Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3030163 A1 2018/01/18

(21) 3 030 163

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2017/07/10
(87) Date publication PCT/PCT Publication Date: 2018/01/18
(85) Entrée phase nationale/National Entry: 2019/01/07
(86) N° demande PCT/PCT Application No.: US 2017/041373
(87) N° publication PCT/PCT Publication No.: 2018/013481
(30) Priorité/Priority: 2016/07/11 (US62/360,745)

(51) Cl.Int./Int.Cl. *C12N 9/00* (2006.01), *C12N 15/11* (2006.01), *C12N 15/70* (2006.01), *C12N 9/86* (2006.01), *C12P 21/02* (2006.01)

(71) Demandeur/Applicant:
BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM, US

(72) Inventeurs/Inventors:
ELLINGTON, ANDREW D., US;
THYER, ROSS, US

(74) Agent: BERESKIN & PARR LLP/S.E.N.C.R.L.,S.R.L.

(54) Titre : POLYPEPTIDES DE RECOMBINAISON COMPRENANT DE LA SELENOCYSTEINE ET LEUR PROCEDE DE PRODUCTION

(54) Title: RECOMBINANT POLYPEPTIDES COMPRISING SELENOCYSTEINE AND METHOD FOR PRODUCING THE SAME

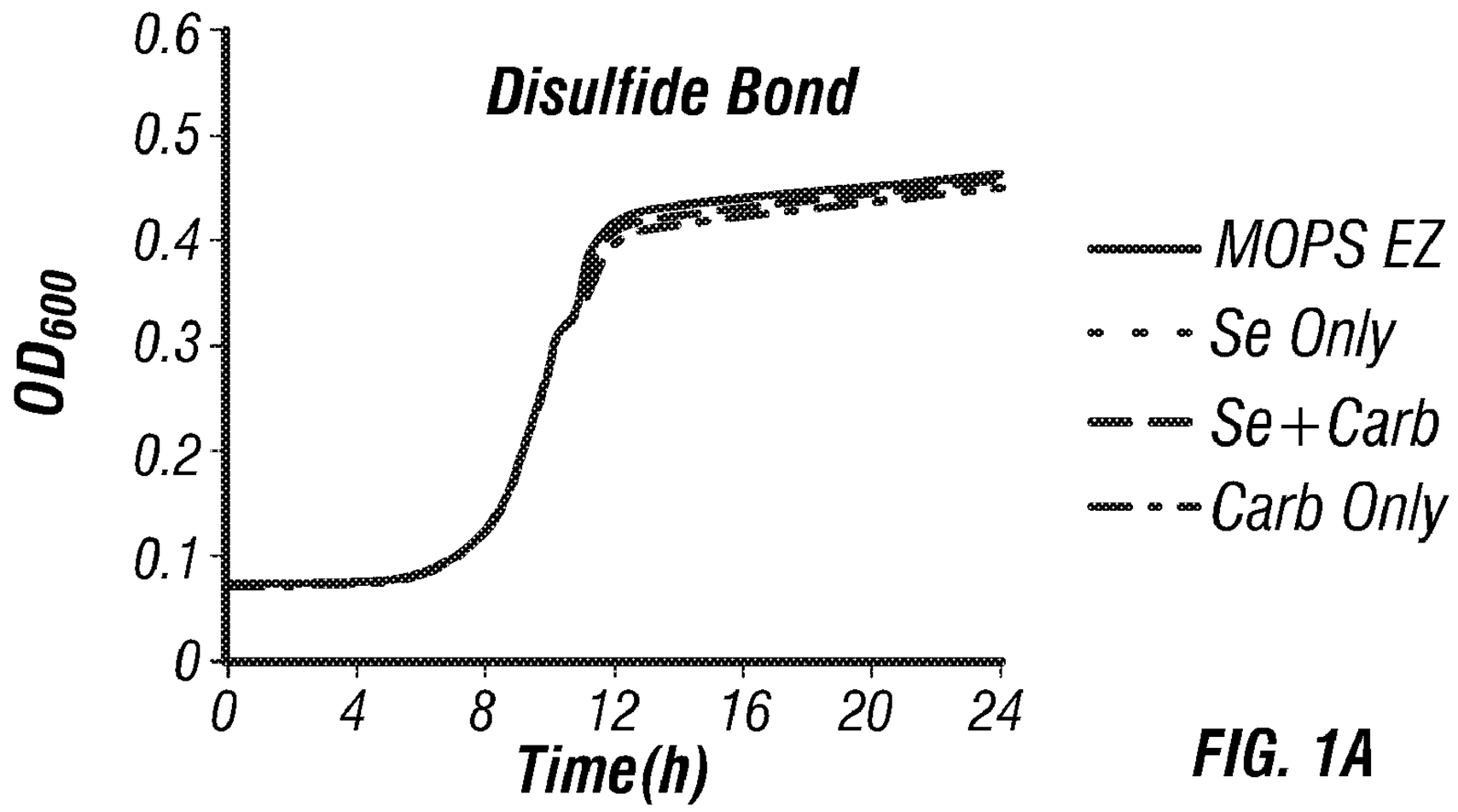

FIG. 1A

(57) Abrégé/Abstract:
Composition comprising purified recombinant selenoproteins, such an antibodies and enzymes, are provided. Method of producing such recombinant polypeptides and bacterial strains for the same are likewise provided.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 January 2018 (18.01.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/013481 A1

(51) International Patent Classification:
C12N 9/00 (2006.01) *C12N 15/70* (2006.01)
C12N 9/86 (2006.01) *C12P 21/02* (2006.01)
C12N 15/11 (2006.01)

(21) International Application Number:
PCT/US2017/041373

(22) International Filing Date:
10 July 2017 (10.07.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/360,745 11 July 2016 (11.07.2016) US

(71) Applicant: **BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM** [US/US]; 201 West 7th Street, Austin, TX 78701 (US).

