carbon ring atoms inclusively and from 1 to 4 hetero ring atoms inclusively selected from nitrogen, sulfur or oxygen within

the ring. Such heterocyclic groups can have a single ring (e.g., piperidinyl or tetrahydrofuryl) or multiple condensed rings (e.g., indolinyl, dihydrobenzofuran or quinuclidinyl). Examples of heterocycles include, but are not limited to, furan, thiophene, thiazole, oxazole, pyrrole, imidazole, pyrazole, pyridine, pyrazine, pyrimidine, pyridazine, indolizine, isoindole, indole, indazole, purine, quinolizine, isoquinoline, quinoline, phthalazine, naphthylpyridine, quinoxaline, quinazoline, cinnoline, pteridine, carbazole, carboline, phenanthridine, acridine, phenanthroline, isothiazole, phenazine, isoxazole, phenoxazine, phenothiazine, imidazolidine, imidazoline, piperidine, piperazine, pyrrolidine, indoline and the like.

[0113] “Heterocycloalkylalkyl” refers to an alkyl substituted with a heterocycloalkyl (i.e., heterocycloalkyl-alkyl- groups), preferably having from 1 to 6 carbon atoms inclusively in the alkyl moiety and from 3 to 12 ring atoms inclusively in the heterocycloalkyl moiety.

[0114] “Membered ring” is meant to embrace any cyclic structure. The number preceding the term “membered” denotes the number of skeletal atoms that constitute the ring. Thus, for example, cyclohexyl, pyridine, pyran and thiopyran are 6-membered rings and cyclopentyl, pyrrole, furan, and thiophene are 5-membered rings.

[0115] “Fused bicyclic ring” as used herein refers to both unsubstituted and substituted carbocyclic and/or heterocyclic ring moieties having 5 to 8 atoms in each ring, the rings having 2 common atoms.

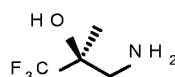
[0116] “Optionally substituted” as used herein with respect to the foregoing chemical groups means that positions of the chemical group occupied by hydrogen can be substituted with another atom (unless otherwise specified) exemplified by, but not limited to carbon, oxygen, nitrogen, or sulfur, or a chemical group, exemplified by, but not limited to, hydroxy, oxo, nitro, methoxy, ethoxy, alkoxy, substituted alkoxy, trifluoromethoxy, haloalkoxy, fluoro, chloro, bromo, iodo, halo, methyl, ethyl, propyl, butyl, alkyl, alkenyl, alkynyl, substituted alkyl, trifluoromethyl, haloalkyl, hydroxyalkyl, alkoxyalkyl, thio, alkylthio, acyl, carboxy, alkoxycarbonyl, carboxamido, substituted carboxamido, alkylsulfonfyl, alkylsulfanyl, alkylsulfonfylamino, sulfonamido, substituted sulfonamido, cyano, amino, substituted amino, alkylamino, dialkylamino, aminoalkyl, acylamino, amidino, amidoximo, hydroxamoyl, phenyl, aryl, substituted aryl, aryloxy, arylalkyl, arylalkenyl, arylalkynyl, pyridyl, imidazolyl, heteroaryl, substituted heteroaryl, heteroaryloxy, heteroarylalkyl, heteroarylalkenyl, heteroarylalkynyl, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloalkyl, cycloalkenyl, cycloalkylalkyl, substituted cycloalkyl, cycloalkyloxy, pyrrolidinyl, piperidinyl, morpholino, heterocycle, (heterocycle)oxy, and (heterocycle)alkyl; where preferred heteroatoms are oxygen, nitrogen, and sulfur. Additionally, where open valences exist on these substitute chemical groups they can be further substituted with alkyl, cycloalkyl, aryl, heteroaryl, and/or heterocycle groups, that where these open valences exist on carbon they can be further substituted by halogen and by oxygen-, nitrogen-, or sulfur-bonded substituents, and where multiple such open valences exist, these groups can be joined to form a ring, either by direct formation of a bond or by formation of bonds to a new heteroatom, preferably oxygen, nitrogen, or sulfur. It is further contemplated that the above substitutions can be made provided that replacing the hydrogen with the substituent does not introduce unacceptable instability to the molecules of the present invention, and is otherwise chemically

reasonable. One of ordinary skill in the art would understand that with respect to any chemical group described as optionally substituted, only sterically practical and/or synthetically feasible chemical groups are meant to be included. “Optionally substituted” as used herein refers to all subsequent modifiers in a term or series of chemical groups. For example, in the term “optionally substituted arylalkyl,” the “alkyl” portion and the “aryl” portion of the molecule may or may not be substituted, and for the series “optionally substituted alkyl, cycloalkyl, aryl and heteroaryl,” the alkyl, cycloalkyl, aryl, and heteroaryl groups, independently of the others, may or may not be substituted.

[0117] “Reaction” as used herein refers to a process in which one or more substances or compounds or substrates is converted into one or more different substances, compounds, or processes.

Conversion of Nitromethane and Trifluoroacetone to a β -nitro Alcohol

[0118] Compound (4), (*S*)-3-amino-1,1,1-trifluoro-2-methylpropan-2-ol, is an intermediate in the synthesis of a novel compound for the treatment of cystic fibrosis.



Compound (4)

[0119] Enzymatic synthesis of compound (4), a β -hydroxy- α -amino alcohol, presents an attractive alternative to traditional chemical synthesis. Use of enzyme biocatalysts may reduce chemical waste and allow a more efficient synthesis method.

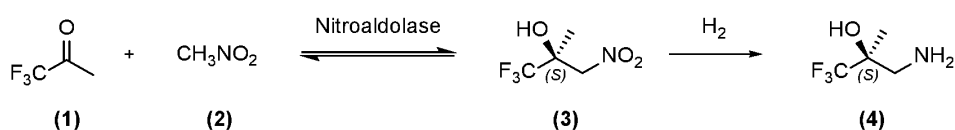
[0120] Generation of chiral β -amino alcohols via a β -hydroxy- α -amino acid intermediate using a two-enzyme cascade has been reported by others (Dückers et al. Appl Microbiol Biotechnol, 2010, 88:409–424). In the first step, a threonine aldolase catalyzes the conversion of glycine (or another nucleophile) and acetaldehyde (or another electrophile) to threonine in a PLP-dependent mechanism. In the second step, an amino acid decarboxylase catalyzes the removal of the carboxyl group to generate the amino alcohol (which is also a PLP-dependent mechanism). We previously reported a two enzyme cascade for production of compound (4), utilizing engineered threonine aldolases and engineered amino acid decarboxylases (US 63/285,377).

[0121] However, other biocatalytic routes have been proposed and may provide additional utility and efficiency under industrial synthesis conditions. These include use of an oxynitrilase (also known as hydroxynitrile lyase) to produce a cyanohydrin or β -nitro alcohol intermediate, followed by reduction to the β -hydroxy- α -amino alcohol of compound (4). The present disclosure details the use of oxynitrilases to perform the nitroaldol conversion of nitromethane and trifluoroacetone to a β -nitro alcohol intermediate. In this disclosure, the term nitroaldolase is used to describe the engineered oxynitrilases/hydroxynitrile lyases to highlight the improvements in enzyme functionality toward to the nitro aldol reaction.

[0122] Oxynitrilases are found mainly in plants, often stone fruit plants (e.g., almond, apricot), where the conversion of naturally occurring cyanogenic glycoside or cyano lipid cyanohydrins to cyanide is used as a defense mechanism against herbivores and microorganisms. Both (*S*) and (*R*)-selective oxynitrilases are known. As members of the α/β hydrolase fold family, oxynitrilases have a catalytic triad similar to esterases. Native substrates are thought to be acetone and benzaldehyde. Some oxynitrilases contain a non-catalytic FAD, which acts as a structural cofactor.

[0123] A proposed mechanism of oxynitrilases shows the activated serine in the catalytic triad acting as a general acid/base and not as a nucleophile (Gruber et al. J. Biol. Chem., 2004, 279:20501-20510). This is uncommon for members of the α/β hydrolase fold family, where serine usually acts as a nucleophile and forms an acyl-enzyme intermediate.

[0124] The present disclosure provides novel biocatalysts and associated methods of use for the synthesis of chiral β -amino alcohols via a β -nitro alcohol intermediate followed by chemical reduction. The nitroaldolase biocatalysts of the present disclosure are engineered variants of a polypeptide (SEQ ID NO: 2) encoded by a homolog gene from *Baliospermum montanum* (SEQ ID NO: 1). These engineered polypeptides are capable of catalyzing the conversion of trifluoroacetone (compound (1)) and nitromethane (compound (2)) to compound (4) via intermediate compound (3), as depicted below in Scheme 1.



Scheme 1.

[0125] In some embodiments, the present disclosure provides nitroaldolase enzymes having improved activity in the production of compound (3), as compared to a reference polypeptide. In some embodiments, the engineered nitroaldolase enzymes have the activity of Scheme 1.

[0126] As described herein, the engineered polypeptides exhibit stereoselectivity; thus, a nitroaldolase of Scheme 1 can be used to establish one, or more, chiral centers of a product. The nitroaldolases of the present disclosure produce (*R*) or (*S*) enantiomers.

[0127] In some embodiments, the present disclosure provides engineered nitroaldolase enzymes having improved stereoselectivity towards the (*S*)- β -nitro alcohol product (compound (3)) as compared to a reference polypeptide. In some embodiments, the present disclosure provides an engineered polypeptide comprising an amino acid sequence having at least 80% sequence identity to an amino acid reference sequence of SEQ ID NO: 2, 24, 496, 860, 1120, or 1348, and further comprising one or more amino acid residue differences as compared to the reference amino acid sequence, wherein the engineered nitroaldolase polypeptide has increased enantioselectivity towards compound (3).

[0128] In some embodiments, the present disclosure provides an engineered polypeptide comprising an amino acid sequence having at least 80% sequence identity to an amino acid reference sequence of SEQ ID NO: 2, 24, 496, 860, 1120, or 1348, and further comprising one or more amino acid residue differences as compared to the reference amino acid sequence, wherein the engineered nitroaldolase polypeptide has increased activity in the conversion of compound (1) and compound (2) to compound (3).

Engineered Nitroaldolase Polypeptides

[0129] The present invention provides nitroaldolase polypeptides, polynucleotides encoding the polypeptides, methods of preparing the polypeptides, and methods for using the polypeptides. Where the description relates to polypeptides, it is to be understood that it can describe the polynucleotides encoding the polypeptides.

[0130] Suitable reaction conditions under which the above-described improved properties of the engineered polypeptides carry out the desired reaction can be determined with respect to concentrations or amounts of polypeptide, substrate, co-substrate, buffer, solvent, pH, conditions including temperature and reaction time, and/or conditions with the polypeptide immobilized on a solid support, as further described below and in the Examples.

[0131] In some embodiments, exemplary engineered nitroaldolases with improved properties, particularly in the stereoselective conversion of compound (1) and compound (2) to compound (3), comprise an amino acid sequence that has one or more residue differences as compared to SEQ ID NO: 2, 24, 496, 860, 1120, or 1348 at the residue positions indicated in Tables 2-1, 2-2, 3-1, 3-2, 4-1, 4-2, 5-1, 5-2, 6-1, 6-2, 7-1, and 7-2.

[0132] The structure and function information for exemplary non-naturally occurring (or engineered) polypeptides of the present invention are based on the conversion of compound (1) and compound (2) to compound (3), the results of which are shown below in Tables 2-1, 2-2, 3-1, 3-2, 4-1, 4-2, 5-1, 5-2, 6-1, 6-2, 7-1, and 7-2, as further described in the Examples. The odd numbered sequence identifiers (i.e., SEQ ID NOs) in these Tables refer to the nucleotide sequence encoding the amino acid sequence provided by the even numbered SEQ ID NOs in these Tables. Exemplary sequences are provided in the electronic sequence listing file accompanying this invention, which is hereby incorporated by reference herein. The amino acid residue differences are based on comparison to the reference sequence of SEQ ID NOs: 2, 24, 496, 860, 1120, and/or 1348, as indicated.

[0133] A nitroaldolase (SEQ ID NO:2) homolog from *Baliospermum montanum* was selected based on conversion of compound (1) and compound (2) to compound (3). SEQ ID NO: 1 is a codon-optimized polynucleotide for expression in *Escherichia coli* that was synthesized based on the polypeptide sequence of SEQ ID NO: 2.

[0134] The activity and stereoselectivity of each engineered nitroaldolase polypeptide relative to the reference polypeptide of SEQ ID NOs: 2, 24, 496, 860, 1120, and/or 1348 was determined as conversion of the substrates described in the Examples herein. In some embodiments, a shake flask powder (SFP) is used as a secondary screen to assess the properties of the engineered nitroaldolases, the results of which

are provided in the Examples. In some embodiments, the SFP forms provide a more purified powder preparation of the engineered polypeptides and can contain the engineered polypeptides that are up to about 30% of total protein.

[0135] In some embodiments, the specific enzyme properties are associated with the residues differences as compared to SEQ ID NOs: 2, 24, 496, 860, 1120, and/or 1348 at the residue positions indicated herein. In some embodiments, residue differences affecting polypeptide expression can be used to increase expression of the engineered nitroaldolase.

[0136] In light of the guidance provided herein, it is further contemplated that any of the exemplary engineered polypeptides comprising the even-numbered sequences of SEQ ID NOs: 4-1484 find use as the starting amino acid sequence for synthesizing other engineered nitroaldolase polypeptides, for example by subsequent rounds of evolution that incorporate new combinations of various amino acid differences from other polypeptides in Tables 2-1, 2-2, 3-1, 3-2, 4-1, 4-2, 5-1, 5-2, 6-1, 6-2, 7-1, and 7-2, 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.

[0137] In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 2 and one or more residue differences as compared to SEQ ID NO: 2, selected from: 7/9, 7/9/33/68/72/158, 7/9/33/72/175/197/230, 7/9/33/72/230, 7/9/68/72/197, 7/9/158/175, 7/9/175, 7/9/197, 7/33, 7/68/72/175, 7/72, 7/72/97/175/230, 9, 9/11/33/72/197/230, 9/33/68/158/197, 9/33/72/230, 9/33/97/230, 9/68/72, 9/97/175/202/218, 9/175, 9/202, 11, 13, 14/33/68/72/97, 18, 18/33/68/97/158/175/239, 18/33/72/97/158/175, 18/33/72/97/158/239, 18/33/72/97/175/202, 18/33/72/175/202/237/239, 18/33/72/197/237, 18/68/72, 18/68/72/158/175/237/239, 18/68/72/175/202/239, 18/68/72/239, 18/68/175/223/237/239, 18/68/239, 18/72/97/175/237/239, 18/72/175/197, 18/72/237, 18/72/239, 18/97/175/237/239, 18/97/223/237/239, 18/175, 18/197/239, 18/239, 20/124, 33/68/72, 33/68/72/175/230, 33/68/223/239, 33/72, 33/72/97/197, 33/72/175, 33/175, 39, 44/97, 52, 68/72, 68/72/97, 68/72/97/197, 68/72/197, 68/72/197/239, 68/97/175, 68/97/175/197/223, 72/97/158/223/237, 72/175/197, 97/158/175/197, 97/175, 103, 121, 122, 124, 125, 128, 133, 148, 175, 197, and 197/202. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 2 and one or more residue differences as compared to SEQ ID NO: 2, selected from: 7V/9V, 7V/9V/33V/68L/72E/158Y, 7V/9V/33V/72E/175P/197V/230I, 7V/9V/33V/72E/230I, 7V/9V/68L/72E/197V, 7V/9V/158Y/175P, 7V/9V/175P, 7V/9V/197V, 7V/33V, 7V/68L/72E/175P, 7V/72E, 7V/72E/97I/175P/230I, 9V, 9V/11G/33V/72E/197V/230I, 9V/33V/68L/158Y/197V, 9V/33V/72E/230I, 9V/33V/97I/230I, 9V/68L/72E, 9V/97I/175P/202I/218M, 9V/175P, 9V/202I, 11S, 13G, 14Y/33V/68L/72E/97I, 18C/33V/68L/97I/158Y/175P/239F, 18C/33V/72E/97I/158Y/175P, 18C/33V/72E/97I/158Y/239F,

18C/33V/72E/97I/175P/202I, 18C/33V/72E/175P/202I/237P/239F, 18C/33V/72E/197V/237P, 18C/68L/72E, 18C/68L/72E/158Y/175P/237P/239F, 18C/68L/72E/175P/202I/239F, 18C/68L/72E/239F, 18C/68L/175P/223P/237P/239F, 18C/68L/239F, 18C/72E/97I/175P/237P/239F, 18C/72E/175P/197V, 18C/72E/237P, 18C/72E/239F, 18C/97I/175P/237P/239F, 18C/97I/223P/237P/239F, 18C/175P, 18C/197V/239F, 18C/239F, 18I, 20H/124C, 33V/68L/72E, 33V/68L/72E/175P/230I, 33V/68L/223P/239F, 33V/72E, 33V/72E/97I/197V, 33V/72E/175P, 33V/175P, 39A, 39G, 44N/97I, 52S, 68L/72E, 68L/72E/97I, 68L/72E/97I/197V, 68L/72E/197V, 68L/72E/197V/239F, 68L/97I/175P, 68L/97I/175P/197V/223P, 72E/97I/158Y/223P/237P, 72E/175P/197V, 97I/158Y/175P/197V, 97I/175P, 103M, 103S, 121Y, 122C, 122Q, 124C, 124E, 124T, 124Y, 125R, 128H, 128Y, 133A, 133C, 148M, 175P, 197V, and 197V/202I. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 2 and one or more residue differences as compared to SEQ ID NO: 2, selected from: I7V/I9V, I7V/I9V/A33V/I68L/K72E/F158Y, I7V/I9V/A33V/K72E/K175P/I197V/V230I, I7V/I9V/A33V/K72E/V230I, I7V/I9V/I68L/K72E/I197V, I7V/I9V/F158Y/K175P, I7V/I9V/K175P, I7V/I9V/I197V, I7V/A33V, I7V/I68L/K72E/K175P, I7V/K72E, I7V/K72E/V97I/K175P/V230I, I9V, I9V/T11G/A33V/K72E/I197V/V230I, I9V/A33V/I68L/F158Y/I197V, I9V/A33V/K72E/V230I, I9V/A33V/V97I/V230I, I9V/I68L/K72E, I9V/V97I/K175P/V202I/Q218M, I9V/K175P, I9V/V202I, T11S, C13G, H14Y/A33V/I68L/K72E/V97I, L18C/A33V/I68L/V97I/F158Y/K175P/I239F, L18C/A33V/K72E/V97I/F158Y/K175P, L18C/A33V/K72E/V97I/F158Y/I239F, L18C/A33V/K72E/V97I/K175P/V202I, L18C/A33V/K72E/K175P/V202I/I237P/I239F, L18C/A33V/K72E/I197V/I237P, L18C/I68L/K72E, L18C/I68L/K72E/F158Y/K175P/I237P/I239F, L18C/I68L/K72E/K175P/V202I/I239F, L18C/I68L/K72E/I239F, L18C/I68L/K175P/K223P/I237P/I239F, L18C/I68L/I239F, L18C/K72E/V97I/K175P/I237P/I239F, L18C/K72E/K175P/I197V, L18C/K72E/I237P, L18C/K72E/I239F, L18C/V97I/K175P/I237P/I239F, L18C/V97I/K223P/I237P/I239F, L18C/K175P, L18C/I197V/I239F, L18C/I239F, L18I, Y20H/V124C, A33V/I68L/K72E, A33V/I68L/K72E/K175P/V230I, A33V/I68L/K223P/I239F, A33V/K72E, A33V/K72E/V97I/I197V, A33V/K72E/K175P, A33V/K175P, V39A, V39G, D44N/V97I, G52S, I68L/K72E, I68L/K72E/V97I, I68L/K72E/V97I/I197V, I68L/K72E/I197V, I68L/K72E/I197V/I239F, I68L/V97I/K175P, I68L/V97I/K175P/I197V/K223P, K72E/V97I/F158Y/K223P/I237P, K72E/K175P/I197V, V97I/F158Y/K175P/I197V, V97I/K175P, H103M, H103S, F121Y, S122C, S122Q, V124C, V124E, V124T, V124Y, F125R, W128H, W128Y, F133A, F133C, L148M, K175P, I197V, and I197V/V202I.

[0138] In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 2 and one or more residue differences as compared to SEQ ID NO: 2, selected from: 7/11/33/72/97, 7/11/33/97, 7/11/72, 7/11/158,

7/11/237, 7/33/72/97/158/175/197/237/239, 9/11/33, 9/11/33/72, 9/11/72, 9/33/68/97/237/239, 9/72/97/158/237/239, 9/218/239, 10, 11, 12, 13, 14, 18, 33/68/72/97/237/239, 33/72/158/239, 39, 40, 41, 48, 49, 51, 54, 54/126, 57, 68/175/239, 82, 84, 85, 86, 103, 104, 105, 107, 115, 117, 119, 119/120, 120, 122, 123, 124, 127, 130, 131, 133, 134, 137/145, 144, 145, 146, 147, 148, 149, 150, 152, 157, 158, 159, 169, 173, 174, 175, 180, 181, 183, 203, 204, 208, 209, 210, 215, 218, 236, and 237. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 2 and one or more residue differences as compared to SEQ ID NO: 2, selected from: 7V/11G/33V/72E/97I, 7V/11G/33V/97I, 7V/11G/72E, 7V/11G/158Y, 7V/11G/237P, 7V/33V/72E/97I/158Y/175P/197V/237P/239F, 9V/11G/33V, 9V/11G/33V/72E, 9V/11G/72E, 9V/33V/68L/97I/237P/239F, 9V/72E/97I/158Y/237P/239F, 9V/218M/239F, 10G, 10P, 10R, 11R, 11S, 12K, 13L, 14R, 18S, 18T, 33V/68L/72E/97I/237P/239F, 33V/72E/158Y/239F, 39R, 39W, 40L, 40S, 40T, 40V, 41C, 48W, 49Q, 49R, 51A, 51Q, 51S, 51T, 54N/126I, 54Y, 57N, 68L/175P/239F, 82T, 84L, 85A, 85L, 85M, 85Q, 86W, 103P, 104A, 104G, 104S, 105D, 107G, 107W, 115C, 115K, 115L, 115V, 117C, 117L, 119L, 119N/120F, 120C, 120L, 120Y, 122G, 122T, 123A, 123P, 123T, 124P, 127A, 127L, 127M, 130M, 130V, 131L, 133W, 134D, 134F, 134L, 134R, 134V, 134W, 137I/145V, 144L, 145F, 145L, 145P, 146D, 146G, 146N, 147R, 148P, 148V, 149W, 150I, 150L, 150V, 152P, 157V, 158I, 158L, 158M, 158R, 158W, 158Y, 159G, 159H, 159L, 169G, 169I, 169S, 173E, 173Y, 174C, 174G, 175D, 180G, 180V, 181C, 183K, 203Q, 204E, 204R, 204V, 208A, 208C, 208E, 208G, 208L, 208T, 208V, 209C, 210V, 215L, 218T, 236G, 236L, 237L, 237S, and 237W. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 2 and one or more residue differences as compared to SEQ ID NO: 2, selected from: I7V/T11G/A33V/K72E/V97I, I7V/T11G/A33V/V97I, I7V/T11G/K72E, I7V/T11G/F158Y, I7V/T11G/I237P, I7V/A33V/K72E/V97I/F158Y/K175P/I197V/I237P/I239F, I9V/T11G/A33V, I9V/T11G/A33V/K72E, I9V/T11G/K72E, I9V/A33V/I68L/V97I/I237P/I239F, I9V/K72E/V97I/F158Y/I237P/I239F, I9V/Q218M/I239F, H10G, H10P, H10R, T11R, T11S, I12K, C13L, H14R, L18S, L18T, A33V/I68L/K72E/V97I/I237P/I239F, A33V/K72E/F158Y/I239F, V39R, V39W, A40L, A40S, A40T, A40V, S41C, L48W, E49Q, E49R, I51A, I51Q, I51S, I51T, W54N/T126I, W54Y, Y57N, I68L/K175P/I239F, G82T, I84L, N85A, N85L, N85M, N85Q, I86W, H103P, N104A, N104G, N104S, A105D, M107G, M107W, A115C, A115K, A115L, A115V, V117C, V117L, K119L, K119N/K120F, K120C, K120L, K120Y, S122G, S122T, E123A, E123P, E123T, V124P, D127A, D127L, D127M, D130M, D130V, S131L, F133W, S134D, S134F, S134L, S134R, S134V, S134W, T137I/A145V, T144L, A145F, A145L, A145P, V146D, V146G, V146N, E147R, L148P, L148V, G149W, D150I, D150L, D150V, T152P, I157V, F158I, F158L, F158M, F158R, F158W, F158Y, S159G, S159H, S159L, A169G, A169I, A169S, V173E, V173Y, R174C, R174G, K175D, E180G, E180V,

Q181C, L183K, Y203Q, G204E, G204R, G204V, Q208A, Q208C, Q208E, Q208G, Q208L, Q208T, Q208V, I209C, F210V, Q215L, Q218T, K236G, K236L, I237L, I237S, and I237W.