(72) Inventors: **ELLINGTON, Andrew D.**; 2106 Elton Lane, Austin, TX 78705 (US). **THYER, Ross**; c/o University of Texas at Austin, 2500 Speedway MBB 3.424, Austin, TX 78712 (US).

(74) Agent: **BYRD, Marshall P.**; Parker Highlander PLLC, 1120 S. Capital of Texas Highway, Building One, Suite 200, Austin, TX 78746 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: RECOMBINANT POLYPEPTIDES COMPRISING SELENOCYSTEINE AND METHOD FOR PRODUCING THE SAME

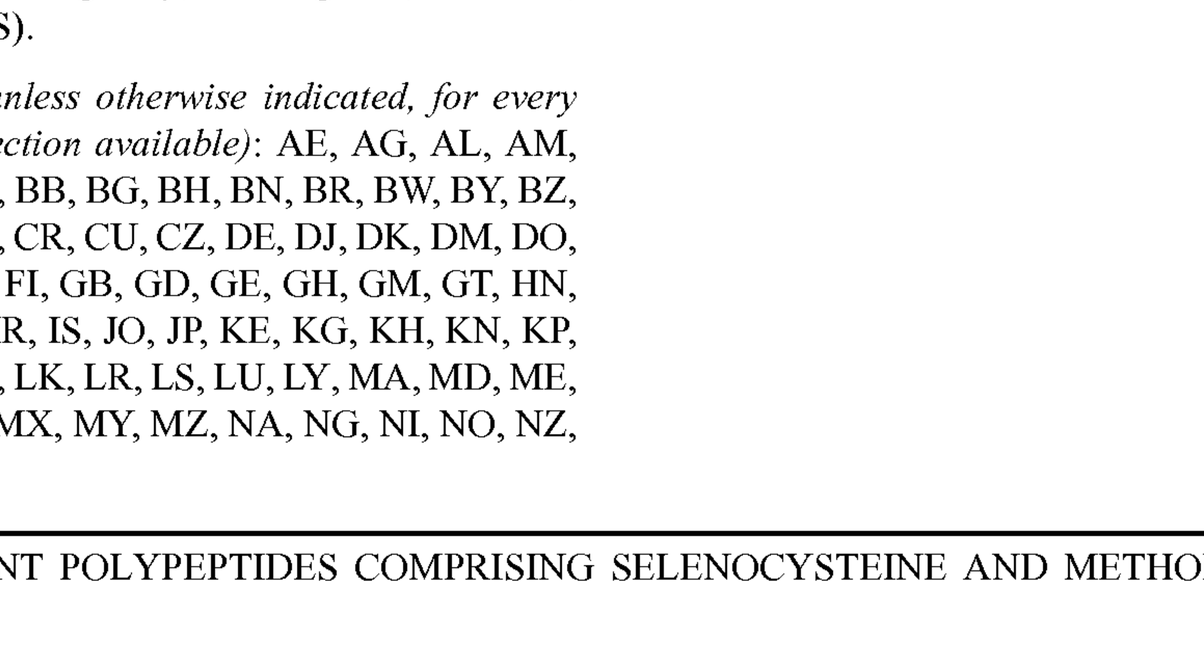

FIG. 1A

(57) Abstract: Composition comprising purified recombinant selenoproteins, such an antibodies and enzymes, are provided. Method of producing such recombinant polypeptides and bacterial strains for the same are likewise provided.

WO 2018/013481 A1

DESCRIPTION

RECOMBINANT POLYPEPTIDES COMPRISING SELENOCYSTEINE AND METHOD FOR PRODUCING THE SAME

[0001] This application claims the benefit of United States Provisional Patent Application No. 62/360,745, filed July 11, 2016, the entirety of which is incorporated herein by reference.

[0002] This invention was made with government support under Grant No. CHE1402753 awarded by the National Science Foundation. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0003] The present invention relates generally to the field of molecular biology. More particularly, it concerns polypeptides comprising non-canonical amino acids.

2. Description of Related Art

[0004] Selenium is used by certain organisms for the expression of selenoproteins. Selenoproteins are a unique group of polypeptides that are found in both prokaryotes and eukaryotes and contain the non-canonical amino acid, selenocysteine. Selenocysteine has a significantly lower pK_a than cysteine (5.2 vs 8.5 for free amino acid) and much stronger nucleophilic properties, making it an attractive target for altering protein chemistry and function. Unfortunately, most selenoproteins have proven difficult to produce in *E. coli*, the standard host for recombinant protein production. This is due to the inherently low efficiency of selenocysteine incorporation in bacteria (4–5% vs termination of protein synthesis). In addition, the requirement that the SECIS element immediately follows the UGA codon, forming part of the coding sequence, greatly limits which proteins are amenable to selenocysteine insertion.

[0005] Recently, an evolved *E. coli* $tRNA^{Sec}$ that is compatible with the canonical translation machinery and can suppress amber stop codons to incorporate selenocysteine with high efficiency was also developed. However, there is an unmet need for systems allow for efficient incorporate selenocysteine to produce commercially relevant amounts of recombinant selenoproteins.

SUMMARY OF THE INVENTION

[0006] A first embodiment of the present disclosure provides a composition comprising purified recombinant polypeptides, said polypeptides comprising at least one selenocysteine residue at a selected position not found in a wild type version of the polypeptide, wherein at least 80% of the recombinant polypeptides in the composition comprise the selenocysteine residue at the selected position. In certain aspects, the recombinant polypeptides comprise an antibody or an enzyme. In certain aspects, the composition comprises at least 10 μg, 50 μg, 100 μg, 500 μg or 1 mg of the purified recombinant polypeptides. In some aspects, the purified recombinant polypeptides are 95%-99.9% pure, *e.g.*, at least about 95%, 96%, 97%, 98%, 99% or 99.5% pure.