[0139] In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 24 and one or more residue differences as compared to SEQ ID NO: 24, selected from: 9/39/122/157, 18, 18/33/39/230, 18/33/175, 18/39/44/118, 18/175, 20, 33/39, 33/39/44/118/121, 33/39/44/121, 33/39/44/152, 33/39/44/175, 33/39/44/197, 33/39/44/210/230, 33/39/118/121/175, 33/39/121/175, 33/39/197/210, 33/39/210, 33/39/210/230, 33/118/121, 33/118/121/175, 39, 39/44/103, 39/44/118, 39/44/152/230, 39/44/175/197/230, 39/103, 39/103/125, 39/103/125/127/146, 39/103/125/127/146/150, 39/103/125/146, 39/103/127/150, 39/103/150, 39/118/121, 39/197, 39/197/210, 43, 44/103, 81, 103, 103/125/127, 103/125/146, 103/127, 103/150, 118/121, 118/121/175, 118/121/197, 121, 152, 152/197, 154, 159, 163, 175, 210, 232, and 238. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 24 and one or more residue differences as compared to SEQ ID NO: 24, selected from: 9V/39W/122T/157V, 18I, 18I/33V/39W/230I, 18I/33V/175P, 18I/39W/44N/118L, 18I/175P, 20H, 33V/39W, 33V/39W/44N/118L/121Y, 33V/39W/44N/121Y, 33V/39W/44N/152F, 33V/39W/44N/175P, 33V/39W/44N/197V, 33V/39W/44N/210M/230I, 33V/39W/118L/121Y/175P, 33V/39W/121Y/175P, 33V/39W/197V/210M, 33V/39W/210M, 33V/39W/210M/230I, 33V/118L/121W/175P, 33V/118L/121Y, 39W, 39W/44N/118L, 39W/44N/152F/230I, 39W/44N/175P/197V/230I, 39W/44Y/103S, 39W/103M, 39W/103M/125R/146L, 39W/103M/150P, 39W/103Q, 39W/103S, 39W/103S/125P/127G/146L/150P, 39W/103S/125R, 39W/103S/125R/127G/146L, 39W/103S/127G/150P, 39W/118L/121Y, 39W/197V, 39W/197V/210M, 43M, 43Q, 43R, 43S, 43T, 44N/103S, 81F, 103L, 103M, 103M/125P/146L, 103M/127G, 103Q, 103S, 103S/125P/127G, 103S/150P, 103V, 118L/121Y, 118L/121Y/175P, 118L/121Y/197V, 121Y, 152A/197V, 152F, 154I, 159Q, 163P, 175P, 210M, 232G, and 238M. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 24 and one or more residue differences as compared to SEQ ID NO: 24, selected from: I9V/V39W/S122T/I157V, L18I, L18I/A33V/V39W/V230I, L18I/A33V/K175P, L18I/V39W/D44N/Y118L, L18I/K175P, Y20H, A33V/V39W, A33V/V39W/D44N/Y118L/F121Y, A33V/V39W/D44N/F121Y, A33V/V39W/D44N/T152F, A33V/V39W/D44N/K175P, A33V/V39W/D44N/I197V, A33V/V39W/D44N/F210M/V230I, A33V/V39W/Y118L/F121Y/K175P, A33V/V39W/F121Y/K175P, A33V/V39W/I197V/F210M, A33V/V39W/F210M, A33V/V39W/F210M/V230I, A33V/Y118L/F121W/K175P, A33V/Y118L/F121Y, V39W, V39W/D44N/Y118L, V39W/D44N/T152F/V230I, V39W/D44N/K175P/I197V/V230I, V39W/D44Y/H103S, V39W/H103M, V39W/H103M/F125R/V146L, V39W/H103M/D150P,

V39W/H103Q, V39W/H103S, V39W/H103S/F125P/D127G/V146L/D150P, V39W/H103S/F125R, V39W/H103S/F125R/D127G/V146L, V39W/H103S/D127G/D150P, V39W/Y118L/F121Y, V39W/I197V, V39W/I197V/F210M, I43M, I43Q, I43R, I43S, I43T, D44N/H103S, G81F, H103L, H103M, H103M/F125P/V146L, H103M/D127G, H103Q, H103S, H103S/F125P/D127G, H103S/D150P, H103V, Y118L/F121Y, Y118L/F121Y/K175P, Y118L/F121Y/I197V, F121Y, T152A/I197V, T152F, A154I, S159Q, I163P, K175P, F210M, S232G, and Q238M.

[0140] In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 24 and one or more residue differences as compared to SEQ ID NO: 24, selected from: 7, 9/13, 9/13/39, 9/13/39/68/72/239, 9/13/39/68/239, 9/13/39/72/122, 9/13/39/97/239, 9/13/68/72/97/239, 9/13/68/97/134/239, 9/13/68/97/239, 9/13/72, 9/13/134, 11, 12/42/125/127, 12/103, 12/103/150, 13, 13/39, 13/39/68, 13/39/68/72/97, 13/39/72/97/122/134/223, 13/39/97/239, 13/39/239, 13/68, 13/68/72/97/122/239, 16, 18/33/39/44/118/121/175/197/210, 18/33/39/44/121, 18/33/39/112/118/121/197, 18/33/39/121, 18/33/39/121/175, 18/33/39/121/175/197, 18/33/39/121/175/210, 18/33/39/121/210, 18/33/118/121, 18/33/121, 18/33/121/197, 18/33/121/210/230, 18/39/44/118/121/230, 18/39/44/121/175/210, 18/39/118/121, 18/39/118/121/210, 18/39/121/175, 18/39/121/175/210, 18/39/121/197/210, 18/118/121, 18/118/121/175, 18/118/121/210, 18/118/121/230, 18/121, 18/121/175, 18/121/210, 18/121/230, 20, 21, 33/39/44/118/121, 33/39/44/118/121/210/230, 33/39/44/121, 33/39/44/121/152, 33/39/44/121/175/210, 33/39/44/121/230, 33/39/73/121/197, 33/39/118/121/175, 33/39/118/121/175/230, 33/39/121, 33/39/121/175, 33/39/121/210/230, 33/118/121, 33/118/121/175, 33/118/121/197, 33/121, 33/121/230, 37/91, 39/44/118/121/175/210, 39/68/72/239, 39/118/121, 39/118/121/210, 39/121, 39/121/210/230, 43, 44, 76, 79, 101, 118/121, 118/121/175, 118/121/197, 121, 121/175/230, 121/210, 148, 153, 159, 161, 172, 200, 207, 233, 245, and 248. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 24 and one or more residue differences as compared to SEQ ID NO: 24, selected from: 7E, 9V/13G, 9V/13G/39W, 9V/13G/39W/68L/72E/239F, 9V/13G/39W/68L/239F, 9V/13G/39W/72E/122T, 9V/13G/39W/97I/239F, 9V/13G/68L/72E/97I/239F, 9V/13G/68L/97I/134D/239F, 9V/13G/68L/97I/239F, 9V/13G/72E, 9V/13G/134D, 11N, 12A/42W/125R/127G, 12A/103S, 12A/103S/150P, 13F, 13G/39W, 13G/39W/68L, 13G/39W/68L/72E/97I, 13G/39W/72E/97I/122T/134D/223P, 13G/39W/97I/239F, 13G/39W/239F, 13G/68L, 13G/68L/72E/97I/122T/239F, 16M, 16V, 18C/33V/39W/112R/118L/121W/197V, 18C/33V/39W/121W/175P/210M, 18C/33V/39W/121W/210M, 18C/33V/39W/121Y/175P, 18C/33V/118L/121W, 18C/33V/118L/121Y, 18C/33V/121Y, 18C/33V/121Y/197V, 18C/33V/121Y/210M/230I, 18C/39W/44N/121Y/175P/210M, 18C/39W/118L/121W, 18C/39W/121W/175P/210M, 18C/39W/121W/197V/210M, 18C/118L/121W, 18C/118L/121Y, 18C/121V, 18C/121V/175P, 18C/121W, 18C/121W/175P, 18C/121W/210M, 18C/121Y, 18C/121Y/230I,

18I/33V/39W/44N/118L/121W/175P/197V/210M, 18I/33V/39W/44N/121T, 18I/33V/39W/121T/175P, 18I/33V/39W/121V/175P/197V, 18I/33V/39W/121W, 18I/33V/118L/121W, 18I/39W/44N/118L/121W/230I, 18I/39W/44N/121V/175P/210M, 18I/39W/118L/121W/210M, 18I/39W/121V/175P, 18I/118L/121T, 18I/118L/121W, 18I/118L/121W/175P, 18I/118L/121W/210M, 18I/118L/121Y, 18I/118L/121Y/230I, 18I/121T/175P, 18I/121W, 18I/121Y/210M, 20G, 20K, 20R, 20T, 21V, 33V/39G/44N/121W/152A, 33V/39W/44N/118L/121W/210M/230I, 33V/39W/44N/118L/121Y, 33V/39W/44N/121V, 33V/39W/44N/121Y, 33V/39W/44N/121Y/175P/210M, 33V/39W/44N/121Y/230I, 33V/39W/73N/121Y/197V, 33V/39W/118L/121W/175P, 33V/39W/118L/121Y/175P, 33V/39W/118L/121Y/175P/230I, 33V/39W/121V, 33V/39W/121W, 33V/39W/121W/210M/230I, 33V/39W/121Y/175P, 33V/118L/121W/175P, 33V/118L/121W/197V, 33V/118L/121Y, 33V/121W, 33V/121Y/230I, 37L/91G, 39G/68L/72E/239F, 39W/44N/118L/121W/175P/210M, 39W/118L/121W/210M, 39W/118L/121Y, 39W/118L/121Y/210M, 39W/121W, 39W/121W/210M/230I, 43E, 43Q, 43S, 43W, 44V, 76P, 79F, 101L, 118L/121W, 118L/121Y, 118L/121Y/175P, 118L/121Y/197V, 121V, 121W, 121Y, 121Y/175P/230I, 121Y/210M, 148F, 153W, 159Q, 161M, 172A, 172E, 172G, 172Q, 172R, 172V, 200G, 207N, 233R, 245D, 245G, 248G, and 248S. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 24 and one or more residue differences as compared to SEQ ID NO: 24, selected from: I7E, I9V/C13G, I9V/C13G/V39W, I9V/C13G/V39W/I68L/K72E/I239F, I9V/C13G/V39W/I68L/I239F, I9V/C13G/V39W/K72E/S122T, I9V/C13G/V39W/V97I/I239F, I9V/C13G/I68L/K72E/V97I/I239F, I9V/C13G/I68L/V97I/S134D/I239F, I9V/C13G/I68L/V97I/I239F, I9V/C13G/K72E, I9V/C13G/S134D, S11N, I12A/G42W/F125R/D127G, I12A/H103S, I12A/H103S/D150P, C13F, C13G/V39W, C13G/V39W/I68L, C13G/V39W/I68L/K72E/V97I, C13G/V39W/K72E/V97I/S122T/S134D/K223P, C13G/V39W/V97I/I239F, C13G/V39W/I239F, C13G/I68L, C13G/I68L/K72E/V97I/S122T/I239F, A16M, A16V, L18C/A33V/V39W/H112R/Y118L/F121W/I197V, L18C/A33V/V39W/F121W/K175P/F210M, L18C/A33V/V39W/F121W/F210M, L18C/A33V/V39W/F121Y/K175P, L18C/A33V/Y118L/F121W, L18C/A33V/Y118L/F121Y, L18C/A33V/F121Y, L18C/A33V/F121Y/I197V, L18C/A33V/F121Y/F210M/V230I, L18C/V39W/D44N/F121Y/K175P/F210M, L18C/V39W/Y118L/F121W, L18C/V39W/F121W/K175P/F210M, L18C/V39W/F121W/I197V/F210M, L18C/Y118L/F121W, L18C/Y118L/F121Y, L18C/F121V, L18C/F121V/K175P, L18C/F121W, L18C/F121W/K175P, L18C/F121W/F210M, L18C/F121Y, L18C/F121Y/V230I, L18I/A33V/V39W/D44N/Y118L/F121W/K175P/I197V/F210M, L18I/A33V/V39W/D44N/F121T, L18I/A33V/V39W/F121T/K175P, L18I/A33V/V39W/F121V/K175P/I197V, L18I/A33V/V39W/F121W, L18I/A33V/Y118L/F121W, L18I/V39W/D44N/Y118L/F121W/V230I, L18I/V39W/D44N/F121V/K175P/F210M, L18I/V39W/Y118L/F121W/F210M,

L18I/V39W/F121V/K175P, L18I/Y118L/F121T, L18I/Y118L/F121W, L18I/Y118L/F121W/K175P, L18I/Y118L/F121W/F210M, L18I/Y118L/F121Y, L18I/Y118L/F121Y/V230I, L18I/F121T/K175P, L18I/F121W, L18I/F121Y/F210M, Y20G, Y20K, Y20R, Y20T, K21V, A33V/V39G/D44N/F121W/T152A, A33V/V39W/D44N/Y118L/F121W/F210M/V230I, A33V/V39W/D44N/Y118L/F121Y, A33V/V39W/D44N/F121V, A33V/V39W/D44N/F121Y, A33V/V39W/D44N/F121Y/K175P/F210M, A33V/V39W/D44N/F121Y/V230I, A33V/V39W/K73N/F121Y/I197V, A33V/V39W/Y118L/F121W/K175P, A33V/V39W/Y118L/F121Y/K175P, A33V/V39W/Y118L/F121Y/K175P/V230I, A33V/V39W/F121V, A33V/V39W/F121W, A33V/V39W/F121W/F210M/V230I, A33V/V39W/F121Y/K175P, A33V/Y118L/F121W/K175P, A33V/Y118L/F121W/I197V, A33V/Y118L/F121Y, A33V/F121W, A33V/F121Y/V230I, D37L/E91G, V39G/I68L/K72E/I239F, V39W/D44N/Y118L/F121W/K175P/F210M, V39W/Y118L/F121W/F210M, V39W/Y118L/F121Y, V39W/Y118L/F121Y/F210M, V39W/F121W, V39W/F121W/F210M/V230I, I43E, I43Q, I43S, I43W, D44V, L76P, E79F, V101L, Y118L/F121W, Y118L/F121Y, Y118L/F121Y/K175P, Y118L/F121Y/I197V, F121V, F121W, F121Y, F121Y/K175P/V230I, F121Y/F210M, L148F, L153W, S159Q, S161M, L172A, L172E, L172G, L172Q, L172R, L172V, V200G, D207N, A233R, L245D, L245G, I248G, and I248S.

[0141] In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 496 and one or more residue differences as compared to SEQ ID NO: 496, selected from: 43/103, 43/103/172/238, 43/103/238/241, 43/238, 81/232, 103, 103/159/238, 103/194/238, 103/238, 122, 232, 238, and 238/241. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 496 and one or more residue differences as compared to SEQ ID NO: 496, selected from: 43E/103S/238M/241R, 43M/103L, 43Q/103S/238M/241R, 43Q/238M, 43S/103S/172R/238M, 43S/238M, 81F/232G, 103L, 103S, 103S/159Q/238M, 103S/194L/238M, 103S/238M, 103V, 122C, 122I, 122K, 122L, 122M, 122Q, 122R, 122T, 122V, 232G, 238M, and 238M/241R. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 496 and one or more residue differences as compared to SEQ ID NO: 496, selected from: I43E/H103S/Q238M/K241R, I43M/H103L, I43Q/H103S/Q238M/K241R, I43Q/Q238M, I43S/H103S/L172R/Q238M, I43S/Q238M, G81F/S232G, H103L, H103S, H103S/S159Q/Q238M, H103S/Y194L/Q238M, H103S/Q238M, H103V, S122C, S122I, S122K, S122L, S122M, S122Q, S122R, S122T, S122V, S232G, Q238M, and Q238M/K241R.

[0142] In some embodiments, the engineered nitroaldolase 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

sequence identity to reference sequence SEQ ID NO: 496 and one or more residue differences as compared to SEQ ID NO: 496, selected from: 12, 43, 43/47/103, 43/47/103/232, 43/103, 43/103/121/172/241, 43/103/172, 43/103/172/238, 43/103/238, 43/103/238/241, 43/121, 43/194/238, 43/238, 44/103/121/238, 44/103/238, 44/103/238/241, 44/121/238/241, 47/103, 63, 103, 103/121, 103/121/172/238, 103/159/238, 103/172/238, 103/194/238, 103/238, 122, 131, 146, 178, 180, and 221. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 496 and one or more residue differences as compared to SEQ ID NO: 496, selected from: 12V, 43E/103S/121W/172R/241R, 43E/103S/238M, 43E/103S/238M/241R, 43M/47G/103L, 43M/47G/103V, 43M/103L, 43M/103S, 43M/103V, 43Q/47G/103S/232G, 43Q/103L/172R, 43Q/103S, 43Q/103S/238M/241R, 43Q/121W, 43Q/238M, 43S, 43S/103L, 43S/103S, 43S/103S/172R, 43S/103S/172R/238M, 43S/103S/238M/241R, 43S/194L/238M, 43S/238M, 44V/103S/121W/238M, 44V/103S/238M, 44V/103S/238M/241R, 44V/121W/238M/241R, 47G/103L, 63P, 103L, 103S, 103S/121W, 103S/121W/172R/238M, 103S/159Q/238M, 103S/172R/238M, 103S/194L/238M, 103S/238M, 103V, 122H, 122I, 122K, 122L, 122M, 122N, 122R, 122T, 122V, 131Y, 146I, 178A, 180L, 180V, and 221G. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 496 and one or more residue differences as compared to SEQ ID NO: 496, selected from: I12V, I43E/H103S/Y121W/L172R/K241R, I43E/H103S/Q238M, I43E/H103S/Q238M/K241R, I43M/Q47G/H103L, I43M/Q47G/H103V, I43M/H103L, I43M/H103S, I43M/H103V, I43Q/Q47G/H103S/S232G, I43Q/H103L/L172R, I43Q/H103S, I43Q/H103S/Q238M/K241R, I43Q/Y121W, I43Q/Q238M, I43S, I43S/H103L, I43S/H103S, I43S/H103S/L172R, I43S/H103S/L172R/Q238M, I43S/H103S/Q238M/K241R, I43S/Y194L/Q238M, I43S/Q238M, N44V/H103S/Y121W/Q238M, N44V/H103S/Q238M, N44V/H103S/Q238M/K241R, N44V/Y121W/Q238M/K241R, Q47G/H103L, T63P, H103L, H103S, H103S/Y121W, H103S/Y121W/L172R/Q238M, H103S/S159Q/Q238M, H103S/L172R/Q238M, H103S/Y194L/Q238M, H103S/Q238M, H103V, S122H, S122I, S122K, S122L, S122M, S122N, S122R, S122T, S122V, S131Y, V146I, F178A, E180L, E180V, and N221G.

[0143] In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 860 and one or more residue differences as compared to SEQ ID NO: 860, selected from: 12, 12/43/63/103/146/180/232/241, 12/43/63/103/146/241, 12/43/81/103/241, 12/43/81/146/232/241, 12/43/81/180/241, 12/43/81/232/241, 12/43/81/241, 12/43/103, 12/43/103/146/180/232/241, 12/43/103/146/180/241, 12/43/103/180/241, 12/43/103/241, 12/43/146/180/232/241, 12/43/180/241, 12/63/81/103/241, 12/81, 12/81/180/241, 12/103/146/241, 12/103/180/241, 12/146, 12/146/180/241, 12/180/232/241, 12/241, 18, 43/81/232/241, 43/103/146/180,

43/103/180/241, 43/103/241, 43/180/232, 43/232/241, 63/103/180/232, 78, 81, 81/103/146, 81/146/180/241, 81/146/241, 81/180/241, 84, 103, 103/146, 103/146/180, 103/146/180/241, 103/180, 103/180/241, 103/261, 144, 145, 146/180/232, 146/180/241, 146/241, 150, 166, 172, 180/232, 232/241, and 241. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 860 and one or more residue differences as compared to SEQ ID NO: 860, selected from: 12V, 12V/43M/63P/103L/146I/180L/232G/241R, 12V/43M/81F/103L/241R, 12V/43M/81F/146I/232G/241R, 12V/43M/81F/180V/241R, 12V/43M/81F/232G/241R, 12V/43M/81F/241R, 12V/43M/103L/180L/241R, 12V/43M/103L/241R, 12V/43M/103V/146I/180V/241R, 12V/43M/103V/241R, 12V/43M/146I/180L/232G/241R, 12V/43M/180L/241R, 12V/43M/180V/241R, 12V/43Q/63P/103L/146I/241R, 12V/43Q/103L/146I/180V/232G/241R, 12V/43Q/103L/146I/180V/241R, 12V/43Q/103L/180V/241R, 12V/43Q/103V, 12V/63P/81F/103L/241R, 12V/81F, 12V/81F/180V/241R, 12V/103L/146I/241R, 12V/103L/180L/241R, 12V/103V/180V/241R, 12V/146I, 12V/146I/180L/241R, 12V/180L/232G/241R, 12V/180V/232G/241R, 12V/241R, 18V, 43M/81F/232G/241R, 43M/103L/146I/180V, 43M/103L/180V/241R, 43M/103L/241R, 43M/103V/146I/180V, 43M/180L/232G, 43M/232G/241R, 43Q/103V/146I/180L, 63P/103L/180L/232G, 78A, 81F/103L/146I, 81F/146I/180L/241R, 81F/146I/241R, 81F/180L/241R, 81Q, 84Y, 103A, 103L, 103L/146I, 103L/146I/180V, 103L/180L, 103L/180V/241R, 103N, 103T, 103T/261V, 103V/146I/180L/241R, 144V, 145V, 146I/180L/241R, 146I/180V/232G, 146I/241R, 150H, 166L, 172L, 172V, 180L/232G, 232G/241R, and 241R. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 860 and one or more residue differences as compared to SEQ ID NO: 860, selected from: I12V, I12V/S43M/T63P/S103L/V146I/E180L/S232G/K241R, I12V/S43M/G81F/S103L/K241R, I12V/S43M/G81F/V146I/S232G/K241R, I12V/S43M/G81F/E180V/K241R, I12V/S43M/G81F/S232G/K241R, I12V/S43M/G81F/K241R, I12V/S43M/S103L/E180L/K241R, I12V/S43M/S103L/K241R, I12V/S43M/S103V/V146I/E180V/K241R, I12V/S43M/S103V/K241R, I12V/S43M/V146I/E180L/S232G/K241R, I12V/S43M/E180L/K241R, I12V/S43M/E180V/K241R, I12V/S43Q/T63P/S103L/V146I/K241R, I12V/S43Q/S103L/V146I/E180V/S232G/K241R, I12V/S43Q/S103L/V146I/E180V/K241R, I12V/S43Q/S103L/E180V/K241R, I12V/S43Q/S103V, I12V/T63P/G81F/S103L/K241R, I12V/G81F, I12V/G81F/E180V/K241R, I12V/S103L/V146I/K241R, I12V/S103L/E180L/K241R, I12V/S103V/E180V/K241R, I12V/V146I, I12V/V146I/E180L/K241R, I12V/E180L/S232G/K241R, I12V/E180V/S232G/K241R, I12V/K241R, L18V, S43M/G81F/S232G/K241R, S43M/S103L/V146I/E180V, S43M/S103L/E180V/K241R, S43M/S103L/K241R, S43M/S103V/V146I/E180V, S43M/E180L/S232G, S43M/S232G/K241R, S43Q/S103V/V146I/E180L, T63P/S103L/E180L/S232G, G78A, G81F/S103L/V146I,

G81F/V146I/E180L/K241R, G81F/V146I/K241R, G81F/E180L/K241R, G81Q, I84Y, S103A, S103L, S103L/V146I, S103L/V146I/E180V, S103L/E180L, S103L/E180V/K241R, S103N, S103T, S103T/A261V, S103V/V146I/E180L/K241R, T144V, A145V, V146I/E180L/K241R, V146I/E180V/S232G, V146I/K241R, D150H, V166L, R172L, R172V, E180L/S232G, S232G/K241R, and K241R.