[0007] In some aspects, 80%-99.9% of the recombinant polypeptides in the composition comprise the selenocysteine residue at the selected position. In certain aspects, 90%-99% of the recombinant polypeptides in the composition comprise the selenocysteine residue at the selected position. In some aspects, at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% of the recombinant polypeptides in the composition comprise the selenocysteine residue at the selected position. In still further aspects, a composition comprises recombinant polypeptides having at least two selenocysteine residues at selected positions and 80%-99.9% (*e.g.*, at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%) of the recombinant polypeptides in the composition comprise both of the selenocysteine residues at the selected positions. In yet further aspects, a composition comprises recombinant polypeptides having at least two selenocysteine residues at selected positions that form a diselenide bond and 80%-99.9% (*e.g.*, at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%) of the recombinant polypeptides in the composition comprise the diselenide bond between the selenocysteine residues at the selected positions.

[0008] In some aspects, the recombinant polypeptides are at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to a human polypeptide. In certain aspects, the human polypeptide is a polypeptide involved in a disease. In some aspects, the human polypeptide is an enzyme, a chemokine, a cytokine, an antibody or T-cell receptor. In some aspects, the antibody is an aglycosylated antibody. In still further aspects, the human polypeptide is a polypeptide that comprises a disulfide bond and the recombinant polypeptide comprises a diselenide bond in place of the disulfide bond.

[0009] In certain aspects, the polypeptide comprises 2, 3, 4, 5, 6, 7, 8, 9 or 10 selenocysteine residues at selected positions. In further aspects, at least about 80%-99.9% (*e.g.*, at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%) of the recombinant polypeptides in the composition comprise selenocysteine residues at each of the selected positions. In still further aspects, the polypeptides comprise at least two selenocysteine residues at selected positions. In certain aspects, the two selenocysteine residues at the selected positions form a diselenide bond.

[0010] A further embodiment provides a pharmaceutical composition comprising a composition of the embodiments comprising purified recombinant polypeptides, said polypeptides comprising at least one selenocysteine residue at a selected position not found in a wild type version of the polypeptide, wherein at least 80%-99.9% (*e.g.*, at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%) of the recombinant polypeptides in the composition comprise the selenocysteine residue at the selected position.

[0011] Further embodiments provide a method of treating a subject comprising administering an effective amount of a pharmaceutical composition of the embodiments comprising purified recombinant polypeptides, said polypeptides comprising at least one selenocysteine residue at a selected position not found in a wild type version of the polypeptide, wherein at least 80%-99.9% (*e.g.*, at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%) of the recombinant polypeptides in the composition comprise the selenocysteine residue at the selected position.

[0012] In yet a further embodiment, there is provided a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to: (a) SEQ ID NO:1 and having an amino acid substitution or deletion at a position corresponding to position 344 of SEQ ID NO:1; (b) SEQ ID NO:2 and having an amino acid substitution or deletion at a position corresponding to position 702 of SEQ ID NO:2; (c) SEQ ID NO:3 and having an amino acid substitution or deletion at a position corresponding to position 655 of SEQ ID NO:3; (d) SEQ ID NO:4 and having an amino acid substitution or deletion at a position corresponding to position 73 of SEQ ID NO:4; (e) SEQ ID NO:5 and having an amino acid substitution or deletion at a position corresponding to position 781 of SEQ ID NO:5; (f) SEQ ID NO:6 and having an amino acid substitution or deletion at a position corresponding to position 136 of SEQ ID NO:6; (g) SEQ ID NO:7 and having an amino acid substitution or deletion at a position

corresponding to position 183 of SEQ ID NO:7; (h) SEQ ID NO:8 and having an amino acid substitution or deletion at a position corresponding to position 1 of SEQ ID NO:8; (i) SEQ ID NO:9 and having an amino acid substitution or deletion at a position corresponding to position 102 of SEQ ID NO:9; (j) SEQ ID NO:10 and having an amino acid substitution or deletion at a position corresponding to position 105 of SEQ ID NO:10; (k) SEQ ID NO:11 and having an amino acid substitution or deletion at a position corresponding to position 673 of SEQ ID NO:11; (l) SEQ ID NO:12 and having an amino acid substitution or deletion at a position corresponding to position 69 of SEQ ID NO:12; (m) SEQ ID NO:13 and having an amino acid substitution or deletion at a position corresponding to position 107 of SEQ ID NO:13; (n) SEQ ID NO:17 and having an amino acid substitution or deletion at a position corresponding to position 246 of SEQ ID NO:17; (o) SEQ ID NO:18 and having an amino acid substitution or deletion at a position corresponding to position 545 of SEQ ID NO:18; and/or (p) SEQ ID NO:19 and having an amino acid substitution or deletion at a position corresponding to position 124 of SEQ ID NO:19.

[0013] A further embodiment provides a recombinant polypeptide comprising an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to: (a) SEQ ID NO:1 and having an amino acid substitution or deletion at a position corresponding to position 344 of SEQ ID NO:1; (b) SEQ ID NO:2 and having an amino acid substitution or deletion at a position corresponding to position 702 of SEQ ID NO:2; (c) SEQ ID NO:3 and having an amino acid substitution or deletion at a position corresponding to position 655 of SEQ ID NO:3; (d) SEQ ID NO:4 and having an amino acid substitution or deletion at a position corresponding to position 73 of SEQ ID NO:4; (e) SEQ ID NO:5 and having an amino acid substitution or deletion at a position corresponding to position 781 of SEQ ID NO:5; (f) SEQ ID NO:6 and having an amino acid substitution or deletion at a position corresponding to position 136 of SEQ ID NO:6; (g) SEQ ID NO:7 and having an amino acid substitution or deletion at a position corresponding to position 183 of SEQ ID NO:7; (h) SEQ ID NO:8 and having an amino acid substitution or deletion at a position corresponding to position 1 of SEQ ID NO:8; (i) SEQ ID NO:9 and having an amino acid substitution or deletion at a position corresponding to position 102 of SEQ ID NO:9; (j) SEQ ID NO:10 and having an amino acid substitution or deletion at a position corresponding to position 105 of SEQ ID NO:10; (k) SEQ ID NO:11 and having an amino acid substitution or deletion at a position corresponding to position 673 of SEQ ID NO:11; (l) SEQ ID NO:12 and having an amino acid substitution or deletion at a position corresponding to position 69