[0144] In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 860 and one or more residue differences as compared to SEQ ID NO: 860, selected from: 9, 12, 12/43/63/103/146/180/232/241, 12/43/63/103/146/241, 12/43/103/146/180/232/241, 12/43/103/146/180/241, 12/43/103/146/232/241, 12/43/103/146/241, 12/43/103/180/241, 12/43/146/180/232/241, 12/43/146/241, 12/43/180/232, 12/43/180/241, 12/63/103/146/232/241, 12/103/146/241, 12/103/180/241, 12/146, 12/146/180/241, 12/146/232/241, 12/146/241, 12/180, 12/180/232/241, 12/180/241, 12/241, 18, 43/63/103/180/232/241, 43/103/146/180, 43/103/180, 43/103/180/232/241, 43/103/180/241, 43/180/232, 63/103/180/232, 103, 103/146/180, 103/146/180/241, 103/180, 103/180/241, 103/241, 118, 119, 122, 144, 145, 146/180/241, 154, 160, 172, 180/232, 180/232/241, 180/241, 208, 211, and 212. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 860 and one or more residue differences as compared to SEQ ID NO: 860, selected from: 9V, 12V, 12V/43M/63P/103L/146I/180L/232G/241R, 12V/43M/103L/146I/241R, 12V/43M/103L/180L/241R, 12V/43M/103V/146I/180V/241R, 12V/43M/103V/146I/232G/241R, 12V/43M/146I/180L/232G/241R, 12V/43M/180L/232G, 12V/43M/180L/241R, 12V/43Q/63P/103V/146I/241R, 12V/43Q/103L/146I/180V/232G/241R, 12V/43Q/103L/146I/180V/241R, 12V/43Q/103L/146I/241R, 12V/43Q/103L/180V/241R, 12V/43Q/146I/241R, 12V/63P/103L/146I/232G/241R, 12V/103L/180L/241R, 12V/103L/180V/241R, 12V/103V/146I/241R, 12V/103V/180L/241R, 12V/103V/180V/241R, 12V/146I, 12V/146I/180L/241R, 12V/146I/232G/241R, 12V/146I/241R, 12V/180L, 12V/180L/232G/241R, 12V/180L/241R, 12V/180V, 12V/180V/232G/241R, 12V/180V/241R, 12V/241R, 18V, 43M/63P/103L/180V/232G/241R, 43M/103L/146I/180V, 43M/103L/180V, 43M/103L/180V/241R, 43M/103V/146I/180V, 43M/180L/232G, 43Q/103L/180V/232G/241R, 43Q/103V/146I/180L, 43Q/103V/180L/241R, 43Q/103V/180V/232G/241R, 63P/103L/180L/232G, 103A, 103L, 103L/146I/180V, 103L/146I/180V/241R, 103L/180L, 103L/180V/241R, 103L/241R, 103N, 103T, 103V/146I/180L/241R, 103V/180V/241R, 118L, 118V, 119H, 122H, 144H, 144V, 145V, 146I/180L/241R, 154R, 160L, 172C, 172L, 172V, 180L/232G, 180L/241R, 180V/232G, 180V/232G/241R, 208F, 211F, and 212P. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 860 and one or more residue differences as

compared to SEQ ID NO: 860, selected from: I9V, I12V, I12V/S43M/T63P/S103L/V146I/E180L/S232G/K241R, I12V/S43M/S103L/V146I/K241R, I12V/S43M/S103L/E180L/K241R, I12V/S43M/S103V/V146I/E180V/K241R, I12V/S43M/S103V/V146I/S232G/K241R, I12V/S43M/V146I/E180L/S232G/K241R, I12V/S43M/E180L/S232G, I12V/S43M/E180L/K241R, I12V/S43Q/T63P/S103V/V146I/K241R, I12V/S43Q/S103L/V146I/E180V/S232G/K241R, I12V/S43Q/S103L/V146I/E180V/K241R, I12V/S43Q/S103L/V146I/K241R, I12V/S43Q/S103L/E180V/K241R, I12V/S43Q/V146I/K241R, I12V/T63P/S103L/V146I/S232G/K241R, I12V/S103L/E180L/K241R, I12V/S103L/E180V/K241R, I12V/S103V/V146I/K241R, I12V/S103V/E180L/K241R, I12V/S103V/E180V/K241R, I12V/V146I, I12V/V146I/E180L/K241R, I12V/V146I/S232G/K241R, I12V/V146I/K241R, I12V/E180L, I12V/E180L/S232G/K241R, I12V/E180L/K241R, I12V/E180V, I12V/E180V/S232G/K241R, I12V/E180V/K241R, I12V/K241R, L18V, S43M/T63P/S103L/E180V/S232G/K241R, S43M/S103L/V146I/E180V, S43M/S103L/E180V, S43M/S103L/E180V/K241R, S43M/S103V/V146I/E180V, S43M/E180L/S232G, S43Q/S103L/E180V/S232G/K241R, S43Q/S103V/V146I/E180L, S43Q/S103V/E180L/K241R, S43Q/S103V/E180V/S232G/K241R, T63P/S103L/E180L/S232G, S103A, S103L, S103L/V146I/E180V, S103L/V146I/E180V/K241R, S103L/E180L, S103L/E180V/K241R, S103L/K241R, S103N, S103T, S103V/V146I/E180L/K241R, S103V/E180V/K241R, Y118L, Y118V, K119H, S122H, T144H, T144V, A145V, V146I/E180L/K241R, A154R, N160L, R172C, R172L, R172V, E180L/S232G, E180L/K241R, E180V/S232G, E180V/S232G/K241R, Q208F, S211F, and R212P.

[0145] In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 1120 and one or more residue differences as compared to SEQ ID NO: 1120, selected from: 3, 7, 7/158, 7/158/175, 7/158/197/202, 27/43/115/172, 27/43/172, 29, 32, 43/110/115/172, 44, 46, 48/172, 55, 59, 67, 71, 72, 73, 94/103/118/146, 103, 103/144/146/172/180, 103/146/154/172/180/232, 103/154, 118, 118/144/215, 118/172/180, 137, 142, 146, 146/154, 154, 158, 172/190, 184, 198, 205, 216, 223, 226, 229, 247, 261, and 263. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 1120 and one or more residue differences as compared to SEQ ID NO: 1120, selected from: 3D, 7V, 7V/158Y, 7V/158Y/175P, 7V/158Y/197V/202I, 27E/43I/115S/172L, 27E/43I/172L, 29G, 29R, 32G, 32R, 43I/110T/115S/172L, 44S, 46H, 46V, 48I/172L, 55A, 55G, 55R, 55S, 59G, 59T, 67N, 71S, 72G, 73R, 94Q/103L/118V/146I, 103L, 103L/144V/146I/172L/180V, 103L/146I/154R/172L/180V/232G, 103L/154R, 118V, 118V/144V/215H, 118V/172L/180V, 137M, 137R, 142V, 146I, 146I/154R, 154R, 158Y, 172L/190S, 184S, 198L, 205G, 205L, 216M, 216V, 223A, 223C, 223E, 223L, 223M, 223S, 226L, 226V, 229L, 229V, 229W, 247V, 261G, 261H, 263E, 263G, 263L, and 263S. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 1120 and one or more residue differences as compared to SEQ ID NO: 1120, selected from: S3D, I7V, I7V/F158Y, I7V/F158Y/K175P, I7V/F158Y/I197V/V202I, Q27E/S43I/A115S/R172L, Q27E/S43I/R172L, A29G, A29R, N32G, N32R, S43I/I110T/A115S/R172L, N44S, R46H, R46V, L48I/R172L, E55A, E55G, E55R, E55S, E59G, E59T, S67N, G71S, K72G, K73R, P94Q/V103L/Y118V/V146I, V103L, V103L/T144V/V146I/R172L/L180V, V103L/V146I/A154R/R172L/L180V/S232G, V103L/A154R, Y118V, Y118V/T144V/Q215H, Y118V/R172L/L180V, T137M, T137R, T142V, V146I, V146I/A154R, A154R, F158Y, R172L/T190S, D184S, R198L, E205G, E205L, L216M, L216V, K223A, K223C, K223E, K223L, K223M, K223S, K226L, K226V, C229L, C229V, C229W, Q247V, A261G, A261H, A263E, A263G, A263L, and A263S.

[0146] In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 1120 and one or more residue differences as compared to SEQ ID NO: 1120, selected from: 73, 80, 103/146/209, 103/172/209, 118/119/172/209, 135, 144/209/232, 172/180/209, 181, and 223. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 1120 and one or more residue differences as compared to SEQ ID NO: 1120, selected from: 73R, 80A, 80G, 103L/146I/209C, 103L/172C/209C, 118V/119H/172C/209C, 135V, 144V/209C/232G, 172C/180V/209C, 181S, and 223E. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 1120 and one or more residue differences as compared to SEQ ID NO: 1120, selected from: K73R, S80A, S80G, V103L/V146I/I209C, V103L/R172C/I209C, Y118V/K119H/R172C/I209C, N135V, T144V/I209C/S232G, R172C/L180V/I209C, Q181S, and K223E.

[0147] In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 1348 and one or more residue differences as compared to SEQ ID NO: 1348, selected from: 2, 7/46, 39, 46, 46/55, 46/55/72, 46/55/92, 46/55/118, 46/55/226/257/263, 46/55/263, 46/70, 46/70/92/118/261/263, 46/70/175/263, 46/72, 46/118, 55/257/261, 55/261/263, 55/263, 105, 118, 118/226/261/263, and 263. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 1348 and one or more residue differences as compared to SEQ ID NO: 1348, selected from: 2G, 7V/46H, 39F, 39G, 39M, 39Q, 39S, 39T, 39V, 46H, 46H/55R, 46H/55R/72C, 46H/55R/92A, 46H/55R/118V, 46H/55R/226V/257Q/263L, 46H/55R/263L, 46H/70N, 46H/70N/92A/118V/261G/263L,

46H/70N/175P/263G, 46H/72C, 46H/118V, 55R/257Q/261G, 55R/261G/263I, 55R/263G, 105G, 118V, 118V/226V/261G/263I, and 263G. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 1348 and one or more residue differences as compared to SEQ ID NO: 1348, selected from: V2G, I7V/R46H, W39F, W39G, W39M, W39Q, W39S, W39T, W39V, R46H, R46H/E55R, R46H/E55R/K72C, R46H/E55R/K92A, R46H/E55R/Y118V, R46H/E55R/K226V/S257Q/A263L, R46H/E55R/A263L, R46H/E70N, R46H/E70N/K92A/Y118V/A261G/A263L, R46H/E70N/K175P/A263G, R46H/K72C, R46H/Y118V, E55R/S257Q/A261G, E55R/A261G/A263I, E55R/A263G, A105G, Y118V, Y118V/K226V/A261G/A263I, and A263G.

[0148] In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 1348 and one or more residue differences as compared to SEQ ID NO: 1348, selected from: 2, 7/46, 7/46/175, 22/46/55/118, 27/46/72, 27/46/146/158/202, 27/46/175, 27/118, 29/70/118, 39, 46, 46/55, 46/55/59/261/263, 46/55/72, 46/55/92, 46/55/118, 46/55/118/263, 46/55/146, 46/55/146/202, 46/55/226/257/263, 46/55/263, 46/70, 46/70/92/118/261/263, 46/70/175/263, 46/72, 46/92, 46/92/175, 46/118, 46/146, 46/146/257/263, 55/70/118, 55/257/261, 55/257/263, 55/261/263, 55/263, 70, 92/118, 105, 118, 118/226/261/263, 146/257/263, 146/261/263, 192/257/261, 202, 257/261/263, 261/263, and 263. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 1348 and one or more residue differences as compared to SEQ ID NO: 1348, selected from: 2G, 7V/46H, 7V/46H/175P, 22V/46H/55R/118V, 27G/46H/72C, 27G/46H/146I/158Y/202I, 27G/46H/175P, 27G/118V, 29V/70N/118V, 39M, 39Q, 39S, 39T, 39V, 46H, 46H/55R, 46H/55R/59T/261G/263L, 46H/55R/72C, 46H/55R/92A, 46H/55R/118V, 46H/55R/118V/263S, 46H/55R/146I, 46H/55R/146I/202I, 46H/55R/226V/257Q/263L, 46H/55R/263L, 46H/70N, 46H/70N/92A/118V/261G/263L, 46H/70N/175P/263G, 46H/72C, 46H/92A, 46H/92A/175P, 46H/118V, 46H/146I, 46H/146I/257Q/263I, 55R/70N/118V, 55R/257Q/261G, 55R/257Q/263L, 55R/261G/263I, 55R/263G, 70N, 92A/118V, 105G, 118V, 118V/226V/261G/263I, 146I/257Q/263S, 146I/261G/263L, 192V/257Q/261G, 202I, 257Q/261G/263G, 261G/263I, and 263G. In some embodiments, the engineered nitroaldolase 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 sequence identity to reference sequence SEQ ID NO: 1348 and one or more residue differences as compared to SEQ ID NO: 1348, selected from: V2G, I7V/R46H, I7V/R46H/K175P, L22V/R46H/E55R/Y118V, Q27G/R46H/K72C, Q27G/R46H/V146I/F158Y/V202I, Q27G/R46H/K175P, Q27G/Y118V, A29V/E70N/Y118V, W39M, W39Q, W39S, W39T, W39V, R46H, R46H/E55R, R46H/E55R/E59T/A261G/A263L, R46H/E55R/K72C, R46H/E55R/K92A, R46H/E55R/Y118V,

R46H/E55R/Y118V/A263S, R46H/E55R/V146I, R46H/E55R/V146I/V202I, R46H/E55R/K226V/S257Q/A263L, R46H/E55R/A263L, R46H/E70N, R46H/E70N/K92A/Y118V/A261G/A263L, R46H/E70N/K175P/A263G, R46H/K72C, R46H/K92A, R46H/K92A/K175P, R46H/Y118V, R46H/V146I, R46H/V146I/S257Q/A263I, E55R/E70N/Y118V, E55R/S257Q/A261G, E55R/S257Q/A263L, E55R/A261G/A263I, E55R/A263G, E70N, K92A/Y118V, A105G, Y118V, Y118V/K226V/A261G/A263I, V146I/S257Q/A263S, V146I/A261G/A263L, E192V/S257Q/A261G, V202I, S257Q/A261G/A263G, A261G/A263I, and A263G.

[0149] As will be appreciated by the skilled artisan, in some embodiments, one or a combination of residue differences above that is selected can be kept constant (i.e., maintained) in the engineered nitroaldolase as a core feature, and additional residue differences at other residue positions incorporated into the sequence to generate additional engineered nitroaldolase polypeptides with improved properties. Accordingly, it is to be understood for any engineered nitroaldolase containing one or a subset of the residue differences above, the present invention contemplates other engineered nitroaldolases that comprise the one or subset of the residue differences, and additionally one or more residue differences at the other residue positions disclosed herein.

[0150] As noted above, the engineered nitroaldolase polypeptides are also capable of converting substrates (e.g., compound **(1)** and compound **(2)**) to products (e.g., compound **(3)**). In some embodiments, the engineered nitroaldolase polypeptide is capable of converting the substrate compounds to the product compound 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 activity relative to the activity of the reference polypeptide of SEQ ID NO: 2, 24, 496, 860, 1120, or 1348.

[0151] In some embodiments, the engineered nitroaldolase polypeptide capable of converting the substrate compounds to the product compounds with at least 2 fold the activity relative to SEQ ID NO: 2, 24, 496, 860, 1120, or 1348, comprises an amino acid sequence selected from: the even-numbered sequences in SEQ ID NOs: 4-1484.

[0152] In some embodiments, the engineered nitroaldolase polypeptide is capable of converting the substrate compounds to the product compound with stereoselectivity towards the (*S*)-product as compared to the reference polypeptide of SEQ ID NO: 2, 24, 496, 860, 1120, or 1348. In some embodiments, the engineered nitroaldolase polypeptide is capable of converting the substrate compounds to the product compound with stereoselectivity towards the (*S*)-product of at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or greater (as measured by the amount of (*S*)-product to the total amount of product).

[0153] In some embodiments, the engineered nitroaldolase polypeptide capable of converting the substrate compounds to the product compounds with stereoselectivity towards the (*S*)-product of at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or greater comprises an amino acid sequence selected from: the even-numbered sequences in SEQ ID NOs: 4-1484.

[0154] In some embodiments, the engineered nitroaldolase has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 2, 24, 496, 860, 1120, or 1348, increases expression of the engineered nitroaldolase activity in a bacterial host cell, particularly in *E. coli*.

[0155] In some embodiments, the engineered nitroaldolase polypeptide with improved properties has an amino acid sequence comprising a sequence selected from the even-numbered sequences in the range of SEQ ID NOs: 4-1484.

[0156] In some embodiments, the engineered nitroaldolase, comprises an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to one of the even-numbered sequences in the range of SEQ ID NOs: 4-1484, and the amino acid residue differences as compared to SEQ ID NO: 2, 24, 496, 860, 1120, or 1348, present in any one of the even-numbered sequences in the range of SEQ ID NOs: 4-1484, as provided in the Examples.

[0157] In addition to the residue positions specified above, any of the nitroaldolase polypeptides disclosed herein can further comprise other residue differences relative to SEQ ID NOs: 2, 24, 496, 860, 1120, and/or 1348, at other residue positions (i.e., residue positions other than those included herein). Residue differences at these other residue positions can provide for additional variations in the amino acid sequence without adversely affecting the ability of the polypeptide to carry out the conversion of substrate to product. Accordingly, in some embodiments, in addition to the amino acid residue differences present in any one of the engineered nitroaldolase polypeptides selected from the even-numbered sequences in the range of SEQ ID NOs: 4-1484, the sequence can further comprise 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-14, 1-15, 1-16, 1-18, 1-20, 1-22, 1-24, 1-26, 1-30, 1-35, 1-40, 1-45, or 1-50 residue differences at other amino acid residue positions as compared to the SEQ ID NO: 2, 24, 496, 860, 1120, or 1348. In some embodiments, the number of amino acid residue differences as compared to the reference sequence can be 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, 35, 40, 45 or 50 residue positions. In some embodiments, the number of amino acid residue differences as compared to the reference sequence can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, or 25 residue positions. The residue differences at these other positions can be conservative changes or non-conservative changes. In some embodiments, the residue differences can comprise conservative substitutions and non-conservative substitutions as compared to the nitroaldolase polypeptide of SEQ ID NO: 2, 24, 496, 860, 1120, or 1348.

[0158] In some embodiments, the present invention also provides engineered polypeptides that comprise a fragment of any of the engineered nitroaldolase polypeptides described herein that retains the functional activity and/or improved property of that engineered nitroaldolase. Accordingly, in some embodiments, the present invention provides a polypeptide fragment capable of converting substrate to product under suitable reaction conditions, wherein the fragment comprises at least about 90%, 95%, 96%, 97%, 98%, or 99% of a full-length amino acid sequence of an engineered nitroaldolase of the present invention, such as an exemplary engineered nitroaldolase polypeptide selected from the even-numbered sequences in the range of SEQ ID NOs: 4-1484. In some embodiments, the engineered nitroaldolase can have an amino

acid sequence comprising a deletion in any one of the nitroaldolase polypeptide sequences described herein, such as the exemplary engineered polypeptides of the even-numbered sequences in the range of SEQ ID NOs: 4-1484.

[0159] Thus, for each and every embodiment of the engineered nitroaldolases of the invention, the amino acid sequence can comprise deletions of one or more amino acids, 2 or more amino acids, 3 or more amino acids, 4 or more amino acids, 5 or more amino acids, 6 or more amino acids, 8 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, up to 20% of the total number of amino acids, or up to 30% of the total number of amino acids of the nitroaldolase polypeptides, where the associated functional activity and/or improved properties of the engineered nitroaldolase described herein are maintained. In some embodiments, the deletions can comprise 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 residues. In some embodiments, the number of deletions can be 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 residues. In some embodiments, the deletions can comprise deletions of 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 residues.

[0160] In some embodiments, the engineered nitroaldolase polypeptide described herein can have an amino acid sequence comprising an insertion as compared to any one of the engineered nitroaldolase polypeptides described herein, such as the exemplary engineered polypeptides of the even-numbered sequences in the range of SEQ ID NOs: 4-1484. Thus, for each and every embodiment of the nitroaldolase polypeptides of the disclosure, the insertions can comprise one or more amino acids, 2 or more amino acids, 3 or more amino acids, 4 or more amino acids, 5 or more amino acids, 6 or more amino acids, 8 or more amino acids, 10 or more amino acids, 15 or more amino acids, 20 or more amino acids, 30 or more amino acids, 40 or more amino acids, or 50 or more amino acids, where the associated functional activity and/or improved properties of the engineered nitroaldolase described herein is maintained. The insertions can be to amino or carboxy terminus, or internal portions of the nitroaldolase polypeptide.

[0161] In some embodiments, the engineered nitroaldolase herein can have an amino acid sequence comprising a sequence selected from the even-numbered sequences in the range of SEQ ID NOs: 4-1484, and optionally 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 number of 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.

[0162] In the above embodiments, the suitable reaction conditions for the engineered polypeptides are provided as described in the Examples herein.

[0163] In some embodiments, the polypeptides of the present invention are fusion polypeptides in which the engineered polypeptides are fused to other polypeptides, such as, by way of example and not limitation, antibody tags (e.g., myc epitope), purification sequences (e.g., His tags for binding to metals), and cell localization signals (e.g., secretion signals). Thus, the engineered polypeptides described herein can be used with or without fusions to other polypeptides.

[0164] It is to be understood that the polypeptides described herein are not restricted to the genetically encoded amino acids. In addition to the genetically encoded amino acids, the polypeptides described herein may be comprised, either in whole or in part, of naturally occurring and/or synthetic non-encoded amino acids. Certain commonly encountered non-encoded amino acids of which the polypeptides described herein may be comprised include, but are not limited to: the D-stereoisomers of the genetically encoded amino acids; 2,3-diaminopropionic acid (Dpr); α -aminoisobutyric acid (Aib); ϵ -aminohexanoic acid (Aha); δ -aminovaleric acid (Ava); N-methylglycine or sarcosine (MeGly or Sar); ornithine (Orn); citrulline (Cit); t-butylalanine (Bua); t-butylglycine (Bug); N-methylisoleucine (MeIle); phenylglycine (Phg); cyclohexylalanine (Cha); norleucine (Nle); naphthylalanine (Nal); 2-chlorophenylalanine (Ocf); 3-chlorophenylalanine (Mcf); 4-chlorophenylalanine (Pcf); 2-fluorophenylalanine (Off); 3-fluorophenylalanine (Mff); 4-fluorophenylalanine (Pff); 2-bromophenylalanine (Obf); 3-bromophenylalanine (Mbf); 4-bromophenylalanine (Pbf); 2-methylphenylalanine (OmF); 3-methylphenylalanine (MmF); 4-methylphenylalanine (PmF); 2-nitrophenylalanine (Onf); 3-nitrophenylalanine (Mnf); 4-nitrophenylalanine (Pnf); 2-cyanophenylalanine (Ocf); 3-cyanophenylalanine (Mcf); 4-cyanophenylalanine (Pcf); 2-trifluoromethylphenylalanine (Otf); 3-trifluoromethylphenylalanine (Mtf); 4-trifluoromethylphenylalanine (Ptf); 4-aminophenylalanine (Paf); 4-iodophenylalanine (Pif); 4-aminomethylphenylalanine (Pamf); 2,4-dichlorophenylalanine (Opef); 3,4-dichlorophenylalanine (Mpcf); 2,4-difluorophenylalanine (Opff); 3,4-difluorophenylalanine (Mpff); pyrid-2-ylalanine (2pAla); pyrid-3-ylalanine (3pAla); pyrid-4-ylalanine (4pAla); naphth-1-ylalanine (1nAla); naphth-2-ylalanine (2nAla); thiazolylalanine (taAla); benzothienylalanine (bAla); thienylalanine (tAla); furylalanine (fAla); homophenylalanine (hPhe); homotyrosine (hTyr); homotryptophan (hTrp); pentafluorophenylalanine (5ff); styrylalanine (sAla); authrylalanine (aAla); 3,3-diphenylalanine (Dfa); 3-amino-5-phenylpentanoic acid (Afp); penicillamine (Pen); 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid (Tic); β -2-thienylalanine (Thi); methionine sulfoxide (Mso); N(w)-nitroarginine (nArg); homolysine (hLys); phosphonomethylphenylalanine (pmPhe); phosphoserine (pSer); phosphothreonine (pThr); homoaspartic acid (hAsp); homoglutamic acid (hGlu); 1-aminocyclopent-(2 or 3)-ene-4 carboxylic acid; pipercolic acid (PA); azetidine-3-carboxylic acid (ACA); 1-aminocyclopentane-3-carboxylic acid; allylglycine (aGly); propargylglycine (pgGly); homoalanine (hAla); norvaline (nVal); homoleucine (hLeu), homovaline (hVal); homoisoleucine (hIle); homoarginine (hArg); N-acetyl lysine (AcLys); 2,4-diaminobutyric acid (Dbu); 2,3-diaminobutyric acid (Dab); N-methylvaline (MeVal); homocysteine (hCys); homoserine

(hSer); hydroxyproline (Hyp) and homoproline (hPro). Additional non-encoded amino acids of which the polypeptides described herein may be comprised will be apparent to those of skill in the art (See e.g., the various amino acids provided in Fasman, CRC Practical Handbook of Biochemistry and Molecular Biology, CRC Press, Boca Raton, FL, pp. 3-70 [1989], and the references cited therein, all of which are incorporated by reference). These amino acids may be in either the L- or D-configuration.

[0165] Those of skill in the art will recognize that amino acids or residues bearing side chain protecting groups may also comprise the polypeptides described herein. Non-limiting examples of such protected amino acids, which in this case belong to the aromatic category, include (protecting groups listed in parentheses), but are not limited to: Arg(tos), Cys(methylbenzyl), Cys (nitropyridinesulfonyl), Glu(δ -benzylester), Gln(xanthyl), Asn(N- δ -xanthyl), His(bom), His(benzyl), His(tos), Lys(fmoc), Lys(tos), Ser(O-benzyl), Thr (O-benzyl) and Tyr(O-benzyl).

[0166] Non-encoding amino acids that are conformationally constrained of which the polypeptides described herein may be composed include, but are not limited to, N-methyl amino acids (L-configuration); 1-aminocyclopent-(2 or 3)-ene-4-carboxylic acid; pipecolic acid; azetidine-3-carboxylic acid; homoproline (hPro); and 1-aminocyclopentane-3-carboxylic acid.

[0167] In some embodiments, the engineered polypeptides can be in various forms, for example, such as an isolated preparation, as a substantially purified enzyme, whole cells transformed with gene(s) encoding the enzyme, and/or as cell extracts and/or lysates of such cells. The enzymes can be lyophilized, spray-dried, precipitated or be in the form of a crude paste, as further discussed below.

[0168] In some embodiments, the engineered polypeptides can be in the form of a biocatalytic composition. In some embodiments, the biocatalytic composition comprises (a) a means for conversion of a ketone compound and an alkane compound to a β -nitro alcohol intermediate by contact with a nitroaldolase polypeptide and (b) a suitable buffer. In some embodiments, the biocatalytic composition comprises a nitroaldolase having activity on a ketone substrate. In some further embodiments, the β -nitro alcohol intermediate is further reduced to a β -hydroxy- α -amino alcohol. In some embodiments, the biocatalytic composition comprises a sodium citrate buffer.

[0169] In some embodiments, the engineered polypeptides can be provided on a solid support, such as a membrane, resin, solid carrier, or other solid phase material. A solid support can be composed of organic polymers such as polystyrene, polyethylene, polypropylene, polyfluoroethylene, polyethyleneoxy, and polyacrylamide, as well as co-polymers and grafts thereof. A solid support can also be inorganic, such as glass, silica, controlled pore glass (CPG), reverse phase silica or metal, such as gold or platinum. The configuration of a solid support can be in the form of beads, spheres, particles, granules, a gel, a membrane or a surface. Surfaces can be planar, substantially planar, or non-planar. Solid supports can be porous or non-porous, and can have swelling or non-swelling characteristics. A solid support can be configured in the form of a well, depression, or other container, vessel, feature, or location.

[0170] In some embodiments, the engineered nitroaldolase polypeptides of the present invention can be immobilized on a solid support such that they retain their improved activity, and/or other improved

properties relative to the reference polypeptide of SEQ ID NO: 2, 24, 496, 860, 1120, or 1348. In such embodiments, the immobilized polypeptides can facilitate the biocatalytic conversion of the substrate compounds or other suitable substrates to the product and after the reaction is complete are easily retained (e.g., by retaining beads on which polypeptide is immobilized) and then reused or recycled in subsequent reactions. Such immobilized enzyme processes allow for further efficiency and cost reduction. Accordingly, it is further contemplated that any of the methods of using the nitroaldolase polypeptides of the present invention can be carried out using the nitroaldolase polypeptides bound or immobilized on a solid support.

[0171] Methods of enzyme immobilization are well-known in the art. The engineered polypeptides can be bound non-covalently or covalently. Various methods for conjugation and immobilization of enzymes to solid supports (e.g., resins, membranes, beads, glass, etc.) are well known in the art (See e.g., Yi et al., *Proc. Biochem.*, 42(5): 895-898 [2007]; Martin et al., *Appl. Microbiol. Biotechnol.*, 76(4): 843-851 [2007]; Koszelewski et al., *J. Mol. Cat. B: Enzymatic*, 63: 39-44 [2010]; Truppo et al., *Org. Proc. Res. Dev.*, published online: [dx.doi.org/10.1021/op200157c](https://doi.org/10.1021/op200157c); Hermanson, *Bioconjugate Techniques*, 2nd ed., Academic Press, Cambridge, MA [2008]; Mateo et al., *Biotechnol. Prog.*, 18(3):629-34 [2002]; and “Bioconjugation Protocols: Strategies and Methods,” *In Methods in Molecular Biology*, Niemeyer (ed.), Humana Press, New York, NY [2004]; the disclosures of each which are incorporated by reference herein). Solid supports useful for immobilizing the engineered nitroaldolase of the present invention include but are not limited to beads or resins comprising polymethacrylate with epoxide functional groups, polymethacrylate with amino epoxide functional groups, styrene/DVB copolymer or polymethacrylate with octadecyl functional groups. Exemplary solid supports useful for immobilizing the engineered nitroaldolase polypeptides of the present invention include, but are not limited to, chitosan beads, Eupergit C, and SEPABEADs (Mitsubishi), including the following different types of SEPABEAD: EC-EP, EC-HFA/S, EXA252, EXE119 and EXE120.

[0172] In some embodiments, the polypeptides described herein are provided in the form of kits. The enzymes in the kits may be present individually or as a plurality of enzymes. The kits can further include reagents for carrying out the enzymatic reactions, substrates for assessing the activity of enzymes, as well as reagents for detecting the products. The kits can also include reagent dispensers and instructions for use of the kits.

[0173] In some embodiments, the kits of the present invention include arrays comprising a plurality of different nitroaldolase polypeptides at different addressable position, wherein the different polypeptides are different variants of a reference sequence each having at least one different improved enzyme property. In some embodiments, a plurality of polypeptides immobilized on solid supports are configured on an array at various locations, addressable for robotic delivery of reagents, or by detection methods and/or instruments. The array can be used to test a variety of substrate compounds for conversion by the polypeptides. Such arrays comprising a plurality of engineered polypeptides and methods of their use are known in the art (See e.g., WO2009/008908A2).

Polynucleotides Encoding Engineered Nitroaldolase, Expression Vectors and Host Cells

[0174] In another aspect, the present invention provides polynucleotides encoding the engineered nitroaldolase polypeptides described herein. The polynucleotides may be 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 nitroaldolase are introduced into appropriate host cells to express the corresponding nitroaldolase polypeptide.

[0175] 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 improved nitroaldolase enzymes. 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 amino acid sequences presented in Tables 2-1, 2-2, 3-1, 3-2, 4-1, 4-2, 5-1, 5-2, 6-1, 6-2, 7-1, and 7-2, and disclosed in the sequence listing incorporated by reference herein as the even-numbered sequences in the range of SEQ ID NOs: 4-1484.

[0176] 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 to express the gene in bacteria; preferred codons used in yeast are used for expression in yeast; and preferred codons used in mammals are used for expression in mammalian cells. In some embodiments, all codons need not be replaced to optimize the codon usage of the nitroaldolase since the natural sequence will comprise preferred codons and because use of preferred codons may not be required for all amino acid residues. Consequently, codon optimized polynucleotides encoding the nitroaldolase enzymes may contain preferred codons at about 40%, 50%, 60%, 70%, 80%, or greater than 90% of codon positions of the full length coding region.

[0177] In some embodiments, the polynucleotide comprises a codon optimized nucleotide sequence encoding the nitroaldolase polypeptide amino acid sequence, as represented by SEQ ID NO: 2, 24, 496, 860, 1120, or 1348. In some embodiments, the polynucleotide has a nucleic acid sequence comprising at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to the codon optimized nucleic acid sequences encoding the even-numbered sequences in the range of SEQ ID NOs: 4-1484. In some embodiments, the polynucleotide has a nucleic acid sequence comprising at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to the codon optimized nucleic acid sequences in the odd-numbered sequences in the range of SEQ ID NOs: 3-1483. In some embodiments, the codon optimized sequences of the odd-numbered sequences in the range of SEQ ID

NOs: 3-1483, enhance expression of the encoded nitroaldolase, providing preparations of enzyme capable of converting substrate to product.

[0178] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference sequence selected from the odd-numbered sequences in SEQ ID NOs: 3-1483, or a complement thereof, and encode a nitroaldolase.

[0179] In some embodiments, as described above, the polynucleotide encodes an engineered nitroaldolase polypeptide with improved properties as compared to SEQ ID NO: 2, 24, 496, 860, 1120, or 1348, 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 selected from SEQ ID NO: 2, 24, 496, 860, 1120, or 1348, and one or more residue differences as compared to SEQ ID NO: 2, 24, 496, 860, 1120, or 1348, wherein the sequence is selected from the even-numbered sequences in the range of SEQ ID NOs: 4-1484. In some embodiments, the reference amino acid sequence is selected from the even-numbered sequences in the range of SEQ ID NOs: 4-1484. In some embodiments, the reference amino acid sequence is SEQ ID NO: 2, while in some other embodiments, the reference sequence is SEQ ID NO: 24, while in some other embodiments, the reference sequence is SEQ ID NO: 496. In some embodiments, the reference amino acid sequence is SEQ ID NO: 860, while in some other embodiments, the reference sequence is SEQ ID NO: 1120, while in some other embodiments, the reference sequence is SEQ ID NO: 1348.

[0180] In some embodiments, the polynucleotide encodes a nitroaldolase polypeptide capable of converting one or more substrates to product with improved properties as compared to SEQ ID NO: 2, 24, 496, 860, 1120, or 1348, 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 sequence identity to reference sequence SEQ ID NO: 2, 24, 496, 860, 1120, or 1348.

[0181] In some embodiments, the polynucleotide encoding the engineered nitroaldolase comprises a polynucleotide sequence comprising at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOs: 1, 23, 495, 860, 1119, and/or 1347. In some embodiments, the polynucleotide encoding the engineered nitroaldolase comprises a polynucleotide sequence comprising at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483. In some embodiments, the polynucleotide encoding the engineered nitroaldolase comprises a polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483.

[0182] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483, or a complement thereof, and encode a nitroaldolase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a nitroaldolase 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 identity to SEQ ID NO: 2, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 2, at residue positions selected from: 7/9, 7/9/33/68/72/158, 7/9/33/72/175/197/230, 7/9/33/72/230, 7/9/68/72/197, 7/9/158/175, 7/9/175, 7/9/197, 7/33, 7/68/72/175, 7/72, 7/72/97/175/230, 9, 9/11/33/72/197/230, 9/33/68/158/197, 9/33/72/230, 9/33/97/230, 9/68/72, 9/97/175/202/218, 9/175, 9/202, 11, 13, 14/33/68/72/97, 18, 18/33/68/97/158/175/239, 18/33/72/97/158/175, 18/33/72/97/158/239, 18/33/72/97/175/202, 18/33/72/175/202/237/239, 18/33/72/197/237, 18/68/72, 18/68/72/158/175/237/239, 18/68/72/175/202/239, 18/68/72/239, 18/68/175/223/237/239, 18/68/239, 18/72/97/175/237/239, 18/72/175/197, 18/72/237, 18/72/239, 18/97/175/237/239, 18/97/223/237/239, 18/175, 18/197/239, 18/239, 20/124, 33/68/72, 33/68/72/175/230, 33/68/223/239, 33/72, 33/72/97/197, 33/72/175, 33/175, 39, 44/97, 52, 68/72, 68/72/97, 68/72/97/197, 68/72/197, 68/72/197/239, 68/97/175, 68/97/175/197/223, 72/97/158/223/237, 72/175/197, 97/158/175/197, 97/175, 103, 121, 122, 124, 125, 128, 133, 148, 175, 197, and 197/202.

[0183] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483, or a complement thereof, and encode a nitroaldolase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a nitroaldolase 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 identity to SEQ ID NO: 2, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 2, at residue positions selected from: 7/11/33/72/97, 7/11/33/97, 7/11/72, 7/11/158, 7/11/237, 7/33/72/97/158/175/197/237/239, 9/11/33, 9/11/33/72, 9/11/72, 9/33/68/97/237/239, 9/72/97/158/237/239, 9/218/239, 10, 11, 12, 13, 14, 18, 33/68/72/97/237/239, 33/72/158/239, 39, 40, 41, 48, 49, 51, 54, 54/126, 57, 68/175/239, 82, 84, 85, 86, 103, 104, 105, 107, 115, 117, 119, 119/120, 120, 122, 123, 124, 127, 130, 131, 133, 134, 137/145, 144, 145, 146, 147, 148, 149, 150, 152, 157, 158, 159, 169, 173, 174, 175, 180, 181, 183, 203, 204, 208, 209, 210, 215, 218, 236, and 237.

[0184] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483, or a complement thereof, and encode a nitroaldolase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a nitroaldolase 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 identity to SEQ ID NO: 24, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 24, at residue positions selected from: 9/39/122/157, 18, 18/33/39/230, 18/33/175, 18/39/44/118, 18/175, 20, 33/39, 33/39/44/118/121, 33/39/44/121, 33/39/44/152, 33/39/44/175, 33/39/44/197, 33/39/44/210/230, 33/39/118/121/175, 33/39/121/175,

33/39/197/210, 33/39/210, 33/39/210/230, 33/118/121, 33/118/121/175, 39, 39/44/103, 39/44/118, 39/44/152/230, 39/44/175/197/230, 39/103, 39/103/125, 39/103/125/127/146, 39/103/125/127/146/150, 39/103/125/146, 39/103/127/150, 39/103/150, 39/118/121, 39/197, 39/197/210, 43, 44/103, 81, 103, 103/125/127, 103/125/146, 103/127, 103/150, 118/121, 118/121/175, 118/121/197, 121, 152, 152/197, 154, 159, 163, 175, 210, 232, and 238.

[0185] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483, or a complement thereof, and encode a nitroaldolase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a nitroaldolase 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 identity to SEQ ID NO: 24, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 24, at residue positions selected from: 7, 9/13, 9/13/39, 9/13/39/68/72/239, 9/13/39/68/239, 9/13/39/72/122, 9/13/39/97/239, 9/13/68/72/97/239, 9/13/68/97/134/239, 9/13/68/97/239, 9/13/72, 9/13/134, 11, 12/42/125/127, 12/103, 12/103/150, 13, 13/39, 13/39/68, 13/39/68/72/97, 13/39/72/97/122/134/223, 13/39/97/239, 13/39/239, 13/68, 13/68/72/97/122/239, 16, 18/33/39/44/118/121/175/197/210, 18/33/39/44/121, 18/33/39/112/118/121/197, 18/33/39/121, 18/33/39/121/175, 18/33/39/121/175/197, 18/33/39/121/175/210, 18/33/39/121/210, 18/33/118/121, 18/33/121, 18/33/121/197, 18/33/121/210/230, 18/39/44/118/121/230, 18/39/44/121/175/210, 18/39/118/121, 18/39/118/121/210, 18/39/121/175, 18/39/121/175/210, 18/39/121/197/210, 18/118/121, 18/118/121/175, 18/118/121/210, 18/118/121/230, 18/121, 18/121/175, 18/121/210, 18/121/230, 20, 21, 33/39/44/118/121, 33/39/44/118/121/210/230, 33/39/44/121, 33/39/44/121/152, 33/39/44/121/175/210, 33/39/44/121/230, 33/39/73/121/197, 33/39/118/121/175, 33/39/118/121/175/230, 33/39/121, 33/39/121/175, 33/39/121/210/230, 33/118/121, 33/118/121/175, 33/118/121/197, 33/121, 33/121/230, 37/91, 39/44/118/121/175/210, 39/68/72/239, 39/118/121, 39/118/121/210, 39/121, 39/121/210/230, 43, 44, 76, 79, 101, 118/121, 118/121/175, 118/121/197, 121, 121/175/230, 121/210, 148, 153, 159, 161, 172, 200, 207, 233, 245, and 248.

[0186] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483, or a complement thereof, and encode a nitroaldolase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a nitroaldolase 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 identity to SEQ ID NO: 496, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 496, at residue positions selected from: 43/103, 43/103/172/238, 43/103/238/241, 43/238, 81/232, 103, 103/159/238, 103/194/238, 103/238, 122, 232, 238, and 238/241.

[0187] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483, or a complement thereof, and encode a nitroaldolase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a nitroaldolase 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 identity to SEQ ID NO: 496, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 496, at residue positions selected from: 12, 43, 43/47/103, 43/47/103/232, 43/103, 43/103/121/172/241, 43/103/172, 43/103/172/238, 43/103/238, 43/103/238/241, 43/121, 43/194/238, 43/238, 44/103/121/238, 44/103/238, 44/103/238/241, 44/121/238/241, 47/103, 63, 103, 103/121, 103/121/172/238, 103/159/238, 103/172/238, 103/194/238, 103/238, 122, 131, 146, 178, 180, and 221.

[0188] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483, or a complement thereof, and encode a nitroaldolase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a nitroaldolase 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 identity to SEQ ID NO: 860, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 860, at residue positions selected from: 12, 12/43/63/103/146/180/232/241, 12/43/63/103/146/241, 12/43/81/103/241, 12/43/81/146/232/241, 12/43/81/180/241, 12/43/81/232/241, 12/43/81/241, 12/43/103, 12/43/103/146/180/232/241, 12/43/103/146/180/241, 12/43/103/180/241, 12/43/103/241, 12/43/146/180/232/241, 12/43/180/241, 12/63/81/103/241, 12/81, 12/81/180/241, 12/103/146/241, 12/103/180/241, 12/146, 12/146/180/241, 12/180/232/241, 12/241, 18, 43/81/232/241, 43/103/146/180, 43/103/180/241, 43/103/241, 43/180/232, 43/232/241, 63/103/180/232, 78, 81, 81/103/146, 81/146/180/241, 81/146/241, 81/180/241, 84, 103, 103/146, 103/146/180, 103/146/180/241, 103/180, 103/180/241, 103/261, 144, 145, 146/180/232, 146/180/241, 146/241, 150, 166, 172, 180/232, 232/241, and 241.

[0189] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483, or a complement thereof, and encode a nitroaldolase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a nitroaldolase 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 identity to SEQ ID NO: 860, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 860, at residue positions selected from: 9, 12, 12/43/63/103/146/180/232/241, 12/43/63/103/146/241, 12/43/103/146/180/232/241,

12/43/103/146/180/241, 12/43/103/146/232/241, 12/43/103/146/241, 12/43/103/180/241, 12/43/146/180/232/241, 12/43/146/241, 12/43/180/232, 12/43/180/241, 12/63/103/146/232/241, 12/103/146/241, 12/103/180/241, 12/146, 12/146/180/241, 12/146/232/241, 12/146/241, 12/180, 12/180/232/241, 12/180/241, 12/241, 18, 43/63/103/180/232/241, 43/103/146/180, 43/103/180, 43/103/180/232/241, 43/103/180/241, 43/180/232, 63/103/180/232, 103, 103/146/180, 103/146/180/241, 103/180, 103/180/241, 103/241, 118, 119, 122, 144, 145, 146/180/241, 154, 160, 172, 180/232, 180/232/241, 180/241, 208, 211, and 212.

[0190] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483, or a complement thereof, and encode a nitroaldolase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a nitroaldolase 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 identity to SEQ ID NO: 1120, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 1120, at residue positions selected from: 3, 7, 7/158, 7/158/175, 7/158/197/202, 27/43/115/172, 27/43/172, 29, 32, 43/110/115/172, 44, 46, 48/172, 55, 59, 67, 71, 72, 73, 94/103/118/146, 103, 103/144/146/172/180, 103/146/154/172/180/232, 103/154, 118, 118/144/215, 118/172/180, 137, 142, 146, 146/154, 154, 158, 172/190, 184, 198, 205, 216, 223, 226, 229, 247, 261, and 263.

[0191] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483, or a complement thereof, and encode a nitroaldolase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a nitroaldolase 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 identity to SEQ ID NO: 1120, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 1120, at residue positions selected from: 73, 80, 103/146/209, 103/172/209, 118/119/172/209, 135, 144/209/232, 172/180/209, 181, and 223.

[0192] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483, or a complement thereof, and encode a nitroaldolase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a nitroaldolase 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 identity to SEQ ID NO: 1348, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 1348, at residue positions selected from: 2, 7/46, 39, 46,

46/55, 46/55/72, 46/55/92, 46/55/118, 46/55/226/257/263, 46/55/263, 46/70, 46/70/92/118/261/263, 46/70/175/263, 46/72, 46/118, 55/257/261, 55/261/263, 55/263, 105, 118, 118/226/261/263, and 263.

[0193] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-1483, or a complement thereof, and encode a nitroaldolase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a nitroaldolase 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 identity to SEQ ID NO: 1348, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 1348, at residue positions selected from: 2, 7/46, 7/46/175, 22/46/55/118, 27/46/72, 27/46/146/158/202, 27/46/175, 27/118, 29/70/118, 39, 46, 46/55, 46/55/59/261/263, 46/55/72, 46/55/92, 46/55/118, 46/55/118/263, 46/55/146, 46/55/146/202, 46/55/226/257/263, 46/55/263, 46/70, 46/70/92/118/261/263, 46/70/175/263, 46/72, 46/92, 46/92/175, 46/118, 46/146, 46/146/257/263, 55/70/118, 55/257/261, 55/257/263, 55/261/263, 55/263, 70, 92/118, 105, 118, 118/226/261/263, 146/257/263, 146/261/263, 192/257/261, 202, 257/261/263, 261/263, and 263.

[0194] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered nitroaldolase polypeptide with improved properties 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 identity to SEQ ID NO: 2, 24, 496, 860, 1120, or 1348. In some embodiments, the polynucleotides encode the polypeptides described herein but have at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity at the nucleotide level to a reference polynucleotide encoding the engineered nitroaldolase. In some embodiments, the reference polynucleotide sequence is selected from SEQ ID NOs: 3-1483.

[0195] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered nitroaldolase polypeptide with improved properties 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 identity to SEQ ID NO: 2. In some embodiments, the polynucleotides encode the polypeptides described herein but have at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity at the nucleotide level to a reference polynucleotide encoding the engineered nitroaldolase. In some embodiments, the reference polynucleotide sequence is selected from SEQ ID NOs: 3-461.

[0196] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered nitroaldolase polypeptide with improved properties 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 identity to SEQ ID NO: 24. In some embodiments, the polynucleotides

encode the polypeptides described herein but have at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity at the nucleotide level to a reference polynucleotide encoding the engineered nitroaldolase. In some embodiments, the reference polynucleotide sequence is selected from SEQ ID NOs: 459-841.

[0197] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered nitroaldolase polypeptide having with improved properties 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 identity to SEQ ID NO: 496. In some embodiments, the polynucleotides encode the polypeptides described herein but have at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity at the nucleotide level to a reference polynucleotide encoding the engineered nitroaldolase. In some embodiments, the reference polynucleotide sequence is selected from SEQ ID NOs: 843-979.

[0198] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered nitroaldolase polypeptide having with improved properties 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 identity to SEQ ID NO: 860. In some embodiments, the polynucleotides encode the polypeptides described herein but have at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity at the nucleotide level to a reference polynucleotide encoding the engineered nitroaldolase. In some embodiments, the reference polynucleotide sequence is selected from SEQ ID NOs: 981-1211.

[0199] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered nitroaldolase polypeptide having with improved properties 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 identity to SEQ ID NO: 1120. In some embodiments, the polynucleotides encode the polypeptides described herein but have at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity at the nucleotide level to a reference polynucleotide encoding the engineered nitroaldolase. In some embodiments, the reference polynucleotide sequence is selected from SEQ ID NOs: 1213-1365.

[0200] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered nitroaldolase polypeptide having with improved properties 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 identity to SEQ ID NO: 1348. In some embodiments, the polynucleotides encode the polypeptides described herein but have at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity at the nucleotide level to a reference polynucleotide encoding the engineered nitroaldolase. In some embodiments, the reference polynucleotide sequence is selected from SEQ ID NOs: 1367-1483.

[0201] In some embodiments, an isolated polynucleotide encoding any of the engineered nitroaldolase 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.

[0202] 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]).

[0203] 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*).

[0204] 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). 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]).

[0205] 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 nitroaldolase 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.

[0206] 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.

[0207] 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.

[0208] The present invention also provides recombinant expression vectors comprising a polynucleotide encoding an engineered nitroaldolase 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 combined 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 nitroaldolase 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.

[0209] 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 nitroaldolase 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.

[0210] 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.