of SEQ ID NO:12; (m) SEQ ID NO:13 and having an amino acid substitution or deletion at a position corresponding to position 107 of SEQ ID NO:13; (n) SEQ ID NO:17 and having an amino acid substitution or deletion at a position corresponding to position 246 of SEQ ID NO:17; (o) SEQ ID NO:18 and having an amino acid substitution or deletion at a position corresponding to position 545 of SEQ ID NO:18; or (p) SEQ ID NO:19 and having an amino acid substitution or deletion at a position corresponding to position 124 of SEQ ID NO:19. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0014] In some aspects, the polypeptide in accordance with the foregoing paragraph comprises a Pro substitution at a position corresponding to position 344 of SEQ ID NO:1. In certain aspects, the polypeptide comprises a His substitution at a position corresponding to position 702 of SEQ ID NO:2. In some aspects, the polypeptide comprises an Ala substitution at a position corresponding to position 655 of SEQ ID NO:3. In some aspects, the polypeptide comprises an Ala substitution at a position corresponding to position 73 of SEQ ID NO:4. In certain aspects, the polypeptide comprises a Gly substitution at a position corresponding to position 781 of SEQ ID NO:5. In some aspects, the polypeptide comprises a Val substitution at a position corresponding to position 136 of SEQ ID NO:6. In some aspects, the polypeptide comprises an Ala substitution at a position corresponding to position 183 of SEQ ID NO:7. In certain aspects, the polypeptide comprises an Arg substitution at a position corresponding to position 1 of SEQ ID NO:8. In some aspects, the polypeptide comprises a Cys substitution at a position corresponding to position 102 of SEQ ID NO:9. In some aspects, the polypeptide comprises a Cys substitution at a position corresponding to position 105 of SEQ ID NO:10. In certain aspects, the polypeptide comprises a Leu substitution at a position corresponding to position 673 of SEQ ID NO:11. In some aspects, the polypeptide comprises a Gly substitution at a position corresponding to position 69 of SEQ ID NO:12. In certain aspects, the polypeptide comprises a Ser substitution at a position corresponding to position 107 of SEQ ID NO:13. In some aspects, the polypeptide comprises an Ala substitution at a position corresponding to position 246 of SEQ ID NO:17. In certain aspects, the polypeptide comprises an Ile substitution at a position corresponding to position 545 of SEQ ID NO:18. In some aspects, the polypeptide comprises a Pro substitution at a position corresponding to position 124 of SEQ ID NO:19. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0015] In some aspects, the polypeptide further comprises an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:1 and a Thr substitution at a position corresponding to position 999, a Thr substitution at a position corresponding to position 457, a Pro substitution at a position corresponding to position 591, a Thr substitution at a position corresponding to position 183, a Leu substitution at a position corresponding to position 358, a Arg substitution at a position corresponding to position 23, a Ile substitution at a position corresponding to position 902, a Val substitution at a position corresponding to position 889, a Cys substitution at a position corresponding to position 620, and/or a Gly substitution at a position corresponding to position 174 of SEQ ID NO:1. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0016] In certain aspects, the polypeptide further comprises an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:3 and a Cys substitution at a position corresponding to position 398, an Ala substitution at a position corresponding to position 652, a Cys substitution at a position corresponding to position 264, and/or an Ala substitution at a position corresponding to position 21 of SEQ ID NO:3. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0017] In some aspects, the polypeptide further comprises an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:4 and an Asn substitution at a position corresponding to position 45, a Leu substitution at a position corresponding to position 290, a Asp substitution at a position corresponding to position 271, a Tyr substitution at a position corresponding to position 153, a Val substitution at a position corresponding to position 45, a Pro substitution at a position corresponding to position 284, an Ile substitution at a position corresponding to position 73, a Leu substitution at a position corresponding to position 68, an Ile substitution at a position corresponding to position 69, a Cys substitution at a position corresponding to position 305, a Ser substitution at a position corresponding to position 144, and/or a Val substitution at a position corresponding to position 281 of SEQ ID NO:4. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0018] In certain aspects, the polypeptide further comprises an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID

NO:5 and a Gly substitution at a position corresponding to position 916, a His substitution at a position corresponding to position 938, a His substitution at a position corresponding to position 860, a Asp substitution at a position corresponding to position 925, and/or a Met substitution at a position corresponding to position 470 of SEQ ID NO:5. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0019] In some aspects, the polypeptide further comprises an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:6 and an Ala substitution at a position corresponding to position 115, an Arg substitution at a position corresponding to position 386, an Arg substitution at a position corresponding to position 155, a Ser substitution at a position corresponding to position 98, an Ala substitution at a position corresponding to position 201, a Thr substitution at a position corresponding to position 294, a Tyr substitution at a position corresponding to position 159, and/or an Ile substitution at a position corresponding to position 112 of SEQ ID NO:6. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0020] In some aspects, the polypeptide further comprises an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:18 and a Gly substitution at a position corresponding to position 76, a Val substitution at a position corresponding to position 293, a Thr substitution at a position corresponding to position 637, a Val substitution at a position corresponding to position 3, a Ser substitution at a position corresponding to position 311, a Thr substitution at a position corresponding to position 471, a Val substitution at a position corresponding to position 228, a Ser substitution at a position corresponding to position 311, and/or a Thr substitution at a position corresponding to position 257 of SEQ ID NO: 18. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0021] In certain aspects, the polypeptide further comprises an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:7 and a Val substitution at a position corresponding to position 173, a Leu substitution at a position corresponding to position 196, a Phe substitution at a position corresponding to position 180, an Ala substitution at a position corresponding to position 249, a Val substitution at a position corresponding to position 5, a Leu substitution at a position

corresponding to position 273, and/or an Asn substitution at a position corresponding to position 176 of SEQ ID NO:7. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0022] In some aspects, the polypeptide further comprises n amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:9 and a Cys substitution at a position corresponding to position 15 and/or a substitution at a position corresponding to position 30 of SEQ ID NO:9. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0023] In certain aspects, the polypeptide further comprises n amino acid sequence at least 90% identical to SEQ ID NO:19 and a Ile substitution at a position corresponding to position 193, a Thr substitution at a position corresponding to position 233, an Ala substitution at a position corresponding to position 300, and/or an Arg substitution at a position corresponding to position 199 of SEQ ID NO:19.

[0024] In some aspects, the polypeptide further comprises an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:17 and an Ala substitution at a position corresponding to position 246 of SEQ ID NO:17. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0025] In some aspects, the polypeptide further comprises an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:11 and a Val substitution at a position corresponding to position 119, a Pro substitution at a position corresponding to position 535, an Arg substitution at a position corresponding to position 373, a Ser substitution at a position corresponding to position 535, a Thr substitution at a position corresponding to position 119, an Ala substitution at a position corresponding to position 601, a Lys substitution at a position corresponding to position 103, an Asp substitution at a position corresponding to position 31, an Ile substitution at a position corresponding to position 662, a Lys substitution at a position corresponding to position 359, and/or an Asp substitution at a position corresponding to position 519 of SEQ ID NO:11. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0026] Another embodiment provides a nucleic acid molecule encoding a polypeptide comprising: (i) an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:14, and (ii) an Ile substitution at a position corresponding to position 212, an Asn substitution at a position corresponding to position 162, an Ala substitution at a position corresponding to position 299, and/or a Arg substitution at a position corresponding to position 220. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0027] In yet another embodiment, there is provided a recombinant polypeptide comprising: (i) an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:14, and (ii) an Ile substitution at a position corresponding to position 212, an Asn substitution at a position corresponding to position 162, an Ala substitution at a position corresponding to position 299, and/or a Arg substitution at a position corresponding to position 220. In still further aspects, there is provided a nucleic acid molecule encoding one of the foregoing polypeptides.

[0028] In a further embodiment, there is provided a nucleic acid molecule encoding a polypeptide comprising: (i) an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:15, and (ii) an Ala substitution at a position corresponding to position 184, an Asn substitution at a position corresponding to position 1730, an Asp substitution at a position corresponding to position 1888, a Thr substitution at a position corresponding to position 352, a Ser substitution at a position corresponding to position 374, an Arg substitution at a position corresponding to position 1423, a Glu substitution at a position corresponding to position 1502, a Gln substitution at a position corresponding to position 285, a Lys substitution at a position corresponding to position 470, a Thr substitution at a position corresponding to position 939, a Phe substitution at a position corresponding to position 1669, an Ile substitution at a position corresponding to position 2034, an Asn substitution at a position corresponding to position 1713, a Thr substitution at a position corresponding to position 704, and/or an Ile substitution at a position corresponding to position 1084.

[0029] In yet a further embodiment, there is provided a recombinant polypeptide comprising: (i) an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:15, and (ii) an Ala substitution at a position corresponding to position 184, an Asn substitution at a position corresponding to position

1730, an Asp substitution at a position corresponding to position 1888, a Thr substitution at a position corresponding to position 352, a Ser substitution at a position corresponding to position 374, an Arg substitution at a position corresponding to position 1423, a Glu substitution at a position corresponding to position 1502, a Gln substitution at a position corresponding to position 285, a Lys substitution at a position corresponding to position 470, a Thr substitution at a position corresponding to position 939, a Phe substitution at a position corresponding to position 1669, an Ile substitution at a position corresponding to position 2034, an Asn substitution at a position corresponding to position 1713, a Thr substitution at a position corresponding to position 704, and/or an Ile substitution at a position corresponding to position 1084.

[0030] In an embodiment, there is provided a nucleic acid molecule encoding a polypeptide comprising: (i) an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:16, and (ii) a Phe substitution at a position corresponding to position 112, an Ala substitution at a position corresponding to position 126, an Leu substitution at a position corresponding to position 978, a Gly substitution at a position corresponding to position 199, a Thr substitution at a position corresponding to position 476, and/or a Val substitution at a position corresponding to position 735.

[0031] A further embodiment provides a recombinant polypeptide comprising: (i) an amino acid sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO:16, and (ii) a Phe substitution at a position corresponding to position 112, an Ala substitution at a position corresponding to position 126, an Leu substitution at a position corresponding to position 978, a Gly substitution at a position corresponding to position 199, a Thr substitution at a position corresponding to position 476, and/or a Val substitution at a position corresponding to position 735.

[0032] An embodiment also provides a bacterial strain comprising at least one nucleic acid molecule of the embodiments or expressing at least one polypeptides of the embodiments (*e.g.*, a polypeptide at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to one of the polypeptides of SEQ ID NO: 1 and comprising one of the substitutions listed in Table 1). In some aspects, the bacterial strain comprises at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 or 19 nucleic acid molecules encoding the polypeptides of Table 1. In some aspects, the bacterial strain expresses the polypeptide

encoded by at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 or 19 nucleic acid molecules.

[0033] In further aspects, the bacterial strain comprises a nucleic acid encoding a tRNA and an aminoacyl-tRNA synthetase for incorporation of selenocysteine. In some aspects, the tRNA recognizes a UAG codon. In some aspects, the bacterial strain comprises a nucleic acid that encodes a tRNA is at least 90% identical to SEQ ID NO: 20, SEQ ID NO: 21, or SEQ ID NO: 22. In some aspects, the bacterial strain further comprises one or more of the following features: (i) a G or C at a position corresponding to position 7; (ii) a T at a position corresponding to position 49; (iii) a A or C at a position corresponding to position 50; (iv) a T at a position corresponding to position 64; (v) a G or A at a position corresponding to position 65; and/or (vi) a G, T or C at a position corresponding to position 66. In some aspects, the molecule encodes a tRNA comprising the sequence at least about 90% identical to SEQ ID NO: 20; and comprises one or more of the following features: (i) a G at a position corresponding to position 7; (ii) a T at a position corresponding to position 49; (iii) a C at a position corresponding to position 50; (iv) a T at a position corresponding to position 64; (v) a G at a position corresponding to position 65; and/or (vi) a C at a position corresponding to position 66. In further aspects, the molecule encodes a tRNA comprising a sequence that is at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20 or SEQ ID NO: 4. In specific aspects, the molecule encodes a tRNA comprising SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20 or SEQ ID NO: 4.