[0211] 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 auxotrophy, and the like. Examples of bacterial selectable markers include, but are not limited to the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers, which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol or tetracycline resistance. Suitable markers for yeast host cells include, but are not limited to ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Selectable markers for use in a filamentous fungal host cell include, but are not limited to, *amdS* (acetamidase), *argB* (ornithine carbamoyltransferases), *bar* (phosphinothricin acetyltransferase), *hph* (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (orotidine-5'-phosphate decarboxylase), *sC* (sulfate adenylyltransferase), and *trpC* (anthranilate synthase), as well as equivalents thereof. In another aspect, the present invention provides a host cell comprising a polynucleotide encoding at least one engineered nitroaldolase polypeptide of the present invention, the polynucleotide being operatively linked to one or more control sequences for expression of the engineered nitroaldolase 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 (e.g., W3110 (Δ hlyA) and BL21).

[0212] Accordingly, in another aspect, the present invention provides methods for producing the engineered nitroaldolase polypeptides, where the methods comprise culturing a host cell capable of expressing a polynucleotide encoding the engineered nitroaldolase polypeptide under conditions suitable for expression of the polypeptide. In some embodiments, the methods further comprise the steps of isolating and/or purifying the nitroaldolase polypeptides, as described herein.

[0213] Appropriate culture media and growth conditions for the above-described host cells are well known in the art. Polynucleotides for expression of the nitroaldolase 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.

[0214] The engineered nitroaldolases with the properties disclosed herein can be obtained by subjecting the polynucleotide encoding the naturally occurring or engineered nitroaldolase 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]).

[0215] 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, 5,811,238, 5,830,721, 5,834,252, 5,837,458, 5,928,905, 6,096,548, 6,117,679, 6,132,970, 6,165,793, 6,180,406, 6,251,674, 6,265,201, 6,277,638, 6,287,861, 6,287,862, 6,291,242, 6,297,053, 6,303,344, 6,309,883, 6,319,713, 6,319,714, 6,323,030, 6,326,204, 6,335,160, 6,335,198, 6,344,356, 6,352,859, 6,355,484, 6,358,740, 6,358,742, 6,365,377, 6,365,408, 6,368,861, 6,372,497, 6,337,186, 6,376,246, 6,379,964, 6,387,702, 6,391,552, 6,391,640, 6,395,547, 6,406,855, 6,406,910, 6,413,745, 6,413,774, 6,420,175, 6,423,542, 6,426,224, 6,436,675, 6,444,468, 6,455,253, 6,479,652, 6,482,647, 6,483,011, 6,484,105, 6,489,146, 6,500,617, 6,500,639, 6,506,602, 6,506,603, 6,518,065, 6,519,065, 6,521,453, 6,528,311, 6,537,746, 6,573,098, 6,576,467, 6,579,678, 6,586,182, 6,602,986, 6,605,430, 6,613,514, 6,653,072, 6,686,515, 6,703,240, 6,716,631, 6,825,001, 6,902,922, 6,917,882, 6,946,296, 6,961,664, 6,995,017, 7,024,312, 7,058,515, 7,105,297, 7,148,054, 7,220,566, 7,288,375, 7,384,387, 7,421,347, 7,430,477, 7,462,469, 7,534,564, 7,620,500, 7,620,502, 7,629,170, 7,702,464, 7,747,391, 7,747,393, 7,751,986, 7,776,598, 7,783,428, 7,795,030, 7,853,410, 7,868,138, 7,783,428, 7,873,477, 7,873,499, 7,904,249, 7,957,912, 7,981,614, 8,014,961, 8,029,988, 8,048,674, 8,058,001, 8,076,138, 8,108,150, 8,170,806, 8,224,580, 8,377,681, 8,383,346, 8,457,903, 8,504,498, 8,589,085, 8,762,066, 8,768,871, 9,593,326, and all related US, as well as PCT and non-US counterparts; Ling et al., Anal. Biochem., 254(2):157-78 [1997]; Dale et al., Meth. Mol. Biol., 57:369-74 [1996]; Smith, Ann. Rev. Genet., 19:423-462 [1985]; Botstein et al., Science, 229:1193-1201 [1985]; Carter, Biochem. J., 237:1-7 [1986]; Kramer et al., Cell, 38:879-887 [1984]; Wells et al., Gene, 34:315-323 [1985]; Minshull et al., Curr. Op. Chem. Biol., 3:284-290 [1999]; Christians et al., Nat. Biotechnol., 17:259-264 [1999]; Cramer et al., Nature, 391:288-291 [1998]; Cramer et al., Nat. Biotechnol., 15:436-438 [1997]; Zhang et al., Proc. Nat. Acad. Sci. U.S.A., 94:4504-4509 [1997]; Cramer et al., Nat. Biotechnol., 14:315-319 [1996]; Stemmer, Nature, 370:389-391

[1994]; Stemmer, Proc. Nat. Acad. Sci. USA, 91:10747-10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767; and WO 2009/152336, all of which are incorporated herein by reference).

[0216] 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 nitroaldolase 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).

[0217] In some embodiments, the clones obtained following mutagenesis treatment can be screened for engineered nitroaldolases having one or more desired improved enzyme properties (e.g., improved regioselectivity). Measuring enzyme activity from the expression libraries can be performed using the standard biochemistry techniques, such as GC analysis, HPLC analysis and/or derivatization of products (pre or post separation), for example, using dansyl chloride or OPA (See e.g., Yaegaki et al., J Chromatogr. 356(1):163-70 [1986]).

[0218] When the sequence of the engineered polypeptide is known, 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 encoding portions of the nitroaldolase can be prepared by chemical synthesis as known in the art (e.g., the classical phosphoramidite method of Beaucage et al., Tet. Lett. 22:1859-69 [1981], or the method described by Matthes et al., EMBO J. 3:801-05 [1984]) as 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. In addition, essentially any nucleic acid can be obtained from any of a variety of commercial sources. In some embodiments, additional variations can be created by synthesizing oligonucleotides containing deletions, insertions, and/or substitutions, and combining the oligonucleotides in various permutations to create engineered nitroaldolases with improved properties.

[0219] Accordingly, in some embodiments, a method for preparing the engineered nitroaldolase polypeptide comprises: (a) synthesizing a polynucleotide encoding a polypeptide comprising an amino acid sequence having at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity to an amino acid sequence selected from the even-numbered sequences of SEQ ID NOs: 4-1484, and having one or more residue differences as compared to

SEQ ID NO: 2, 24, 496, 860, 1120, or 1348; and (b) expressing the nitroaldolase polypeptide encoded by the polynucleotide.

[0220] In some embodiments of the method, the polynucleotide encodes an engineered nitroaldolase that has optionally 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.

[0221] In some embodiments, any of the engineered nitroaldolase enzymes expressed in a host cell can be 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. Suitable solutions for lysing and the high efficiency extraction of proteins from bacteria, such as *E. coli*, are commercially available (e.g., CellLytic B™, Sigma-Aldrich, St. Louis MO).

[0222] Chromatographic techniques for isolation of the nitroaldolase polypeptide include, among others, reverse phase chromatography high performance liquid chromatography, ion exchange chromatography, gel electrophoresis, and affinity chromatography. Conditions for purifying a particular enzyme will depend, 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.

[0223] In some embodiments, affinity techniques may be used to isolate the improved nitroaldolase enzymes. For affinity chromatography purification, any antibody which specifically binds the nitroaldolase polypeptide may be used. For the production of antibodies, various host animals, including but not limited to rabbits, mice, rats, etc., may be immunized by injection with a nitroaldolase polypeptide, or a fragment thereof. The nitroaldolase polypeptide or fragment may be 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. In some embodiments, the affinity purification can use a specific ligand bound by the nitroaldolase or dye affinity column (See e.g., EP0641862; Stellwagen, "Dye Affinity Chromatography," *In Current Protocols in Protein Science*, Unit 9.2-9.2.16 [2001]).

Methods of Using the Engineered Nitroaldolase Enzymes

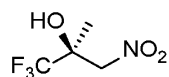
[0224] In some embodiments, the nitroaldolase enzymes described herein find use in processes for conversion of one or more suitable substrates to a product.

[0225] In another aspect, the engineered nitroaldolase polypeptides disclosed herein can be used in a process for the conversion of the substrate compound (1), or structural analogs thereof, and of the

substrate compound (2), or structural analogs thereof, to the product of compound (3) or the corresponding structural analog.

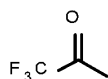
[0226] Structural analogs of compound (1) include other ketones with halogen modifications and/or ketones with various modifications at the alpha carbons. Structural analogs of compound (2) include other alkanes with nitro groups or other nitro containing compounds with an available α -hydrogen. Structural analogs of compound (3) include various β -nitro alcohols.

[0227] In some embodiments, the present disclosure provides a process of preparing compound (3);



Compound (3)

the process comprising a step of contacting a substrate of compound (1)



Compound (1)

with a substrate of compound (2)



Compound (2)

with an engineered nitroaldolase polypeptide as disclosed herein under suitable reaction conditions, such that a product of compound (3) is prepared.

[0228] In the embodiments provided herein and illustrated in the Examples, various ranges of suitable reaction conditions that can be used in the processes, include but are not limited to, substrate loading, co-substrate loading, pH, temperature, buffer, solvent system, polypeptide loading, and reaction time. Further suitable reaction conditions for carrying out the process for biocatalytic conversion of substrate compounds to product compounds using an engineered nitroaldolase described herein can be readily optimized in view of the guidance provided herein by routine experimentation that includes, but is not limited to, contacting the engineered nitroaldolase polypeptide and one or more substrate compounds under experimental reaction conditions of concentration, pH, temperature, and solvent conditions, and detecting the product compound.

[0229] The substrate compound(s) in the reaction mixtures can be varied, taking into consideration, for example, the desired amount of product compound, the effect of each substrate concentration on enzyme

activity, stability of enzyme under reaction conditions, and the percent conversion of each substrate to product. In some embodiments, the suitable reaction conditions comprise a substrate compound loading for each of one of more substrates of at least about 0.5 to about 60 g/L, 1 to about 40 g/L, 20 to about 35 g/L, about 10 to about 25 g/L, or 25 to about 35 g/L. In some embodiments, the suitable reaction conditions comprise a substrate compound loading for each of one of more substrates of at least about 0.5 g/L, at least about 1 g/L, at least about 5 g/L, at least about 10 g/L, at least about 15 g/L, at least about 20 g/L, at least about 30 g/L, at least about 35 g/L, or at least about 40 g/L, or even greater.

[0230] In carrying out the nitroaldolase mediated processes described herein, the engineered polypeptide may be added to the reaction mixture in the form of a purified enzyme, partially purified enzyme, whole cells transformed with gene(s) encoding the enzyme, as cell extracts and/or lysates of such cells, and/or as an enzyme immobilized on a solid support. Whole cells transformed with gene(s) encoding the engineered nitroaldolase enzyme or cell extracts, lysates thereof, and isolated enzymes may be employed in a variety of different forms, including solid (e.g., lyophilized, spray-dried, and the like) or semisolid (e.g., a crude paste). The cell extracts or cell lysates may be partially purified by precipitation (ammonium sulfate, polyethyleneimine, heat treatment or the like, followed by a desalting procedure prior to lyophilization (e.g., ultrafiltration, dialysis, etc.). Any of the enzyme preparations (including whole cell preparations) may be stabilized by crosslinking using known crosslinking agents, such as, for example, glutaraldehyde or immobilization to a solid phase (e.g., Eupergit C, and the like).

[0231] The gene(s) encoding the engineered nitroaldolase polypeptides can be transformed into host cell separately or together into the same host cell. For example, in some embodiments one set of host cells can be transformed with gene(s) encoding one engineered nitroaldolase polypeptide and another set can be transformed with gene(s) encoding another engineered nitroaldolase polypeptide. Both sets of transformed cells can be utilized together in the reaction mixture in the form of whole cells, or in the form of lysates or extracts derived therefrom. In other embodiments, a host cell can be transformed with gene(s) encoding multiple engineered nitroaldolase polypeptides. In some embodiments the engineered polypeptides can be expressed in the form of secreted polypeptides and the culture medium containing the secreted polypeptides can be used for the nitroaldolase reaction.

[0232] In some embodiments, the improved activity and/or stereoselectivity of the engineered nitroaldolase polypeptides disclosed herein provides for processes wherein higher percentage conversion can be achieved with lower concentrations of the engineered polypeptide. In some embodiments of the process, the suitable reaction conditions comprise an engineered polypeptide amount of about 1% (w/w), 2% (w/w), 5% (w/w), 10% (w/w), 20% (w/w), 30% (w/w), 40% (w/w), 50% (w/w), 75% (w/w), 100% (w/w), 150% (w/w), 175% (w/w), 200% (w/w) or more of substrate compound loading.

[0233] In some embodiments, the engineered polypeptide is present at about 0.01 g/L to about 130 g/L; about 0.05 g/L to about 100 g/L; about 0.1 g/L to about 40 g/L; about 1 g/L to about 40 g/L; about 2 g/L to about 40 g/L; about 5 g/L to about 40 g/L; about 5 g/L to about 30 g/L; about 0.1 g/L to about 10 g/L; about 0.5 g/L to about 10 g/L; about 1 g/L to about 10 g/L; about 0.1 g/L to about 5 g/L; about 0.5 g/L to

about 5 g/L; or about 0.1 g/L to about 2 g/L. In some embodiments, the nitroaldolase polypeptide is present at about 0.01 g/L, 0.05 g/L, 0.1 g/L, 0.2 g/L, 0.5 g/L, 1, 2 g/L, 5 g/L, 10 g/L, 15 g/L, 20 g/L, 25 g/L, 30 g/L, 35 g/L, 40 g/L, 50 g/L, 80 g/L, or 100 g/L.

[0234] During the course of the reaction, the pH of the reaction mixture may change. The pH of the reaction mixture may be maintained at a desired pH or within a desired pH range. This may be done by the addition of an acid or a base, before and/or during the course of the reaction. Alternatively, the pH may be controlled by using a buffer. Accordingly, in some embodiments, the reaction condition comprises a buffer. Suitable buffers to maintain desired pH ranges are known in the art and include, by way of example and not limitation, sodium citrate, borate, phosphate, 2-(N-morpholino)ethanesulfonic acid (MES), 3-(N-morpholino)propanesulfonic acid (MOPS), acetate, triethanolamine, and 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris), and the like. In some embodiments, the reaction conditions comprise water as a suitable solvent with no buffer present.

[0235] In the embodiments of the process, the reaction conditions comprise a suitable pH. The desired pH or desired pH range can be maintained by use of an acid or base, an appropriate buffer, or a combination of buffering and acid or base addition. The pH of the reaction mixture can be controlled before and/or during the course of the reaction. In some embodiments, the suitable reaction conditions comprise a solution pH from about 4 to about 10, pH from about 4 to about 6, pH from about 5 to about 6, pH from about 6 to about 9, pH from about 5 to about 7. In some embodiments, the reaction conditions comprise a solution pH of about 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, or 10.

[0236] In the embodiments of the processes herein, a suitable temperature is used for the reaction conditions, for example, taking into consideration the increase in reaction rate at higher temperatures, and the activity of the enzyme during the reaction time period. Accordingly, in some embodiments, the suitable reaction conditions comprise a temperature of about 10°C to about 60°C, about 10°C to about 55°C, about 15°C to about 60°C, about 20°C to about 60°C, about 20°C to about 55°C, about 25°C to about 55°C, or about 30°C to about 50°C. In some embodiments, the suitable reaction conditions comprise a temperature of about 10°C, 15°C, 20°C, 22°C, 25°C, 30°C, 35°C, 40°C, 45°C, 50°C, 55°C, or 60°C. In some embodiments, the temperature during the enzymatic reaction can be maintained at a specific temperature throughout the course of the reaction. In some embodiments, the temperature during the enzymatic reaction can be adjusted over a temperature profile during the course of the reaction.

[0237] In some embodiments, the processes of the invention are carried out in a solvent. Suitable solvents include water, aqueous buffer solutions, organic solvents, polymeric solvents, and/or co-solvent systems, which generally comprise aqueous solvents, organic solvents and/or polymeric solvents. The aqueous solvent (water or aqueous co-solvent system) may be pH-buffered or unbuffered. In some embodiments, the processes using the engineered nitroaldolase polypeptides can be carried out in an aqueous co-solvent system comprising an organic solvent (e.g., ethanol, isopropanol (IPA), dimethyl sulfoxide (DMSO), dimethylformamide (DMF) ethyl acetate, butyl acetate, 1-octanol, heptane, octane, methyl tert-butyl ether (MTBE), toluene, and the like), ionic or polar solvents (e.g., 1-ethyl-4-

methylimidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium hexafluorophosphate, glycerol, polyethylene glycol, and the like). In some embodiments, the co-solvent can be a polar solvent, such as a polyol, dimethylsulfoxide (DMSO), or lower alcohol. The non-aqueous co-solvent component of an aqueous co-solvent system may be miscible with the aqueous component, providing a single liquid phase, or may be partly miscible or immiscible with the aqueous component, providing two liquid phases. Exemplary aqueous co-solvent systems can comprise water and one or more co-solvents selected from an organic solvent, polar solvent, and polyol solvent. In general, the co-solvent component of an aqueous co-solvent system is chosen such that it does not adversely inactivate the nitroaldolase enzyme under the reaction conditions. Appropriate co-solvent systems can be readily identified by measuring the enzymatic activity of the specified engineered nitroaldolase enzyme with a defined substrate of interest in the candidate solvent system, utilizing an enzyme activity assay, such as those described herein.

[0238] In some embodiments of the process, the suitable reaction conditions comprise an aqueous co-solvent, where the co-solvent comprises DMSO at about 1% to about 50% (v/v), about 1 to about 40% (v/v), about 2% to about 40% (v/v), about 5% to about 30% (v/v), about 10% to about 30% (v/v), or about 10% to about 20% (v/v). In some embodiments of the process, the suitable reaction conditions can comprise an aqueous co-solvent comprising ethanol at about 1% (v/v), about 5% (v/v), about 10% (v/v), about 15% (v/v), about 20% (v/v), about 25% (v/v), about 30% (v/v), about 35% (v/v), about 40% (v/v), about 45% (v/v), or about 50% (v/v).

[0239] In some embodiments, the reaction conditions comprise a surfactant for stabilizing or enhancing the reaction. Surfactants can comprise non-ionic, cationic, anionic and/or amphiphilic surfactants. Exemplary surfactants, include by way of example and not limitation, nonyl phenoxypolyethoxylethanol (NP40), TRITON™ X-100 polyethylene glycol *tert*-octylphenyl ether, polyoxyethylene-stearylamine, cetyltrimethylammonium bromide, sodium oleylamidosulfate, polyoxyethylene-sorbitanmonostearate, hexadecyldimethylamine, etc. Any surfactant that may stabilize or enhance the reaction may be employed. The concentration of the surfactant to be employed in the reaction may be generally from 0.1 to 50 mg/ml, particularly from 1 to 20 mg/ml.

[0240] In some embodiments, the reaction conditions include an antifoam agent, which aids in reducing or preventing formation of foam in the reaction solution, such as when the reaction solutions are mixed or sparged. Anti-foam agents include non-polar oils (e.g., minerals, silicones, etc.), polar oils (e.g., fatty acids, alkyl amines, alkyl amides, alkyl sulfates, etc.), and hydrophobic (e.g., treated silica, polypropylene, etc.), some of which also function as surfactants. Exemplary anti-foam agents include, Y-30® (Dow Corning), poly-glycol copolymers, oxy/ethoxylated alcohols, and polydimethylsiloxanes. In some embodiments, the anti-foam can be present at about 0.001% (v/v) to about 5% (v/v), about 0.01% (v/v) to about 5% (v/v), about 0.1% (v/v) to about 5% (v/v), or about 0.1% (v/v) to about 2% (v/v). In some embodiments, the anti-foam agent can be present at about 0.001% (v/v), about 0.01% (v/v), about 0.1%

(v/v), about 0.5% (v/v), about 1% (v/v), about 2% (v/v), about 3% (v/v), about 4% (v/v), or about 5% (v/v) or more as desirable to promote the reaction.

[0241] The quantities of reactants used in the nitroaldolase reaction will generally vary depending on the quantities of product desired, and concomitantly the amount of nitroaldolase substrate employed. Those having ordinary skill in the art will readily understand how to vary these quantities to tailor them to the desired level of productivity and scale of production.

[0242] In some embodiments, the order of addition of reactants is not critical. The reactants may be added together at the same time to a solvent (e.g., monophasic solvent, biphasic aqueous co-solvent system, and the like), or alternatively, some of the reactants may be added separately, and some together at different time points. For example, the cofactor, co-substrate and substrate may be added first to the solvent.

[0243] The solid reactants (e.g., enzyme, salts, etc.) may be provided to the reaction in a variety of different forms, including powder (e.g., lyophilized, spray dried, and the like), solution, emulsion, suspension, and the like. The reactants can be readily lyophilized or spray dried using methods and equipment that are known to those having ordinary skill in the art. For example, the protein solution can be frozen at -80°C in small aliquots, then added to a pre-chilled lyophilization chamber, followed by the application of a vacuum.

[0244] For improved mixing efficiency when an aqueous co-solvent system is used, the nitroaldolase and co-substrate may be added and mixed into the aqueous phase first. The nitroaldolase substrate may be added and mixed in, followed by the organic phase or the substrate may be dissolved in the organic phase and mixed in. Alternatively, the nitroaldolase substrate may be premixed in the organic phase, prior to addition to the aqueous phase.

[0245] The processes of the present invention are generally allowed to proceed until further conversion of substrate to product does not change significantly with reaction time (e.g., less than 10% of substrate being converted, or less than 5% of substrate being converted). In some embodiments, the reaction is allowed to proceed until there is complete or near complete conversion of substrate to product. Transformation of substrate to product can be monitored using known methods by detecting substrate and/or product, with or without derivatization. Suitable analytical methods include gas chromatography, HPLC, MS, and the like.

[0246] In some embodiments of the process, the suitable reaction conditions comprise a substrate loading for each of one or more substrates of at least about 5 g/L, 10 g/L, 20 g/L, 30 g/L, 40 g/L, or more, and wherein the method results in at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or greater conversion of substrate compound to product compound in about 24 h or less, in about 12 h or less, in about 6 h or less, or in about 4 h or less.

[0247] The engineered nitroaldolase polypeptides of the present invention when used in the process under suitable reaction conditions result in an excess of the desired product in at least 30%, 40%, 50%, 60%, or greater enantiomeric excess over undesired product(s).

[0248] In some further embodiments of the processes for converting one or more substrate compounds to product compound using the engineered nitroaldolase polypeptides, the suitable reaction conditions can comprise an initial substrate loading for each of one or more substrates to the reaction solution which is then contacted by the polypeptide. This reaction solution is then further supplemented with additional substrate compound as a continuous or batchwise addition over time at a rate of at least about 1 g/L/h, at least about 2 g/L/h, at least about 4 g/L/h, at least about 6 g/L/h, or higher for each of one or more substrate compounds. Thus, according to these suitable reaction conditions, polypeptide is added to a solution having an initial substrate loading of at least about 1 g/L, 5 g/L, or 10 g/L for each of one or more substrate compounds. This addition of polypeptide is then followed by continuous addition of further substrate to the solution at a rate of about 2 g/L/h, 4 g/L/h, or 6 g/L/h for each of one or more substrate compounds until a much higher final substrate loading of at least about 30 g/L or more for each of one or more substrate compounds, is reached. Accordingly, in some embodiments of the process, the suitable reaction conditions comprise addition of the polypeptide to a solution having an initial substrate loading of at least about 1 g/L, 5 g/L, or 10 g/L followed by addition of further substrate to the solution at a rate of about 2 g/L/h, 4 g/L/h, or 6 g/L/h until a final substrate loading of at least about 30 g/L, or more, is reached for each of one or more substrate compounds. This substrate supplementation reaction condition allows for higher substrate loadings to be achieved while maintaining high rates of conversion of substrate to product of at least about 5%, 25%, 50%, 75%, 90% or greater conversion of substrate for either or both of one or more substrate compounds.

[0249] Any of the processes disclosed herein using the engineered polypeptides for the preparation of compound (3) can be carried out under a range of suitable reaction conditions, including but not limited to ranges of ketone substrates, ranges of nitroalkane substrates, temperature, pH, solvent system, substrate loading, polypeptide loading, and reaction time. In one example, in some embodiments, the preparation of compound (3) can be carried out wherein the suitable reaction conditions comprise: (a) ketone substrate loading of about 0.01 M to 1 M of substrate compound; (b) amino acid substrate loading of about 0.01 M to 1 M of substrate compound; (c) of about 0.5 g/L to 100 g/L of engineered polypeptide; (d) 0.01 M-1 M sodium citrate buffer; (e) pH at 5-8; and (f) temperature of about 15°C to 60°C. In some embodiments, the suitable reaction conditions comprise: (a) about 0.5 M of compound (1) substrate compound; (b) about 0.3 M of compound (2) substrate compound; (c) about 50 g/L of each engineered polypeptide; (d) 0.1 M sodium citrate buffer; (e) static pH at 6, and (g) about 22°C.