[0034] In some aspects, the bacterial strain further comprises an expressible nucleic sequence encoding a polypeptide of interest having at least one position in the coding sequence with a TAG codon for selenocysteine incorporation. In certain aspects, the expressible nucleic sequence encodes a human polypeptide. In some aspects, the expressible nucleic sequence encodes an enzyme or an antibody. In certain aspects, the expressible nucleic sequence comprises a T7 RNA polymerase promoter. In some aspects, the bacterial strain further comprises a nucleic acid sequence encoding T7 RNA polymerase.

[0035] In further aspects, the bacterial strain is a gram negative bacteria, such as an *E. coli* strain. In a further embodiment provides an *E. coli* bacterial strain deposited at NCIMB under the accession no. 42595. In yet a further embodiment there is provided a recombinant polypeptide comprising at least one selenocysteine residue at a selected position produced by

a expressing a nucleic acid encoding polypeptide in a bacterial strain according the embodiments and in the presence of selenium source and purifying the recombinant polypeptide from the bacteria.

[0036] A further embodiment provides a culture of bacteria comprising an expressed recombinant polypeptide in an amount of 5 to 100 mg/L of the culture. In certain aspects, the expressed recombinant polypeptide comprises at least one selenocysteine residue. In some aspects, the culture of bacteria comprises an expressed recombinant polypeptide in an amount of 10 to 40 mg/L of the culture, said expressed recombinant polypeptide comprising at least one selenocysteine residue. In certain aspects, the expressed recombinant polypeptide is present in an amount of 5 to 50 mg/L, 10 to 80 mg/L, 15 to 60 mg/L, 10 to 30 mg/L, 20 to 80 mg/L, 30 to 90 mg/L, 40 to 80 mg/L, 50 to 70 mg/L, 60 to 90 mg/L, 70 to 80 mg/L, or 90 to 100 mg/L of the culture. In certain aspects, the expressed recombinant polypeptide is present in an amount of 5 to 10 mg/L, 7 to 15 mg/L, 10 to 20 mg/L, 15 to 30 mg/L, 20 to 35 mg/L, 30 to 40 mg/L, 35 to 45 mg/L, 40 to 50 mg/L, 45 to 55 mg/L, 50 to 50 mg/L, 55 to 65 mg/L, 50 to 60 mg/L, 65 to70 mg/L, 75 to 85 mg/L, 85 to 90 mg/L, 80 to 95 mg/L, 85 to 98 mg/L, or 95 to 100 mg/L of the culture. In certain aspects, the expressed recombinant polypeptide is in an amount of at least 1 mg/L, 5 mg/L, 10 mg/L, 15 mg/L, 20 mg/L, 25 mg/L, 30 mg/L, 35 mg/L, 40 mg/L, 45 mg/L, 50 mg/L, 55 mg/L, 60 mg/L, 65 mg/L, 70 mg/L, 75 mg/L, 80 mg/L, 85 mg/L, 90 mg/L, 95 mg/L, 100 mg/L or higher in the culture.

[0037] In some aspects, the expressed recombinant polypeptide is a polypeptide of the embodiments (*e.g.*, a polypeptide at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to one of the polypeptides of SEQ ID NO: 1 and comprising one of the substitutions listed in Table 1). In certain aspects, the expressed recombinant polypeptide comprises at least one selenocysteine residue at a selected position not found a wild type version of the polypeptide. In some aspects, at least 80% of the expressed recombinant polypeptides in the culture comprise the selenocysteine residue at the selected position. In certain aspects, 80%-99.9% of the recombinant polypeptides in the culture comprise the selenocysteine residue at the selected position. In some aspects, 90%-99% of the expressed recombinant polypeptides in the culture comprise the selenocysteine residue at the selected position. In some aspects, at least 95% of the expressed recombinant polypeptides in the culture comprise the selenocysteine residue at the selected position. In some aspects, at least 99% of the expressed recombinant polypeptides in the culture comprise the selenocysteine

residue at the selected position. In some aspects, the expressed recombinant polypeptide is at least 90% identical to a human polypeptide. In certain aspects, the human polypeptide is a polypeptides involved in a disease. In some aspects, the expressed recombinant polypeptide comprises an antibody or an enzyme. In certain aspects, the polypeptide comprises at least two selenocysteine residues at selected positions. In some aspects, the two selenocysteine residues at the selected positions form a diselenide bond. In specific aspects, the polypeptide comprises 2, 3, 4, 5, 6, 7, 8, 9 or 10 selenocysteine residues at selected positions. An even further embodiment provides a polypeptide comprising at least a first selenocysteine residues purified from a culture of the embodiments. In yet another embodiment, there is provided a method of expressing a polypeptides comprising at least one selenocysteine residue comprising: (a) expressing a nucleic acid encoding the polypeptide in a bacterial strain of the embodiments and in the presence of a selenium source; and (b) purifying the recombinant polypeptide from the bacteria. In another embodiment, there is provided a recombinant polypeptide comprising at least one selenocysteine residue at a selected position produced by a method comprising: (a) expressing a nucleic acid encoding the polypeptide in a bacterial strain of the embodiments and in the presence of a selenium source; and (b) purifying the recombinant polypeptide from the bacteria.

[0038] In an even further embodiment, there is provided the use of a bacterial strain of the embodiments as a host for production of a polypeptide comprising at least one selenocysteine residue. In further aspects, the bacterial strain is cultured in a medium comprising a selenium source.

[0039] In a further embodiment there is provided a transgenic bacterial strain comprising heterologous nucleic acids encoding translation components for incorporation of at least a first non-canonical amino acid and a screenable or selectable marker polypeptide that exhibits enhanced activity when at least one position of the marker polypeptide is said first non-canonical amino acid. In specific aspects, the screenable marker is a fluorescent or luminescent polypeptide. In further aspects, the bacterial strain comprises at least one nucleic acid molecule of the embodiments or expressing at least one polypeptide of the embodiments (*e.g.*, a polypeptide at least 90% identical to one of the polypeptides of SEQ ID NO: 1 and comprising one of the substitutions listed in Table 1).

[0040] In other aspects, the bacterial strain comprises a heterologous nucleic acid encoding a selectable marker that exhibits enhanced activity when at least one position of the

marker polypeptide is said first non-canonical amino acid. In further aspects, the selectable marker is a polypeptide that provides antibiotic resistance. In particular aspects, the selectable marker is a beta-lactamase enzyme.

[0041] In some aspects, the bacterial stain is a Gram positive or a Gram negative bacterial cell. In some specific aspects, the bacterial cell is an *E. coli* cell. In other particular aspects, the bacterial cell is an *Enterobacter* or *Serratia* bacteria. In further aspects, the bacterial cell is an *Enterobacter cloacae or Serratia marcescens* bacterial cell.

[0042] In certain aspects, the translation components for incorporation of the first non-canonical amino acid comprise a nucleic acid encoding a tRNA and an aminoacyl-tRNA synthetase for the first non-canonical amino acid. In some aspects, the tRNA recognizes a UAG codon. In further aspects, the tRNA is at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22 or SEQ ID NO: 23. In specific aspects, the tRNA comprises SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22 or SEQ ID NO: 23.

[0043] In still further aspects, the translation components for incorporation of the first non-canonical amino acid further comprise a nucleic acid encoding an enzyme for synthesis for the first non-canonical amino acid. In a particular aspect, the non-canonical amino acid is selenocysteine. In other aspects, the cell comprises a nucleic acid encoding *selA, selB* and/*or selC*. In a specific aspect, the cell comprises a nucleic acid encoding *selA*. In some aspects, the bacterial cell comprises an inactivated or deleted *prfA* gene. In certain aspects, the cell has been engineered to lack endogenous Amber (TAG) codons. In some particular aspects, the cell is or is derived from a *E. coli* C321.ΔA.

[0044] In a further embodiment the invention provides a population of bacterial cells in accordance with the embodiments and aspects described above. In certain aspects, the population comprises 1 x 10^3 to 1 x 10^{12} bacterial cells.

[0045] In still a further embodiment, there is provided a method of producing a commercial polypeptide comprising at least a first non-canonical amino acid comprising (i) obtaining a bacterial strain according to the embodiments and an expression cassette encoding the commercial polypeptide; and (ii) incubating the bacterial strain in conditions that allow expression of the commercial polypeptide. In some aspects, the expression cassette encoding the commercial polypeptide is under the control of an inducible promoter. In

certain aspects, the method further comprises isolating the expressed commercial polypeptide.

[0046] In yet still a further embodiment, the invention provides a method of screening for a polypeptide having a desired activity comprising (i) obtaining a population of bacterial cells according to the embodiments, said cells encoding a library of candidate polypeptides, said polypeptides comprising at least a first non-canonical amino acid position; and (ii) screening the population of bacteria to identify a candidate polypeptide having the desired biological activity. In specific aspects, the population of bacterial cells comprises nucleic acid constructs encoding 100 to 10,000,000 different candidate polypeptides.

[0047] In still a further embodiment, there is provided a recombinant nucleic acid molecule, wherein the molecule encodes a tRNA that is at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 20, SEQ ID NO: 21, or SEQ ID NO: 22 and comprising one or more of the following features: a G or C at a position corresponding to position 7; a T at a position corresponding to position 49; an A or C at a position corresponding to position 50; a T at a position corresponding to position 64; a G or A at a position corresponding to position 65; and/or a G, T or C at a position corresponding to position 66. In some particular aspects, the molecule encodes a tRNA comprising the sequence at least about 90% identical to SEQ ID NO: 20; and comprises one or more of the features listed above. In further specific aspects, the molecule comprises 2, 3, 4, 5 or 6 of the features listed above. In certain aspects, the molecule encodes a tRNA comprising the sequence at least about 90% identical to SEQ ID NO: 21 or SEQ ID NO: 19. In particular aspects, the molecule encodes a tRNA comprising the sequence of SEQ ID NO: 20, SEQ ID NO: 21, or SEQ ID NO: 22.

[0048] In yet still a further embodiment there is provided a recombinant polypeptide, encoding a beta-lactamase enzyme, said enzyme comprising a disulfide bond between two cysteine positions that is required for activity of the enzyme, where at least one of said two cysteine positions is substituted with a selenocysteine. In some aspects, both of said cysteine residues are substituted with a selenocysteine. In further aspects, the beta-lactamase enzyme is a SME-type beta-lactamase or NMC-A beta-lactamase. For example, the beta-lactamase can comprises a sequence least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 24 and wherein the positions corresponding to C69 and/or

C238 are selenocysteine. In certain aspects, the positions corresponding to C69 and C238 are selenocysteine.

[0049] Further embodiments of the invention provide a recombinant nucleic acid molecule encoding the polypeptide according to the embodiments and aspects described above. In some aspects, the codons corresponding to the selenocysteine position(s) is a UAG codon. In other aspects, the sequence is at least about 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 25.

[0050] As used herein, "essentially free," in terms of a specified component, is used herein to mean that none of the specified component has been purposefully formulated into a composition and/or is present only as a contaminant or in trace amounts. The total amount of the specified component resulting from any unintended contamination of a composition is therefore well below 0.01%. Most preferred is a composition in which no amount of the specified component can be detected with standard analytical methods.

[0051] As used herein the specification, "a" or "an" may mean one or more. As used herein in the claim(s), when used in conjunction with the word "comprising," the words "a" or "an" may mean one or more than one.