[0250] In some embodiments, additional reaction components or additional techniques carried out to supplement the reaction conditions. These can include taking measures to stabilize or prevent inactivation of the enzyme, reduce product inhibition, shift reaction equilibrium to formation of the desired product.

[0251] In further embodiments, any of the above described process for the conversion of one or more substrate compounds to product compound can further comprise one or more steps selected from: extraction; isolation; purification; and crystallization of product compound. Methods, techniques, and protocols for extracting, isolating, purifying, and/or crystallizing the product from biocatalytic reaction

mixtures produced by the above disclosed processes are known to the ordinary artisan and/or accessed through routine experimentation. Additionally, illustrative methods are provided in the Examples below.

[0252] Various features and embodiments of the invention are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

EXPERIMENTAL

[0253] 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.

[0254] In the Examples below, the following abbreviations apply: ppm (parts per million); M (molar); mM (millimolar), uM and μ M (micromolar); nM (nanomolar); mol (moles); gm and g (gram); mg (milligrams); ug and μ g (micrograms); L and l (liter); ml and mL (milliliter); cm (centimeters); mm (millimeters); um and μ m (micrometers); sec. (seconds); min(s) (minute(s)); h(s) and hr(s) (hour(s)); U (units); MW (molecular weight); rpm (rotations per minute); psi and PSI (pounds per square inch); °C (degrees Centigrade); RT and rt (room temperature); CAM and cam (chloramphenicol); DMSO (dimethylsulfoxide); PMBS (polymyxin B sulfate); IPTG (isopropyl β -D-1-thiogalactopyranoside); LB (Luria-Bertani broth); TB (Terrific Broth; 12 g/L bacto-tryptone, 24 g/L yeast extract, 4 mL/L glycerol, 65 mM potassium phosphate, pH 7.0, 1 mM MgSO_4); PLP (pyridoxal 5'-phosphate), TEOA (triethanolamine buffer), HEPES (HEPES zwitterionic buffer; 4-(2-hydroxyethyl)-piperazineethanesulfonic acid); SFP (shake flask powder); CDS (coding sequence); DNA (deoxyribonucleic acid); RNA (ribonucleic acid); *E. coli* W3110 (commonly used laboratory *E. coli* strain, available from the Coli Genetic Stock Center [CGSC], New Haven, CT); HTP (high throughput); HPLC (high pressure liquid chromatography); GC (gas chromatography); FIOPC (fold improvements over positive control); Microfluidics (Microfluidics, Corp., Westwood, MA); Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO); Difco (Difco Laboratories, BD Diagnostic Systems, Detroit, MI); Agilent (Agilent Technologies, Inc., Santa Clara, CA); Corning (Corning, Inc., Palo Alto, CA); Dow Corning (Dow Corning, Corp., Midland, MI); and Gene Oracle (Gene Oracle, Inc., Mountain View, CA).

EXAMPLE 1

Production of Engineered Polypeptides in pCK110900

[0255] The polynucleotide (SEQ ID NO: 1) encoding the polypeptide having nitroaldolase activity (SEQ ID NO: 2), was cloned into the pCK110900 vector system (See e.g., US Pat. No. 9,714,437, which is hereby incorporated by reference in its entirety) and subsequently expressed in *E. coli* W3110/*fluA* under the control of the *lac* promoter. This polynucleotide, and associated polypeptide, was derived from an oxynitrilase found in *Baliospermum montanum*.

[0256] In a 96-well format, single colonies were picked and grown in 190 μ L LB media containing 1% glucose and 30 μ g/mL CAM, at 30°C, 200 rpm with 85% humidity. Following overnight growth, 20 μ L

of the grown cultures were transferred into a deep-well plate containing 380 μ L of TB media with 30 μ g/mL CAM. The cultures were grown at 30°C, 250 rpm with 85% humidity for approximately 2.5 hours. When the optical density (OD₆₀₀) of the cultures reached 0.4–0.6, expression of the nitroaldolase gene was induced by the addition of IPTG to a final concentration of 1 mM. Following induction, growth continued for 18–20 hours at 30°C, 250 rpm with 85% humidity. Cells were harvested by centrifugation at 4,000 rpm and 4°C for 10 minutes; the supernatant was then discarded. The cell pellets were stored at -80°C until ready for use.

[0257] Prior to performing the assay, the cell pellets were thawed and resuspended in 200 μ L of lysis buffer containing 1 g/L lysozyme, 0.5 g/L PMBS and 0.025 μ L/mL of commercial DNase (New England BioLabs, M0303L) in 0.1 M sodium citrate buffer at pH 6. The plates were agitated with medium-speed shaking for 2 hours on a microtiter plate shaker at room temperature. The plates were then centrifuged at 4,000 rpm for 15 minutes at 4°C, and the clarified supernatants were used in the HTP assay reaction described in the following examples.

[0258] Shake-flask procedures can be used to generate engineered nitroaldolase or polypeptide shake-flask powders (SFP), which are useful for secondary screening assays and/or use in the biocatalytic processes described herein. Shake flask powder preparation of enzymes provides a more purified preparation (e.g., up to 30% of total protein) of the engineered enzyme, as compared to the cell lysate used in HTP assays and also allows for the use of more concentrated enzyme solutions. To start the culture, a single colony of *E. coli* containing a plasmid encoding an engineered polypeptide of interest was inoculated into 5 mL of LB cell culture media with 30 μ g/mL CAM and 1% glucose. The culture was grown overnight (at least 16 hours) in an incubator at 37°C, with shaking at 250 rpm. The grown culture was then added to 250 mL of TB media with 30 μ g/mL CAM, in a 1 L shake-flask. The 250 mL culture was grown at 30°C and 250 rpm, for 3.5 hours until OD₆₀₀ reached 0.6–0.8. Expression of the nitroaldolase gene was induced by the addition of IPTG to a final concentration of 1 mM, and growth was continued for an additional 18-20 hours. Cells were harvested by transferring the culture into a pre-weighed centrifuge bottle which was then centrifuged at 4,000 rpm for 20 minutes at 4°C. The supernatant was discarded, and the remaining cell pellet was weighed. In some embodiments, the cell pellet was stored at -80°C until ready to use. For lysis, the cell pellet was resuspended in 6 mL/g wet cell weight of 25 mM triethanolamine-HCl buffer at pH 7 or 25 mM sodium citrate buffer at pH 6 and lysed using a 110L MICROFLUIDIZER[®] processor system (Microfluidics). Cell debris was removed by centrifugation at 10,000 rpm for 60 minutes at 4°C. The clarified lysate was collected, frozen at -80°C, and then lyophilized, using standard methods known in the art. Lyophilization of frozen clarified lysate provides a dry shake-flask powder comprising crude engineered polypeptide.

EXAMPLE 2

Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 2 for Improved Production of Compound (3)

[0259] The engineered polynucleotide (SEQ ID NO: 1) encoding the polypeptide with nitroaldolase activity of SEQ ID NO: 2 was used to generate the engineered polypeptides of Table 2-1. These polypeptides displayed improved nitroaldolase activity under the desired conditions e.g., the improvement in the formation of the β -nitro alcohol, compound (3), from the substrates trifluoroacetone and nitromethane, compounds (1) and (2), respectively, as compared to the starting polypeptide. Similarly, the engineered polypeptides of Table 2-2 displayed improved nitroaldolase selectivity towards the desired (S)-alcohol. The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the “backbone” amino acid sequence of SEQ ID NO: 2, as described below.

[0260] Directed evolution began with the polynucleotide set forth in SEQ ID NO: 1. Libraries of engineered polypeptides were generated using various well-known techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods that measured the polypeptide’s ability to produce compound (3).

[0261] The enzyme assays were carried out in 96-well deep-well (1.1 mL total volume) plates, in 100 μ L total reaction volume per well. The reactions contained 67 v/v% of undiluted nitroaldolase lysate, prepared as described in EXAMPLE 1, 40 mM trifluoroacetone (1), and 500 mM nitromethane (2) dissolved in 100 mM sodium citrate at pH 6. The reaction plates were heat-sealed and shaken at 250 rpm at 22°C for ~22 hours.

[0262] After overnight incubation (~22 hours), 300 μ L/well of methyl *tert*-butyl ether (MTBE) was added to the reaction plates and mixed well. The plates were sealed and centrifuged at 4,000 rpm for 10 min. An aliquot of the upper organic phase was removed and added to shallow 96 well plates for analysis by gas chromatography (GC) as described in EXAMPLE 8.

[0263] Hit variants were grown in 250 mL shake flasks as described in EXAMPLE 1 to generate lyophilized enzyme powders. The activity of the enzyme powders were evaluated by measuring the enzyme dose response at 0-25 g/L of the nitroaldolase shake flask powder under assay conditions, as described in this EXAMPLE. These hit variants are denoted with a * in Tables 2-1 and 2-2.

Table 2-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2)	Fold Improvement in Activity Relative to SEQ ID NO: 2
3/4*	I68L/V97I/K175P/I197V/K223P	+++
5/6*	A33V/K72E/V97I/I197V	+++

Table 2-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2)	Fold Improvement in Activity Relative to SEQ ID NO: 2
7/8	I9V/A33V/V97I/V230I	+++
9/10*	H103S	+++
11/12	L18C/V97I/K223P/I237P/I239F	+++
13/14	K72E/K175P/I197V	+++
15/16	I68L/K72E/V97I	+++
17/18	I68L/K72E/V97I/I197V	+++
19/20	I9V/I68L/K72E	+++
21/22	C13G	+++
23/24*	T11S	+++
25/26	L18C/K72E/V97I/K175P/I237P/I239F	+++
27/28	I7V/K72E/V97I/K175P/V230I	+++
29/30	A33V/K72E/K175P	+++
31/32	I7V/I68L/K72E/K175P	+++
33/34	I7V/I9V/I68L/K72E/I197V	+++
35/36	L18C/A33V/K72E/V97I/K175P/V202I	+++
37/38	A33V/I68L/K72E	++
39/40	L18C/A33V/K72E/K175P/V202I/I237P/I239F	++
41/42	I9V/V202I	++
43/44	I7V/I9V/K175P	++
45/46	I9V/K175P	++
47/48	L18C/K72E/K175P/I197V	++
49/50	L18C/I68L/K72E/K175P/V202I/I239F	++
51/52	I9V/A33V/K72E/V230I	++
53/54	I9V	++
55/56	A33V/I68L/K72E/K175P/V230I	++
57/58	H103M	++
59/60	L18C/A33V/K72E/V97I/F158Y/I239F	++
61/62	L18C/A33V/K72E/I197V/I237P	++
63/64	I7V/I9V/A33V/K72E/K175P/I197V/V230I	++
65/66	L18C/I68L/K72E/F158Y/K175P/I237P/I239F	++
67/68	I68L/K72E/I197V	++
69/70	L18C/A33V/I68L/V97I/F158Y/K175P/I239F	++
71/72	I68L/K72E	++

Table 2-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2)	Fold Improvement in Activity Relative to SEQ ID NO: 2
73/74	I7V/I9V	++
75/76	I7V/I9V/A33V/K72E/V230I	++
77/78	L18I	++
79/80	L18C/K72E/I239F	++
81/82	I68L/V97I/K175P	++
83/84	V97I/K175P	++
85/86	I7V/I9V/I197V	++
87/88	I9V/A33V/I68L/F158Y/I197V	++
89/90	L18C/I68L/K175P/K223P/I237P/I239F	+
91/92	L18C/I68L/K72E/I239F	+
93/94	V39A	+
95/96	I9V/T11G/A33V/K72E/I197V/V230I	+
97/98	L18C/A33V/K72E/V97I/F158Y/K175P	+
99/100	V97I/F158Y/K175P/I197V	+
101/102	I7V/K72E	+
103/104	I7V/I9V/A33V/I68L/K72E/F158Y	+
105/106	L18C/I197V/I239F	+
107/108	L18C/I68L/K72E	+
109/110	I9V/V97I/K175P/V202I/Q218M	+
111/112	A33V/K175P	+
113/114	I68L/K72E/I197V/I239F	+
115/116	A33V/K72E	+
117/118	L18C/K72E/I237P	+
119/120	S122C	+
121/122	V124Y	+
123/124	A33V/I68L/K223P/I239F	+
125/126	I197V/V202I	+
127/128	I197V	+
129/130	K72E/V97I/F158Y/K223P/I237P	+
131/132	V124E	+
133/134	L18C/I68L/I239F	+
135/136	V124T	+
137/138	S122Q	+

Table 2-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2)	Fold Improvement in Activity Relative to SEQ ID NO: 2
139/140	H14Y/A33V/I68L/K72E/V97I	+
141/142	K175P	+
143/144	L18C/I239F	+
145/146	L18C/K175P	+
147/148	V39G	+
149/150	W128Y	+
151/152	L18C/V97I/K175P/I237P/I239F	+
153/154	I7V/I9V/F158Y/K175P	+
155/156	I7V/A33V	+
157/158	F121Y	+
159/160	L148M	+
161/162	D44N/V97I	+
163/164	V124C	+
165/166	F125R	+
167/168	W128H	+
169/170	G52S	+
171/172	Y20H/V124C	+
173/174	F133C	+
175/176	F133A	+
Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 2 and defined as follows: "+" 1.14 to 1.62, "++" > 1.62, "+++" > 1.99		

Table 2-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2)	Improvement in Selectivity Relative to SEQ ID NO: 2
177/178*	I7V/T11G/K72E	+++
179/180	I9V/T11G/A33V/K72E	+++
181/182	I7V/T11G/A33V/K72E/V97I	+++
183/184	I9V/T11G/K72E	+++
185/186	I7V/T11G/I237P	+++
187/188	I7V/T11G/A33V/V97I	+++

Table 2-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2)	Improvement in Selectivity Relative to SEQ ID NO: 2
189/190	I9V/T11G/A33V	+++
191/192	M107G	+++
193/194	F158R	+++
195/196	I7V/T11G/F158Y	+++
197/198	I86W	+++
199/200	Q218T	+++
201/202*	F158I	+++
203/204	I12K	+++
205/206	R174G	+++
207/208	G204R	+++
209/210	V39W	+++
211/212	S134D	+++
213/214	S122T	+++
215/216	A40L	+++
217/218	I237W	+++
219/220	V146D	+++
221/222	A105D	+++
223/224	C13L	+++
225/226	D130V	+++
227/228	G82T	+++
229/230	G204V	+++
231/232	V117L	+++
233/234	M107W	+++
235/236	K236L	++
237/238	L183K	++
239/240	K120L	++
241/242	K119N/K120F	++
243/244	H14R	++
245/246	H10R	++
247/248	K236G	++
249/250	G204E	++
251/252	W54Y	++

Table 2-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2)	Improvement in Selectivity Relative to SEQ ID NO: 2
253/254	T11R	++
255/256	Y203Q	++
257/258	I237S	++
259/260	I68L/K175P/I239F	++
261/262	N104G	++
263/264	A169I	++
265/266	E123P	++
267/268	D130M	++
269/270	N85Q	++
271/272	H103P	++
273/274	W54N/T126I	++
275/276	N85M	++
277/278	A115K	++
279/280	N85L	++
281/282	Q215L	++
283/284	A115V	++
285/286	A115L	++
287/288	T152P	++
289/290	Y57N	++
291/292	H10G	++
293/294	S131L	++
295/296	F210V	++
297/298	R174C	++
299/300	G149W	++
301/302	L18T	++
303/304	K120C	++
305/306	K119L	++
307/308	S159L	++
309/310	A40V	++
23/24*	T11S	++
311/312	Q208V	++
313/314	N85A	++

Table 2-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2)	Improvement in Selectivity Relative to SEQ ID NO: 2
315/316	S159H	++
317/318	E180G	++
319/320	F158L	+
321/322	I7V/A33V/K72E/V97I/F158Y/K175P/I197V/I237P/I239F	+
323/324	A33V/I68L/K72E/V97I/I237P/I239F	+
325/326	F158W	+
327/328	I9V/A33V/I68L/V97I/I237P/I239F	+
329/330	I51Q	+
331/332	I237L	+
333/334	V146N	+
335/336	A145P	+
337/338	S159G	+
339/340	N104A	+
341/342	Q208C	+
343/344	V117C	+
345/346	Q208G	+
347/348	Q208T	+
349/350	A115C	+
351/352	L148P	+
353/354	A40T	+
355/356	S134W	+
357/358	T137I/A145V	+
359/360	H10P	+
361/362	I51A	+
363/364	L18S	+
365/366	I51S	+
367/368	A169G	+
369/370	E49R	+
371/372	V173Y	+
373/374	L48W	+
375/376	Q208A	+

Table 2-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2)	Improvement in Selectivity Relative to SEQ ID NO: 2
377/378	F158M	+
379/380	D150I	+
381/382	Q208L	+
383/384	I209C	+
385/386	A169S	+
387/388	D150L	+
389/390	K120Y	+
391/392	I9V/K72E/V97I/F158Y/I237P/I239F	+
393/394	S134L	+
395/396	V124P	+
397/398	K175D	+
399/400	F133W	+
401/402	V146G	+
403/404	T144L	+
405/406	S41C	+
407/408	A145F	+
409/410	F158Y	+
411/412	I9V/Q218M/I239F	+
413/414	A33V/K72E/F158Y/I239F	+
415/416	Q208E	+
417/418	Q181C	+
419/420	S122G	+
421/422	D150V	+
423/424	D127L	+
425/426	I51T	+
427/428	E49Q	+
429/430	I84L	+
431/432	S134V	+
433/434	E180V	+
435/436	E123A	+
437/438	V173E	+
439/440	E147R	+

Table 2-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2)	Improvement in Selectivity Relative to SEQ ID NO: 2
441/442	D127M	+
443/444	V39R	+
445/446	S134F	+
447/448	L148V	+
449/450	A145L	+
451/452	E123T	+
453/454	D127A	+
455/456	N104S	+
457/458	I157V	+
459/460	A40S	+
461/462	S134R	+
Levels of increased selectivity (fraction of product that is the (S)-alcohol) were determined relative to the reference polypeptide of SEQ ID NO: 2 and defined as follows: "+" 0.40 to 0.47, "++" > 0.47, "+++" > 0.52		

EXAMPLE 3

Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 24 for Improved Production of Compound (3)

[0264] The engineered polynucleotide (SEQ ID NO: 23) encoding the polypeptide with nitroaldolase activity of SEQ ID NO: 24 was used to generate the engineered polypeptides of Table 3-1. These polypeptides displayed improved nitroaldolase activity under the desired conditions e.g., the improvement in the formation of the β -nitro alcohol, compound (3), from the substrates trifluoroacetone and nitromethane, compounds (1) and (2), respectively, as compared to the starting polypeptide. Similarly, the engineered polypeptides of Table 3-2 displayed improved nitroaldolase selectivity towards the desired (S)-alcohol. The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the “backbone” amino acid sequence of SEQ ID NO: 24, as described below.

[0265] Directed evolution began with the polynucleotide set forth in SEQ ID NO: 23. Libraries of engineered polypeptides were generated using various well-known techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods that measured the polypeptide's ability to produce compound (3).

[0266] The enzyme assay and analysis were carried using a similar method as described in EXAMPLE 2. The high throughput cell pellets were lysed in 400 μ L/well of lysis buffer and the reaction was carried out with 93 v/v% clarified lysate, 0.3 M trifluoroacetone (1) and 0.5 M nitromethane (2).

[0267] Hit variants were grown in 250 mL shake flasks, and enzyme powders were generated as described in EXAMPLE 1. The activity of the enzyme powders was evaluated at 0-100 g/L of the nitroaldolase SFP, 0.3 M trifluoroacetone (1), 0.5 M nitromethane (2), 0.1 M sodium citrate, pH 6, 22°C, for 22 hours, using similar assay as described in EXAMPLE 2. These hit variants are denoted with a * in Tables 3-1 and 3-2.

Table 3-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 24		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 24)	Fold Improvement in Activity Relative to SEQ ID NO: 24
463/464*	A33V/V39W/Y118L/F121Y/K175P	+++
465/466	A33V/V39W/I197V/F210M	+++
467/468*	Q238M	+++
469/470	G81F	+++
471/472	A33V/V39W/D44N/I197V	+++
473/474	V39W/H103S	+++
475/476	H103V	+++
477/478	S232G	+++
479/480	V39W/D44Y/H103S	+++
481/482	A33V/Y118L/F121Y	+++
483/484*	A33V/V39W/F121Y/K175P	+++
485/486	H103S	+++
487/488*	A33V/V39W/D44N/Y118L/F121Y	+++
489/490	D44N/H103S	++
491/492	H103S/D150P	++
493/494	V39W/D44N/K175P/I197V/V230I	++
495/496*	A33V/V39W/D44N/F121Y	++
497/498	A33V/V39W/F210M/V230I	++
499/500	Y118L/F121Y/I197V	++
501/502	I43M	++
503/504	H103L	++
505/506	T152F	++
507/508	V39W/H103S/F125P/D127G/V146L/D150P	++

Table 3-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 24		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 24)	Fold Improvement in Activity Relative to SEQ ID NO: 24
509/510	I9V/V39W/S122T/I157V	++
511/512	Y118L/F121Y/K175P	++
513/514	A33V/V39W/D44N/T152F	++
515/516*	I43S	++
517/518	H103Q	++
519/520	V39W/D44N/T152F/V230I	++
521/522	S159Q	++
523/524	A33V/V39W/F210M	++
525/526	A33V/V39W	++
527/528	H103M/F125P/V146L	++
529/530	V39W/H103S/D127G/D150P	+
531/532	Y118L/F121Y	+
533/534	A33V/V39W/D44N/F210M/V230I	+
535/536	L18I/A33V/K175P	+
537/538	I43R	+
539/540	I43T	+
541/542	F121Y	+
543/544	V39W	+
545/546	K175P	+
547/548	L18I/A33V/V39W/V230I	+
549/550	V39W/H103M/D150P	+
551/552	V39W/H103S/F125R	+
553/554	F210M	+
555/556	A33V/Y118L/F121W/K175P	+
557/558	H103M	+
559/560	L18I/K175P	+
561/562	I43Q	+
563/564	A33V/V39W/D44N/K175P	+
565/566	V39W/H103M/F125R/V146L	+
567/568	V39W/H103S/F125R/D127G/V146L	+
569/570	V39W/H103M	+
571/572	T152A/I197V	+

Table 3-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 24		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 24)	Fold Improvement in Activity Relative to SEQ ID NO: 24
573/574	A154I	+
575/576	V39W/Y118L/F121Y	+
577/578	H103S/F125P/D127G	+
579/580	V39W/H103Q	+
581/582	L18I	+
583/584	V39W/I197V/F210M	+
585/586	V39W/I197V	+
587/588	L18I/V39W/D44N/Y118L	+
589/590	H103M/D127G	+
591/592	V39W/D44N/Y118L	+
593/594	Y20H	+
595/596	I163P	+
Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 24 and defined as follows: "+" 1.20 to 1.35, "++" > 1.35, "+++" > 1.55		

Table 3-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 24		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 24)	Improvement in Selectivity Relative to SEQ ID NO: 24
597/598	A33V/F121W	+++
495/496*	A33V/V39W/D44N/F121Y	+++
599/600	F121W	+++
601/602	Y118L/F121W	+++
603/604	A33V/Y118L/F121W/I197V	+++
605/606	A33V/Y118L/F121W/K175P	+++
607/608	A33V/F121Y/V230I	+++
487/488*	A33V/V39W/D44N/Y118L/F121Y	+++
609/610	A33V/V39W/D44N/F121Y/V230I	+++
611/612	L18I/A33V/V39W/D44N/F121T	+++
613/614	A33V/V39W/F121W	+++
541/542	F121Y	+++

Table 3-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 24		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 24)	Improvement in Selectivity Relative to SEQ ID NO: 24
615/616	V39W/F121W	+++
481/482	A33V/Y118L/F121Y	+++
617/618	A33V/V39W/Y118L/F121W/K175P	+++
483/484*	A33V/V39W/F121Y/K175P	+++
531/532	Y118L/F121Y	+++
499/500	Y118L/F121Y/I197V	+++
463/464*	A33V/V39W/Y118L/F121Y/K175P	+++
619/620*	I43E	+++
621/622	F121Y/K175P/V230I	+++
623/624	L172R	+++
511/512	Y118L/F121Y/K175P	+++
625/626	L18I/Y118L/F121W/F210M	+++
627/628	I9V/C13G/V39W/K72E/S122T	+++
629/630	L18I/V39W/D44N/Y118L/F121W/V230I	+++
631/632	I43W	++
575/576	V39W/Y118L/F121Y	++
633/634	I9V/C13G	++
635/636	A33V/V39W/Y118L/F121Y/K175P/V230I	++
637/638	L18I/Y118L/F121Y/V230I	++
639/640	A33V/V39W/D44N/F121Y/K175P/F210M	++
641/642	I9V/C13G/I68L/V97I/I239F	++
643/644	I9V/C13G/K72E	++
645/646	I9V/C13G/S134D	++
647/648	F121Y/F210M	++
649/650	I9V/C13G/V39W/V97I/I239F	++
651/652	A33V/V39W/D44N/Y118L/F121W/F210M/V230I	++
653/654	L18I/A33V/V39W/F121T/K175P	++
655/656	L172Q	++
657/658	L18C/A33V/Y118L/F121W	++
659/660	Y20R	++
661/662	A33V/V39W/K73N/F121Y/I197V	++
663/664	L18I/Y118L/F121Y	++

Table 3-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 24		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 24)	Improvement in Selectivity Relative to SEQ ID NO: 24
665/666	V39W/D44N/Y118L/F121W/K175P/F210M	++
667/668	L172E	++
669/670	V39W/Y118L/F121W/F210M	++
671/672	L18I/F121Y/F210M	++
673/674	C13G/V39W/K72E/V97I/S122T/S134D/K223P	++
675/676	L18C/Y118L/F121W	++
677/678	I9V/C13G/V39W	++
679/680	L18I/F121T/K175P	++
681/682	V39W/Y118L/F121Y/F210M	++
561/562	I43Q	++
683/684	Y20G	++
685/686	A33V/V39W/F121W/F210M/V230I	++
687/688	C13G/I68L	++
689/690	A33V/V39W/D44N/F121V	++
691/692	L18C/F121W	++
693/694	L18I/Y118L/F121W	++
695/696	C13G/V39W	++
697/698	I9V/C13G/I68L/K72E/V97I/I239F	++
699/700	C13G/I68L/K72E/V97I/S122T/I239F	++
701/702	L18C/F121Y	++
703/704	C13G/V39W/I68L/K72E/V97I	++
705/706	L18I/A33V/Y118L/F121W	++
707/708	L18C/F121W/F210M	++
515/516*	I43S	+
709/710	L18C/A33V/Y118L/F121Y	+
711/712	L18C/F121Y/V230I	+
713/714	L18C/V39W/Y118L/F121W	+
715/716	V39W/F121W/F210M/V230I	+
717/718	C13G/V39W/V97I/I239F	+
719/720	L18C/A33V/F121Y	+
721/722	L18C/Y118L/F121Y	+
723/724	L18I/Y118L/F121W/K175P	+

Table 3-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 24		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 24)	Improvement in Selectivity Relative to SEQ ID NO: 24
725/726	L18I/F121W	+
727/728	L18C/A33V/F121Y/I197V	+
729/730	L18C/F121W/K175P	+
731/732	L18I/V39W/F121V/K175P	+
733/734	A33V/V39W/F121V	+
735/736	L18I/A33V/V39W/F121W	+
737/738	L18I/A33V/V39W/F121V/K175P/I197V	+
739/740	L18C/A33V/V39W/F121W/F210M	+
741/742	Y20T	+
743/744	L18I/A33V/V39W/D44N/Y118L/F121W/K175P/I197V/F210M	+
745/746	I9V/C13G/V39W/I68L/K72E/I239F	+
747/748	L18C/A33V/V39W/F121Y/K175P	+
749/750	L18C/A33V/F121Y/F210M/V230I	+
751/752	A33V/V39G/D44N/F121W/T152A	+
753/754	C13G/V39W/I239F	+
755/756	I9V/C13G/I68L/V97I/S134D/I239F	+
757/758	I9V/C13G/V39W/I68L/I239F	+
759/760	C13G/V39W/I68L	+
761/762	D44V	+
763/764	L18C/V39W/F121W/I197V/F210M	+
765/766	L18C/A33V/V39W/H112R/Y118L/F121W/I197V	+
767/768	I12A/H103S	+
769/770	I12A/H103S/D150P	+
771/772	L172A	+
773/774	L18C/F121V/K175P	+
775/776	L18C/F121V	+
777/778	L18C/V39W/F121W/K175P/F210M	+
779/780	F121V	+
781/782	V39G/I68L/K72E/I239F	+
783/784	Y20K	+
785/786	C13F	+

Table 3-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 24		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 24)	Improvement in Selectivity Relative to SEQ ID NO: 24
787/788	L18C/A33V/V39W/F121W/K175P/F210M	+
789/790	L18I/V39W/D44N/F121V/K175P/F210M	+
791/792	L18I/V39W/Y118L/F121W/F210M	+
793/794	L148F	+
795/796	L172G	+
797/798	A16V	+
799/800	K21V	+
801/802	S159Q	+
803/804	L245G	+
805/806	D37L/E91G	+
807/808	L18C/V39W/D44N/F121Y/K175P/F210M	+
809/810	L18I/Y118L/F121T	+
811/812	E79F	+
813/814	L153W	+
815/816	I248S	+
817/818	A233R	+
819/820	S161M	+
821/822	S11N	+
823/824	L76P	+
825/826	V101L	+
827/828	I7E	+
829/830	L172V	+
831/832	I12A/G42W/F125R/D127G	+
833/834	A16M	+
835/836	D207N	+
837/838	L245D	+
839/840	V200G	+
841/842	I248G	+
Levels of increased selectivity (fraction of product that is the (S)-alcohol) were determined relative to the reference polypeptide of SEQ ID NO: 24 and defined as follows: "+" 0.50 to 0.52, "++" > 0.52, "+++" > 0.54		

EXAMPLE 4**Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 496 for Improved Production of Compound (3)**

[0268] The engineered polynucleotide (SEQ ID NO: 495) encoding the polypeptide with nitroaldolase activity of SEQ ID NO: 496 was used to generate the engineered polypeptides of Table 4-1. These polypeptides displayed improved nitroaldolase activity under the desired conditions e.g., the improvement in the formation of the β -nitro alcohol, compound (3), from the substrates trifluoroacetone and nitromethane, compounds (1) and (2), respectively, as compared to the starting polypeptide. Similarly, the engineered polypeptides of Table 4-2 displayed improved nitroaldolase selectivity towards the desired (S)-alcohol. The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the “backbone” amino acid sequence of SEQ ID NO: 496, as described below.

[0269] Directed evolution began with the polynucleotide set forth in SEQ ID NO: 495. Libraries of engineered polypeptides were generated using various well-known techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods that measured the polypeptide’s ability to produce compound (3).

[0270] The enzyme assay and analysis were carried using the same method as described in EXAMPLE 3.

[0271] Hit variants were grown in 250 mL shake flasks, and enzyme powders were generated as described in EXAMPLE 1. The activity of the enzyme powders was evaluated as described in EXAMPLE 3. These hit variants are denoted with a * in Tables 4-1 and 4-2.

Table 4-1. Nitroaldolase Activity for the Production of Compound (3)		
Relative to SEQ ID NO: 496		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 496)	Fold Improvement in Activity Relative to SEQ ID NO: 496
843/844	Q238M	+++
845/846	H103S/S159Q/Q238M	+++
847/848	H103S/Q238M	+++
849/850	H103S/Y194L/Q238M	++
851/852	I43Q/Q238M	+
853/854	H103S	+
855/856	I43S/Q238M	+
857/858	Q238M/K241R	+
859/860*	I43S/H103S/L172R/Q238M	+
861/862	I43Q/H103S/Q238M/K241R	+

Table 4-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 496		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 496)	Fold Improvement in Activity Relative to SEQ ID NO: 496
863/864*	I43E/H103S/Q238M/K241R	+
865/866	S232G	++
867/868	H103V	++
869/870	G81F/S232G	++
871/872	I43M/H103L	++
873/874	H103L	+
875/876	S122R	+++
877/878	S122I	++
879/880	S122C	++
881/882	S122Q	++
883/884	S122M	+
885/886	S122T	+
887/888	S122V	+
889/890	S122K	+
891/892	S122L	+
Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 496 and defined as follows: "+" 1.14 to 1.26, "++" > 1.26, "+++" > 1.35		

Table 4-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 496		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 496)	Improvement in Selectivity Relative to SEQ ID NO: 496
859/860*	I43S/H103S/L172R/Q238M	+++
893/894*	I43E/H103S/Q238M	+++
895/896	H103S/Y121W/L172R/Q238M	+++
863/864*	I43E/H103S/Q238M/K241R	+++
897/898	I43S/H103S/Q238M/K241R	+++
861/862	I43Q/H103S/Q238M/K241R	+++
899/900	I43E/H103S/Y121W/L172R/K241R	+++
901/902	H103S/L172R/Q238M	+++
903/904	N44V/H103S/Y121W/Q238M	+++
905/906	N44V/H103S/Q238M/K241R	+++
851/852	I43Q/Q238M	++
855/856	I43S/Q238M	++

Table 4-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 496		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 496)	Improvement in Selectivity Relative to SEQ ID NO: 496
907/908	N44V/H103S/Q238M	++
909/910	I43S/Y194L/Q238M	++
845/846	H103S/S159Q/Q238M	++
911/912	I43Q/Y121W	++
847/848	H103S/Q238M	++
913/914	I43S	+
915/916	N44V/Y121W/Q238M/K241R	+
917/918	H103S/Y194L/Q238M	+
919/920	H103S/Y121W	+
853/854	H103S	+
921/922*	I43Q/H103L/L172R	+++
923/924*	I43Q/Q47G/H103S/S232G	++
925/926*	I43S/H103S	++
927/928	I43S/H103L	++
929/930	I43Q/H103S	++
931/932	I43M/H103V	++
933/934	I43M/Q47G/H103V	++
935/936	H103V	++
937/938	I43M/H103S	+
939/940	I43M/Q47G/H103L	+
871/872	I43M/H103L	+
941/942	H103S	+
943/944	I43S/H103S/L172R	+
945/946	H103L	+
947/948	Q47G/H103L	+
949/950*	S122R	++
883/884	S122M	++
951/952	S122I	+
953/954	S122L	+
955/956	S122H	+
957/958	S122V	+
959/960	S122T	+
961/962	S122K	+
963/964	I12V	+
965/966	S122N	+
967/968*	E180L	+
969/970	E180V	+
971/972	N221G	+
973/974	S131Y	+
975/976	T63P	+

Table 4-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 496		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 496)	Improvement in Selectivity Relative to SEQ ID NO: 496
977/978	F178A	+
979/980	V146I	+
Levels of increased selectivity (fraction of product that is the (S)-alcohol) were determined relative to the reference polypeptide of SEQ ID NO: 496 and defined as follows: "+" 0.55 to 0.58, "++" > 0.58, "+++" > 0.60		

EXAMPLE 5

Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 860 for Improved Production of Compound (3)

[0272] The engineered polynucleotide (SEQ ID NO: 859) encoding the polypeptide with nitroaldolase activity of SEQ ID NO: 860 was used to generate the engineered polypeptides of Table 5-1. These polypeptides displayed improved nitroaldolase activity under the desired conditions e.g., the improvement in the formation of the β -nitro alcohol, compound (3), from the substrates trifluoroacetone and nitromethane, compounds (1) and (2), respectively, as compared to the starting polypeptide. Similarly, the engineered polypeptides of Table 5-2 displayed improved nitroaldolase selectivity towards the desired (S)-alcohol. The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the "backbone" amino acid sequence of SEQ ID NO: 860, as described below.

[0273] Directed evolution began with the polynucleotide set forth in SEQ ID NO: 859. Libraries of engineered polypeptides were generated using various well-known techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods that measured the polypeptide's ability to produce compound (3).

[0274] The enzyme assay and analysis were carried using the same method as described in EXAMPLE 3.

[0275] Hit variants were grown in 250 mL shake flasks, and enzyme powders were generated as described in EXAMPLE 1. The activity of the enzyme powders was evaluated as described in EXAMPLE 3. These hit variants are denoted with a * in Tables 5-1 and 5-2.

Table 5-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 860		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 860)	Fold Improvement in Activity Relative to SEQ ID NO: 860
981/982	S103T	+++

Table 5-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 860		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 860)	Fold Improvement in Activity Relative to SEQ ID NO: 860
983/984	S43M/S103V/V146I/E180V	+++
985/986*	S103V/V146I/E180L/K241R	+++
987/988*	I12V/S43M/T63P/S103L/V146I/E180L/S232G/K241R	+++
989/990	I12V/S43M/G81F/S103L/K241R	+++
991/992	S43Q/S103V/V146I/E180L	+++
993/994	S43M/E180L/S232G	+++
995/996	I12V/S43Q/S103L/V146I/E180V/S232G/K241R	+++
997/998	I12V/V146I	+++
999/1000	S103T/A261V	+++
1001/1002	I12V/E180L/S232G/K241R	+++
1003/1004	I12V/S43M/G81F/V146I/S232G/K241R	+++
1005/1006	I12V/S43M/G81F/K241R	+++
1007/1008	I12V/S43M/E180L/K241R	++
1009/1010	V146I/E180V/S232G	++
1011/1012	I12V/S43Q/S103L/E180V/K241R	++
1013/1014	I12V/S43M/S103V/K241R	++
1015/1016	S43M/S103L/V146I/E180V	++
1017/1018	I12V/S43Q/S103L/V146I/E180V/K241R	++
1019/1020	S103A	++
1021/1022	I12V/S103V/E180V/K241R	++
1023/1024	S43M/S103L/E180V/K241R	++
1025/1026	I12V/S43M/S103L/E180L/K241R	++
1027/1028	I12V/S103L/V146I/K241R	++
1029/1030*	I12V/S43M/V146I/E180L/S232G/K241R	++
1031/1032	T144V	++
1033/1034*	I12V/S43M/S103V/V146I/E180V/K241R	++
1035/1036	I12V/G81F/E180V/K241R	++
1037/1038	I12V/S43M/G81F/S232G/K241R	++
1039/1040	K241R	++
1041/1042	V146I/E180L/K241R	++
1043/1044	V166L	++

Table 5-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 860		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 860)	Fold Improvement in Activity Relative to SEQ ID NO: 860
1045/1046	R172L	++
1047/1048	I12V/K241R	++
1049/1050	V146I/K241R	+
1051/1052	S103L/E180L	+
1053/1054	T63P/S103L/E180L/S232G	+
1055/1056	S103N	+
1057/1058	S43M/G81F/S232G/K241R	+
1059/1060	S103L/E180V/K241R	+
1061/1062	S103L/V146I/E180V	+
1063/1064	S103L/V146I	+
1065/1066	I12V/G81F	+
1067/1068	S103L	+
1069/1070*	I12V/S103L/E180L/K241R	+
1071/1072	G81F/E180L/K241R	+
1073/1074	A145V	+
1075/1076	I12V	+
1077/1078	I12V/V146I/E180L/K241R	+
1079/1080	S43M/S232G/K241R	+
1081/1082	L18V	+
1083/1084	I12V/S43M/G81F/E180V/K241R	+
1085/1086	I12V/E180V/S232G/K241R	+
1087/1088	E180L/S232G	+
1089/1090	G81F/V146I/K241R	+
1091/1092	R172V	+
1093/1094	I12V/S43Q/T63P/S103L/V146I/K241R	+
1095/1096	G81F/S103L/V146I	+
1097/1098	S232G/K241R	+
1099/1100	G81F/V146I/E180L/K241R	+
1101/1102	I12V/S43Q/S103V	+
1103/1104	D150H	+
1105/1106	S43M/S103L/K241R	+
1107/1108	I12V/S43M/S103L/K241R	+

Table 5-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 860		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 860)	Fold Improvement in Activity Relative to SEQ ID NO: 860
1109/1110	I12V/T63P/G81F/S103L/K241R	+
1111/1112	I12V/S43M/E180V/K241R	+
1113/1114	G81Q	+
1115/1116	I84Y	+
1117/1118	G78A	+
Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 860 and defined as follows: "+" 1.12 to 1.28, "++" > 1.28, "+++" > 1.39		

Table 5-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 860		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 860)	Improvement in Selectivity Relative to SEQ ID NO: 860
1119/1120*	I12V/S103V/E180L/K241R	+++
1121/1122*	Q208F	+++
1033/1034*	I12V/S43M/S103V/V146I/E180V/K241R	+++
985/986*	S103V/V146I/E180L/K241R	+++
1123/1124*	S103N	+++
1125/1126	S103A	+++
987/988*	I12V/S43M/T63P/S103L/V146I/E180L/S232G/K241R	+++
1127/1128	S103T	+++
1069/1070*	I12V/S103L/E180L/K241R	+++
1029/1030*	I12V/S43M/V146I/E180L/S232G/K241R	+++
991/992	S43Q/S103V/V146I/E180L	+++
983/984	S43M/S103V/V146I/E180V	+++
1017/1018	I12V/S43Q/S103L/V146I/E180V/K241R	+++
1015/1016	S43M/S103L/V146I/E180V	+++
1041/1042	V146I/E180L/K241R	+++
1129/1130	I12V/S43Q/T63P/S103V/V146I/K241R	++
1021/1022	I12V/S103V/E180V/K241R	++
1131/1132	S43Q/S103V/E180L/K241R	++
1133/1134	I12V/S43M/S103L/V146I/K241R	++
1025/1026	I12V/S43M/S103L/E180L/K241R	++
1135/1136	I12V/S43M/E180L/S232G	++
1031/1032	T144V	++
1077/1078	I12V/V146I/E180L/K241R	++
1137/1138	S43Q/S103V/E180V/S232G/K241R	++

Table 5-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 860		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 860)	Improvement in Selectivity Relative to SEQ ID NO: 860
1061/1062	S103L/V146I/E180V	++
1139/1140	S103L/E180V/K241R	++
1087/1088	E180L/S232G	++
1007/1008	I12V/S43M/E180L/K241R	++
1085/1086	I12V/E180V/S232G/K241R	++
1141/1142	I12V/T63P/S103L/V146I/S232G/K241R	++
1143/1144	I12V/S43Q/S103L/V146I/K241R	++
1145/1146	I12V/S103L/E180V/K241R	++
1147/1148	I12V/E180L/K241R	++
1011/1012	I12V/S43Q/S103L/E180V/K241R	++
1001/1002	I12V/E180L/S232G/K241R	++
1149/1150	S43M/S103L/E180V/K241R	++
995/996	I12V/S43Q/S103L/V146I/E180V/S232G/K241R	++
1151/1152	S103L/V146I/E180V/K241R	++
1153/1154	I12V/E180L	+
1051/1052	S103L/E180L	+
1155/1156	I12V/V146I/S232G/K241R	+
993/994	S43M/E180L/S232G	+
997/998	I12V/V146I	+
1157/1158	E180V/S232G	+
1159/1160	I12V/K241R	+
1161/1162	E180V/S232G/K241R	+
1163/1164	I12V	+
1165/1166	S103L/K241R	+
1167/1168	I12V/V146I/K241R	+
1169/1170	I12V/S43M/S103V/V146I/S232G/K241R	+
1053/1054	T63P/S103L/E180L/S232G	+
1171/1172	I12V/E180V/K241R	+
1173/1174	I12V/S103V/V146I/K241R	+
1175/1176	I12V/S43Q/V146I/K241R	+
1177/1178	E180L/K241R	+
1179/1180	S43M/T63P/S103L/E180V/S232G/K241R	+
1181/1182	S103V/E180V/K241R	+
1183/1184	S43Q/S103L/E180V/S232G/K241R	+
1185/1186	S43M/S103L/E180V	+
1187/1188	I12V/E180V	+
1189/1190	R172C	+
1067/1068	S103L	+
1191/1192	Y118V	+
1193/1194	K119H	+

Table 5-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 860		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 860)	Improvement in Selectivity Relative to SEQ ID NO: 860
1195/1196	A145V	+
1045/1046	R172L	+
1197/1198	S122H	+
1199/1200	R212P	+
1081/1082	L18V	+
1201/1202	S211F	+
1203/1204	T144H	+
1205/1206	Y118L	+
1091/1092	R172V	+
1207/1208	N160L	+
1209/1210	I9V	+
1211/1212	A154R	+
Levels of increased selectivity (fraction of product that is the (S)-alcohol) were determined relative to the reference polypeptide of SEQ ID NO: 860 and defined as follows: "+" 0.56 to 0.62, "++" > 0.62, "+++" > 0.64		

EXAMPLE 6

Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 1120 for Improved Production of Compound (3)

[0276] The engineered polynucleotide (SEQ ID NO: 1119) encoding the polypeptide with nitroaldolase activity of SEQ ID NO: 1120 was used to generate the engineered polypeptides of Table 6-1. These polypeptides displayed improved nitroaldolase activity under the desired conditions e.g., the improvement in the formation of the β -nitro alcohol, compound (3), from the substrates trifluoroacetone and nitromethane, compounds (1) and (2), respectively, as compared to the starting polypeptide. Similarly, the engineered polypeptides of Table 6-2 displayed improved nitroaldolase selectivity towards the desired (S)-alcohol. The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the “backbone” amino acid sequence of SEQ ID NO: 1120, as described below.

[0277] Directed evolution began with the polynucleotide set forth in SEQ ID NO: 1119. Libraries of engineered polypeptides were generated using various well-known techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods that measured the polypeptide’s ability to produce compound (3).

[0278] The enzyme assay and analysis were carried using the method as described in EXAMPLE 3 except that the nitromethane (2) concentration was increased to 0.75 M and the pH was decreased to pH 5.5.

[0279] Hit variants were grown in 250 mL shake flasks, and enzyme powders were generated as described in EXAMPLE 1. The activity of the enzyme powders was evaluated as described in EXAMPLE 3 except the trifluoroacetone (1) concentration was increased to 0.525 M, the nitromethane concentration was increased to 0.75 M and the pH was decreased to pH 5.5; the hit variants are denoted with a * in Tables 6-1 and 6-2.

Table 6-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 1120		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1120)	Fold Improvement in Activity Relative to SEQ ID NO: 1120
1213/1214	I7V/F158Y	+++
1215/1216	S3D	+++
1217/1218	E55R	+++
1219/1220	Y118V/R172L/L180V	+++
1221/1222	K223L	+++
1223/1224	I7V/F158Y/I197V/V202I	+++
1225/1226	A154R	+++
1227/1228	V103L/A154R	+++
1229/1230	V103L	+++
1231/1232	V103L/T144V/V146I/R172L/L180V	+++
1233/1234	Q27E/S43I/R172L	+++
1235/1236	E59T	+++
1237/1238	L48I/R172L	+++
1239/1240	E55A	++
1241/1242	V146I	++
1243/1244	K226V	++
1245/1246	I7V	++
1247/1248	N32G	++
1249/1250	Y118V	++
1251/1252	Y118V/T144V/Q215H	++
1253/1254	A263G	++
1255/1256	V146I/A154R	++
1257/1258	V103L/V146I/A154R/R172L/L180V/S232G	++
1259/1260	P94Q/V103L/Y118V/V146I	++
1261/1262	E55S	++
1263/1264	I7V/F158Y/K175P	++
1265/1266	A29G	++
1267/1268	F158Y	++
1269/1270	T137M	++
1271/1272	R46V	++
1273/1274	R172L/T190S	++
1275/1276	G71S	++
1277/1278	A261G	++

Table 6-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 1120		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1120)	Fold Improvement in Activity Relative to SEQ ID NO: 1120
1279/1280	A263S	+
1281/1282	Q27E/S43I/A115S/R172L	+
1283/1284	K223M	+
1285/1286	K72G	+
1287/1288	C229W	+
1289/1290	S43I/I110T/A115S/R172L	+
1291/1292	K73R	+
1293/1294	R46H	+
1295/1296	K223S	+
1297/1298	R198L	+
1299/1300	E55G	+
1301/1302	A263L	+
1303/1304	K226L	+
1305/1306	T137R	+
1307/1308	A29R	+
1309/1310	E59G	+
1311/1312	K223C	+
1313/1314	L216M	+
1315/1316	K223E	+
1317/1318	A263E	+
1319/1320	N32R	+
1321/1322	K223A	+
1323/1324	C229L	+
1325/1326	L216V	+
1327/1328	T142V	+
1329/1330	S67N	+
1331/1332	E205G	+
1333/1334	Q247V	+
1335/1336	A261H	+
1337/1338	N44S	+
1339/1340	E205L	+
1341/1342	D184S	+
1343/1344	C229V	+
Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1120 and defined as follows: "+" 1.20 to 1.34, "++" > 1.34, "+++" > 1.47		

Table 6-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 1120		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1120)	Improvement in Selectivity Relative to SEQ ID NO: 1120
1345/1346*	S80G	+++
1347/1348*	S80A	+++
1349/1350	V103L/V146I/I209C	++
1351/1352	R172C/L180V/I209C	++
1353/1354	Y118V/K119H/R172C/I209C	++
1355/1356	T144V/I209C/S232G	+
1357/1358	V103L/R172C/I209C	+
1359/1360	N135V	+
1361/1362	Q181S	+
1363/1364	K223E	+
1365/1366	K73R	+
Levels of increased selectivity (fraction of product that is the (S)-alcohol) were determined relative to the reference polypeptide of SEQ ID NO: 1120 and defined as follows: "+" 0.84 to 0.85, "++" > 0.85, "+++" > 0.87		

EXAMPLE 7

Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 1348 for Improved Production of Compound (3)

[0280] The engineered polynucleotide (SEQ ID NO: 1347) encoding the polypeptide with nitroaldolase activity of SEQ ID NO: 1348 was used to generate the engineered polypeptides of Table 6-1. These polypeptides displayed improved nitroaldolase activity under the desired conditions e.g., the improvement in the formation of the β -nitro alcohol, compound (3), from the substrates trifluoroacetone and nitromethane, compounds (1) and (2), respectively, as compared to the starting polypeptide. Similarly, the engineered polypeptides of Table 6-2 displayed improved nitroaldolase selectivity towards the desired (S)-alcohol. The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the "backbone" amino acid sequence of SEQ ID NO: 1348, as described below.

[0281] Directed evolution began with the polynucleotide set forth in SEQ ID NO: 1347. Libraries of engineered polypeptides were generated using various well-known techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods that measured the polypeptide's ability to produce compound (3).

[0282] The enzyme assay and analysis were carried using the method as described in EXAMPLE 6 except that the trifluoroacetone (1) concentration was increased to 0.525 M and the reaction time was decreased to 6 hours.

[0283] Hit variants were grown in 250 mL shake flasks, and enzyme powders were generated as described in EXAMPLE 1. The activity of the enzyme powders was evaluated as described in EXAMPLE 6 except that the reaction time was decreased to 6 hours; the hit variants are denoted with a * in Tables 7-1 and 7-2.

Table 7-1. Nitroaldolase Activity for the Production of Compound (3) Relative to SEQ ID NO: 1348		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1348)	Fold Improvement in Activity Relative to SEQ ID NO: 1348
1367/1368*	W39S	+++
1369/1370*	W39T	+++
1371/1372*	R46H/E55R/A263L	+++
1373/1374*	R46H/E70N/K175P/A263G	+++
1375/1376	R46H/E55R/Y118V	+++
1377/1378	A105G	+++
1379/1380	W39V	++
1381/1382	Y118V/K226V/A261G/A263I	++
1383/1384	R46H/Y118V	++
1385/1386	W39Q	++
1387/1388	R46H/E55R	++
1389/1390	R46H/E55R/K72C	++
1391/1392	E55R/A263G	++
1393/1394	R46H/K72C	++
1395/1396	R46H/E55R/K226V/S257Q/A263L	+
1397/1398	W39F	+
1399/1400	R46H/E70N	+
1401/1402	V2G	+
1403/1404	W39M	+
1405/1406	R46H/E55R/K92A	+
1407/1408	R46H/E70N/K92A/Y118V/A261G/A263L	+
1409/1410	I7V/R46H	+
1411/1412	R46H	+
1413/1414	Y118V	+
1415/1416	A263G	+
1417/1418	E55R/A261G/A263I	+
1419/1420	W39G	+
1421/1422	E55R/S257Q/A261G	+
Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1348 and defined as follows: "+" 1.50 to 1.73, "++" > 1.73, "+++" > 2.01		

Table 7-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 1348		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1348)	Improvement in Selectivity Relative to SEQ ID NO: 1348
1423/1424*	R46H/E55R/Y118V/A263S	+++
1371/1372*	R46H/E55R/A263L	+++
1381/1382	Y118V/K226V/A261G/A263I	+++
1375/1376	R46H/E55R/Y118V	+++
1413/1414	Y118V	+++
1425/1426	W39S	+++
1427/1428	A105G	+++
1383/1384	R46H/Y118V	+++
1373/1374*	R46H/E70N/K175P/A263G	+++
1429/1430	W39T	+++
1401/1402	V2G	++
1421/1422	E55R/S257Q/A261G	++
1411/1412	R46H	++
1431/1432	R46H/K92A	++
1433/1434	R46H/V146I	++
1435/1436	L22V/R46H/E55R/Y118V	++
1437/1438	E55R/S257Q/A263L	++
1439/1440	Q27G/Y118V	++
1441/1442	R46H/E55R/V146I/V202I	++
1409/1410	I7V/R46H	++
1443/1444	I7V/R46H/K175P	++
1399/1400	R46H/E70N	++
1391/1392	E55R/A263G	++
1445/1446	E70N	++
1447/1448	V146I/A261G/A263L	++
1379/1380	W39V	++
1449/1450	R46H/E55R	+
1403/1404	W39M	+
1385/1386	W39Q	+
1451/1452	R46H/E55R/E59T/A261G/A263L	+
1415/1416	A263G	+
1453/1454	A261G/A263I	+
1405/1406	R46H/E55R/K92A	+
1455/1456	V202I	+
1389/1390	R46H/E55R/K72C	+
1457/1458	E55R/E70N/Y118V	+
1459/1460	R46H/K92A/K175P	+
1461/1462	R46H/K72C	+
1463/1464	S257Q/A261G/A263G	+
1465/1466	V146I/S257Q/A263S	+

Table 7-2. Nitroaldolase Selectivity for the Production of Compound (3) Relative to SEQ ID NO: 1348		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1348)	Improvement in Selectivity Relative to SEQ ID NO: 1348
1467/1468	R46H/E55R/V146I	+
1469/1470	K92A/Y118V	+
1417/1418	E55R/A261G/A263I	+
1471/1472	Q27G/R46H/V146I/F158Y/V202I	+
1473/1474	R46H/V146I/S257Q/A263I	+
1407/1408	R46H/E70N/K92A/Y118V/A261G/A263L	+
1475/1476	Q27G/R46H/K72C	+
1395/1396	R46H/E55R/K226V/S257Q/A263L	+
1477/1478	A29V/E70N/Y118V	+
1479/1480	E192V/S257Q/A261G	+
1481/1482	Q27G/R46H/K175P	+
1483/1484*	W39V	+
Levels of increased selectivity (fraction of product that is the (S)-alcohol) were determined relative to the reference polypeptide of SEQ ID NO: 1348 and defined as follows: "+" 0.92 to 0.94, "++" > 0.94, "+++" > 0.95		

EXAMPLE 8

Analytical Detection of Conversion of Compounds (1) and (2) to Compound (3)

[0284] Data described in EXAMPLE 2 through EXAMPLE 4 were collected using the analytical methods provided in Table 8-1 and Table 8-2. The methods provided herein find use in analyzing the variants produced using the present invention. However, it is not intended that present invention be limited to the methods described herein, as there are other suitable methods known in the art that are applicable to the analysis of the variants provided herein and/or produced using the methods provided herein.

Table 8-1. GC Parameters	
Method for Determining Conversion of Compound (1) and (2) to Compound (3)	
Instrument	Agilent 6890 GC
Column	Supelco Astec ChiralDEX B-DM, 30 m x 0.25 mm x 0.12 μ m film (77023AST)
Carrier gas and flow rate	Helium, 3.0 mL/min (constant flow control)
Column temperature	100°C, isothermal
Detector and temperature	FID, 250°C
Inlet temperature and split ratio	250°C, 20:1
Injection volume	1 μ L
Product retention time	(S)-alcohol: 2.71 min (R)-alcohol: 2.81 min

Table 8-2. GC Parameters	
Method for Determining Conversion of Compound (1) and (2) to Compound (3)	
Instrument	Agilent 6890 GC
Column	Restek β DEX _{sm} , 30 m x 0.25 mm x 0.25 μ m film (13105)
Carrier gas and flow rate	Helium, 3.0 mL/min (constant flow control)
Column temperature	120°C, isothermal
Detector and temperature	FID, 250°C
Inlet temperature and split ratio	250°C, 20:1
Injection volume	1 μ L
Product retention time	(S)-alcohol: 2.87 min (R)-alcohol: 2.93 min

[0285] Data described in EXAMPLE 5 were collected using the analytical methods provided in Table 8-3 and Table 8-4. Data described in EXAMPLE 6 and EXAMPLE 7 were collected using the analytical method provided in Table 8-4. The methods provided herein find use in analyzing the variants produced using the present invention. However, it is not intended that present invention be limited to the methods described herein, as there are other suitable methods known in the art that are applicable to the analysis of the variants provided herein and/or produced using the methods provided herein.

Table 8-3. GC Parameters	
Method for Determining Conversion of Compound (1) and (2) to Compound (3)	
Instrument	Agilent 6890 GC
Column	Supelco β -DEX 325, 30 m x 0.25 mm x 0.25 μ m film, (24308)
Carrier gas and flow rate	Helium, 10.0 mL/min (constant flow control)
Column temperature	100°C, isothermal
Detector and temperature	FID, 250°C
Inlet temperature and split ratio	250°C, 60:1
Injection volume	1 μ L
Product retention time	(S)-alcohol: 1.80 min (R)-alcohol: 1.86 min

Table 8-4. GC Parameters	
Method for Determining Conversion of Compound (1) and (2) to Compound (3)	
Instrument	Agilent 6890 GC
Column	Astec CHIRALDEX G-TA, 30 m x 0.25 mm x 0.12 μ m film, (73033AST)
Carrier gas and flow rate	Helium, 10.0 mL/min (constant flow control)
Column temperature	130°C, isothermal

Table 8-4. GC Parameters	
Method for Determining Conversion of Compound (1) and (2) to Compound (3)	
Detector and temperature	FID, 250°C
Inlet temperature and split ratio	250°C, 60:1
Injection volume	1 µL
Product retention time	(S)-alcohol: 1.16 min (R)-alcohol: 1.10 min

[0286] All publications, patents, patent applications and other documents cited in this application are hereby incorporated by reference in their entireties for all purposes to the same extent as if each individual publication, patent, patent application or other document were individually indicated to be incorporated by reference for all purposes.

[0287] While various specific embodiments have been illustrated and described, it will be appreciated that various changes can be made without departing from the spirit and scope of the invention(s).

CLAIMS

What is claimed is:

1. An engineered nitroaldolase polypeptide comprising a polypeptide sequence comprising at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NOs: 2, 24, 496, 860, 1120, and/or 1348, or a functional fragment thereof, wherein said engineered nitroaldolase polypeptide comprises at least one substitution or substitution set in said polypeptide sequence, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NOs: 2, 24, 496, 860, 1120, and/or 1348.

2. The engineered nitroaldolase polypeptide of Claim 1, wherein said polypeptide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 2, wherein said engineered nitroaldolase polypeptide comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 11, 7/9, 7/9/33/68/72/158, 7/9/33/72/175/197/230, 7/9/33/72/230, 7/9/68/72/197, 7/9/158/175, 7/9/175, 7/9/197, 7/33, 7/68/72/175, 7/72, 7/72/97/175/230, 9, 9/11/33/72/197/230, 9/33/68/158/197, 9/33/72/230, 9/33/97/230, 9/68/72, 9/97/175/202/218, 9/175, 9/202, 11, 13, 14/33/68/72/97, 18, 18/33/68/97/158/175/239, 18/33/72/97/158/175, 18/33/72/97/158/239, 18/33/72/97/175/202, 18/33/72/175/202/237/239, 18/33/72/197/237, 18/68/72, 18/68/72/158/175/237/239, 18/68/72/175/202/239, 18/68/72/239, 18/68/175/223/237/239, 18/68/239, 18/72/97/175/237/239, 18/72/175/197, 18/72/237, 18/72/239, 18/97/175/237/239, 18/97/223/237/239, 18/175, 18/197/239, 18/239, 20/124, 33/68/72, 33/68/72/175/230, 33/68/223/239, 33/72, 33/72/97/197, 33/72/175, 33/175, 39, 44/97, 52, 68/72, 68/72/97, 68/72/97/197, 68/72/197, 68/72/197/239, 68/97/175, 68/97/175/197/223, 72/97/158/223/237, 72/175/197, 97/158/175/197, 97/175, 103, 121, 122, 124, 125, 128, 133, 148, 175, 197, and 197/202, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 2.

3. The engineered nitroaldolase polypeptide of Claim 1, wherein said polypeptide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 2, wherein said engineered nitroaldolase polypeptide comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 7/11/33/72/97, 7/11/33/97, 7/11/72, 7/11/158, 7/11/237, 7/33/72/97/158/175/197/237/239, 9/11/33, 9/11/33/72, 9/11/72, 9/33/68/97/237/239, 9/72/97/158/237/239, 9/218/239, 10, 11, 12, 13, 14, 18, 33/68/72/97/237/239, 33/72/158/239, 39, 40, 41, 48, 49, 51, 54, 54/126, 57, 68/175/239, 82, 84, 85, 86, 103, 104, 105, 107, 115, 117, 119, 119/120, 120, 122, 123, 124, 127, 130, 131, 133, 134, 137/145, 144, 145, 146, 147, 148, 149, 150, 152, 157, 158, 159, 169, 173, 174, 175, 180, 181, 183, 203, 204, 208, 209,

210, 215, 218, 236, and 237, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 2.

4. The engineered nitroaldolase polypeptide of Claim 1, wherein said polypeptide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 24, wherein said engineered nitroaldolase polypeptide comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 9/39/122/157, 18, 18/33/39/230, 18/33/175, 18/39/44/118, 18/175, 20, 33/39, 33/39/44/118/121, 33/39/44/121, 33/39/44/152, 33/39/44/175, 33/39/44/197, 33/39/44/210/230, 33/39/118/121/175, 33/39/121/175, 33/39/197/210, 33/39/210, 33/39/210/230, 33/118/121, 33/118/121/175, 39, 39/44/103, 39/44/118, 39/44/152/230, 39/44/175/197/230, 39/103, 39/103/125, 39/103/125/127/146, 39/103/125/127/146/150, 39/103/125/146, 39/103/127/150, 39/103/150, 39/118/121, 39/197, 39/197/210, 43, 44/103, 81, 103, 103/125/127, 103/125/146, 103/127, 103/150, 118/121, 118/121/175, 118/121/197, 121, 152, 152/197, 154, 159, 163, 175, 210, 232, and 238, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 24.

5. The engineered nitroaldolase polypeptide of Claim 1, wherein said polypeptide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO: 24, wherein said engineered nitroaldolase polypeptide comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 7, 9/13, 9/13/39, 9/13/39/68/72/239, 9/13/39/68/239, 9/13/39/72/122, 9/13/39/97/239, 9/13/68/72/97/239, 9/13/68/97/134/239, 9/13/68/97/239, 9/13/72, 9/13/134, 11, 12/42/125/127, 12/103, 12/103/150, 13, 13/39, 13/39/68, 13/39/68/72/97, 13/39/72/97/122/134/223, 13/39/97/239, 13/39/239, 13/68, 13/68/72/97/122/239, 16, 18/33/39/44/118/121/175/197/210, 18/33/39/44/121, 18/33/39/112/118/121/197, 18/33/39/121, 18/33/39/121/175, 18/33/39/121/175/197, 18/33/39/121/175/210, 18/33/39/121/210, 18/33/118/121, 18/33/121, 18/33/121/197, 18/33/121/210/230, 18/39/44/118/121/230, 18/39/44/121/175/210, 18/39/118/121, 18/39/118/121/210, 18/39/121/175, 18/39/121/175/210, 18/39/121/197/210, 18/118/121, 18/118/121/175, 18/118/121/210, 18/118/121/230, 18/121, 18/121/175, 18/121/210, 18/121/230, 20, 21, 33/39/44/118/121, 33/39/44/118/121/210/230, 33/39/44/121, 33/39/44/121/152, 33/39/44/121/175/210, 33/39/44/121/230, 33/39/73/121/197, 33/39/118/121/175, 33/39/118/121/175/230, 33/39/121, 33/39/121/175, 33/39/121/210/230, 33/118/121, 33/118/121/175, 33/118/121/197, 33/121, 33/121/230, 37/91, 39/44/118/121/175/210, 39/68/72/239, 39/118/121, 39/118/121/210, 39/121, 39/121/210/230, 43, 44, 76, 79, 101, 118/121, 118/121/175, 118/121/197, 121, 121/175/230, 121/210, 148, 153, 159, 161, 172, 200, 207, 233, 245, and 248, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 24.

6. The engineered nitroaldolase polypeptide of Claim 1, wherein said polypeptide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 496, wherein said engineered nitroaldolase polypeptide comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 43/103, 43/103/172/238, 43/103/238/241, 43/238, 81/232, 103, 103/159/238, 103/194/238, 103/238, 122, 232, 238, and 238/241, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 496.

7. The engineered nitroaldolase polypeptide of Claim 1, wherein said polypeptide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 496, wherein said engineered nitroaldolase polypeptide comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 12, 43, 43/47/103, 43/47/103/232, 43/103, 43/103/121/172/241, 43/103/172, 43/103/172/238, 43/103/238, 43/103/238/241, 43/121, 43/194/238, 43/238, 44/103/121/238, 44/103/238, 44/103/238/241, 44/121/238/241, 47/103, 63, 103, 103/121, 103/121/172/238, 103/159/238, 103/172/238, 103/194/238, 103/238, 122, 131, 146, 178, 180, and 221, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 496.

8. The engineered nitroaldolase polypeptide of Claim 1, wherein said polypeptide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 860, wherein said engineered nitroaldolase polypeptide comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 12, 12/43/63/103/146/180/232/241, 12/43/63/103/146/241, 12/43/81/103/241, 12/43/81/146/232/241, 12/43/81/180/241, 12/43/81/232/241, 12/43/81/241, 12/43/103, 12/43/103/146/180/232/241, 12/43/103/146/180/241, 12/43/103/180/241, 12/43/103/241, 12/43/146/180/232/241, 12/43/180/241, 12/63/81/103/241, 12/81, 12/81/180/241, 12/103/146/241, 12/103/180/241, 12/146, 12/146/180/241, 12/180/232/241, 12/241, 18, 43/81/232/241, 43/103/146/180, 43/103/180/241, 43/103/241, 43/180/232, 43/232/241, 63/103/180/232, 78, 81, 81/103/146, 81/146/180/241, 81/146/241, 81/180/241, 84, 103, 103/146, 103/146/180, 103/146/180/241, 103/180, 103/180/241, 103/261, 144, 145, 146/180/232, 146/180/241, 146/241, 150, 166, 172, 180/232, 232/241, and 241, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 860.

9. The engineered nitroaldolase polypeptide of Claim 1, wherein said polypeptide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 860, wherein said engineered nitroaldolase polypeptide comprises at least one substitution or substitution set in said polypeptide sequence at one or more

positions selected from 9, 12, 12/43/63/103/146/180/232/241, 12/43/63/103/146/241, 12/43/103/146/180/232/241, 12/43/103/146/180/241, 12/43/103/146/232/241, 12/43/103/146/241, 12/43/103/180/241, 12/43/146/180/232/241, 12/43/146/241, 12/43/180/232, 12/43/180/241, 12/63/103/146/232/241, 12/103/146/241, 12/103/180/241, 12/146, 12/146/180/241, 12/146/232/241, 12/146/241, 12/180, 12/180/232/241, 12/180/241, 12/241, 18, 43/63/103/180/232/241, 43/103/146/180, 43/103/180, 43/103/180/232/241, 43/103/180/241, 43/180/232, 63/103/180/232, 103, 103/146/180, 103/146/180/241, 103/180, 103/180/241, 103/241, 118, 119, 122, 144, 145, 146/180/241, 154, 160, 172, 180/232, 180/232/241, 180/241, 208, 211, and 212, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 860.

10. The engineered nitroaldolase polypeptide of Claim 1, wherein said polypeptide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 1120, wherein said engineered nitroaldolase polypeptide comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 3, 7, 7/158, 7/158/175, 7/158/197/202, 27/43/115/172, 27/43/172, 29, 32, 43/110/115/172, 44, 46, 48/172, 55, 59, 67, 71, 72, 73, 94/103/118/146, 103, 103/144/146/172/180, 103/146/154/172/180/232, 103/154, 118, 118/144/215, 118/172/180, 137, 142, 146, 146/154, 154, 158, 172/190, 184, 198, 205, 216, 223, 226, 229, 247, 261, and 263, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 1120.

11. The engineered nitroaldolase polypeptide of Claim 1, wherein said polypeptide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 1120, wherein said engineered nitroaldolase polypeptide comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 73, 80, 103/146/209, 103/172/209, 118/119/172/209, 135, 144/209/232, 172/180/209, 181, and 223, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 1120.

12. The engineered nitroaldolase polypeptide of Claim 1, wherein said polypeptide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 1348, wherein said engineered nitroaldolase polypeptide comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 2, 7/46, 39, 46, 46/55, 46/55/72, 46/55/92, 46/55/118, 46/55/226/257/263, 46/55/263, 46/70, 46/70/92/118/261/263, 46/70/175/263, 46/72, 46/118, 55/257/261, 55/261/263, 55/263, 105, 118, 118/226/261/263, and 263, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 1348.

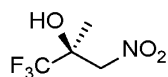
13. The engineered nitroaldolase polypeptide of Claim 1, wherein said polypeptide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 1348, wherein said engineered nitroaldolase polypeptide comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 2, 7/46, 7/46/175, 22/46/55/118, 27/46/72, 27/46/146/158/202, 27/46/175, 27/118, 29/70/118, 39, 46, 46/55, 46/55/59/261/263, 46/55/72, 46/55/92, 46/55/118, 46/55/118/263, 46/55/146, 46/55/146/202, 46/55/226/257/263, 46/55/263, 46/70, 46/70/92/118/261/263, 46/70/175/263, 46/72, 46/92, 46/92/175, 46/118, 46/146, 46/146/257/263, 55/70/118, 55/257/261, 55/257/263, 55/261/263, 55/263, 70, 92/118, 105, 118, 118/226/261/263, 146/257/263, 146/261/263, 192/257/261, 202, 257/261/263, 261/263, and 263, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 1348.

14. The engineered nitroaldolase polypeptide of any of Claims 1-13, wherein said engineered nitroaldolase polypeptide comprises an amino acid sequence with at least 80% sequence identity to any even-numbered sequence set forth in SEQ ID NOs: 4-1484.

15. The engineered nitroaldolase polypeptide of any of Claims 1-14, wherein said engineered nitroaldolase polypeptide comprises a polypeptide sequence set forth in the even numbered sequences of SEQ ID NOs: 4-1484.

16. The engineered nitroaldolase polypeptide of any of Claims 1-15, wherein said engineered nitroaldolase polypeptide comprises a polypeptide sequence that exhibits at least one improved property compared to the engineered nitroaldolase polypeptide of SEQ ID NO: 2.

17. The engineered nitroaldolase polypeptide of Claim 16, wherein said improved property comprises improved production of compound (3)



Compound (3)

18. The engineered nitroaldolase polypeptide of Claim 16, wherein said improved property comprises improved stereoselectivity towards the (*S*)-product of at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or greater.

19. The engineered nitroaldolase polypeptide of Claim 16, wherein said improved property comprises improved activity on the substrates of trifluoroacetone, compound (1), and nitromethane, compound (2).

20. The engineered nitroaldolase polypeptide of any of Claims 1-19, wherein said engineered nitroaldolase polypeptide is purified.

21. A composition comprising at least one engineered nitroaldolase polypeptide provided in any of Claims 1-20.

22. An engineered polynucleotide encoding at least one engineered nitroaldolase polypeptide of any of Claims 1-20.

23. An engineered polynucleotide sequence encoding at least one engineered polypeptide, wherein said polypeptide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NOs: 2, 24, 496, 860, 1120, and/or 1348, wherein the sequence of said polypeptide comprises at least one substitution at one or more positions.

24. The engineered polynucleotide sequence of Claim 22 or 23, wherein said polynucleotide sequence comprises at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NOs: 1, 23, 495, 860, 1119, and/or 1347.

25. The engineered polynucleotide sequence of any of Claims 22-24, wherein said polynucleotide sequence comprises SEQ ID NO: 3, 23, 495, 859, 1119, or 1347.

26. An engineered polynucleotide comprising the odd-numbered sequences set forth in SEQ ID NOs: 3-1483.

27. A vector comprising the engineered polynucleotide of any of Claims 22 to 26.

28. The vector of Claim 27, further comprising at least one control sequence.

29. A host cell comprising the vector of Claim 27 and/or 28.
30. A host cell, wherein said host cell produces at least one engineered nitroaldolase polypeptide of any of Claims 1-20.
31. A method of producing an engineered nitroaldolase polypeptide in a host cell, comprising culturing the host cell of Claim 29 and/or 30, in a culture medium under suitable conditions, such that at least one engineered nitroaldolase polypeptide is produced.
32. The method of Claim 31, further comprising the step of recovering said engineered nitroaldolase polypeptide.
33. The method of Claim 31 and/or 32, further comprising the step of purifying said at least one engineered nitroaldolase polypeptide.