[0052] The use of the term "or" in the claims is used to mean "and/or" unless explicitly indicated to refer to alternatives only or the alternatives are mutually exclusive, although the disclosure supports a definition that refers to only alternatives and "and/or." As used herein "another" may mean at least a second or more.

[0053] Throughout this application, the term "about" is used to indicate that a value includes the inherent variation of error for the device, the method being employed to determine the value, or the variation that exists among the study subjects.

[0054] Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0055] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

[0056] FIGS. 1A-1C: *E. coli* strains for production of selenoproteins are conditionally dependent on selenocysteine for growth and survival. An *E. coli* strain containing an integrated β-lactamase with the native disulfide bond (FIG. 1A) does not require selenium supplementation to survive in the presence of the β-lactam antibiotic carbenicillin. *E. coli* strains with either one (FIG. 1B) or two (FIG. 1C) essential selenocysteine residues, which form either a selenyl-sulfhydryl or diselenide bond, respectively, require selenium for resistance to β-lactam antibiotics. Conditional dependence on selenocysteine incorporation prevents loss or attenuation of the otherwise toxic selenocysteine biosynthesis and incorporation machinery. In FIGS. 1B and 1C, the lines, from top to bottom, at time 16h represent MOPS EZ, Se Only, Se + Carb, and Carb Only.

[0057] FIGS. 2A-2B: *E. coli* strains for the production of selenoproteins contain a set of mutations which significantly enhances cell growth and resistance to β-lactam antibiotics via selenocysteine dependent β-lactamase. A parental *E. coli* strain (FIG. 2A) has moderate resistance to β-lactam antibiotics mediated via a selenocysteine dependent β-lactamase during while growing in a rich medium. An *E. coli* strain for the production of recombinant selenoproteins (FIG. 2B) containing a series of mutations has a higher growth rate, final cell density and resistance to carbenicillin. This improved growth makes the mutant strain a superior host for the production of recombinant selenoproteins. In FIG. 2A, the lines, from top to bottom, at time 16h represent LB, 100 Carb, 1000 Carb, and 10000 Carb with the lines for 1000 Carb and 10000 Carb being overlapping. In FIG. 2B, the lines, from top to bottom, at time 16h represent LB, 100 Carb, 1000 Carb, and 10000 Carb with the lines for 100 Carb and 1000 Carb being overlapping.

[0058] FIGS. 3A-3B: *E. coli* strains for the production of selenoproteins contain a set of mutations which significantly enhances cell growth in defined media. The parental *E. coli* strain (FIG. 3A) has an extended lag phase and displays poor growth in a defined growth medium. An *E. coli* strain for the production of recombinant selenoproteins (FIG. 3B)

containing a series of mutations has a higher growth rate, final cell density and selenocysteine dependent resistance to β-lactam antibiotics. For reproducibility and quality assurance, defined growth media are commonly used for the commercial production of recombinant proteins. In FIGS. 3A and 3B, the lines, from top to bottom, at time 20h represent MOPS EZ, Se Only, Se + Carb, and Carb Only.

[0059] FIGS. 4A-4B: Intact mass spectrum and UVPD fragmentation map of *E. coli* dihydrofolate reductase containing a diselenide bond. The mass spectrum of *E. coli* DHFR (FIG. 4A, average masses shown) confirms incorporation of two selenocysteine residues with approximately 100% efficiency. No masses corresponding to the incorporation of either one or two serine residues were detected. UVPD fragmentation mapping (FIG. 4B, SEQ ID NO: 26) confirms incorporation of selenocysteine (U) at positions 39 and 85 and the formation of a diselenide bond. Diselenide bond formation is indicated by the lack of ions corresponding to fragmentation events between the two selenocysteine residues. Yield of DHFR was 8 mg/L. Other proteins containing a diselenide bond have been expressed at yields exceeding 40 mg/L.

[0060] FIGS. 5A-5C: Mass spectrum and UVPD fragmentation map of anti-MS2 scFv containing two essential diselenide bonds. An *E. coli* strain developed for the expression of recombinant selenoproteins enables the production of diselenide stabilized antibody fragments in the bacterial cytoplasm. Unlike the *E. coli* strains developed for the expression of proteins containing disulfide bonds, this strain does not require an elevated cytoplasmic redox potential. The experimentally determined monoisotopic mass (FIG. 5A) is consistent with a recombinant anti-MS2 scFv containing four selenocysteine residues which have formed two diselenide bonds. Diselenide bond formation is indicated by the loss of four protons (FIG. 5A, row 3 vs. row 2). Intact mass spectrum of the anti-MS2 scFv containing four selenocysteine residues (FIG. 5B) displayed as average masses. No masses corresponding to incorporation of serine residues were detected. UVPD fragmentation mapping (FIG. 5C, SEQ ID NO: 27) confirms incorporation of selenocysteine (U) at positions 42, 116, 179 and 249 and also diselenide bond formation. Diselenide bond formation is indicated by the lack of ions corresponding to fragmentation events between the pairs of selenocysteine residues, U42:U116 and U179:U249).

[0061] FIGS. 6A-6D: UVPD fragment maps and ELISA for wild-type anti-ricin A chain scFv (FIGS. 6A and 6B, respectively) and seleno anti-ricin A chain scFv (FIGS. 6C and

6D, respectively). For the 193 nm UVPD sequence information, covalently bonded (or potentially bonded) cysteine residues (FIG. 6A, SEQ ID NO: 28) and covalently bonded (or potentially bonded) selenocysteine residues (FIG. 6C, SEQ ID NO: 29) are shaded in gray. The gaps in sequence coverage indicate the formation of covalent disulfide bonds between the cysteine residues. The absence of fragmentation between the selenocysteine residues confirms formation of the two diselenide bonds. The bond connectivity is identical to the wild-type anti-ricin A chain scFv. Treatment of the wild-type anti-ricin A chain scFv with DTT leads to significant loss of activity (FIG. 6B). The selenocysteine containing scFv is strongly resistant to reducing conditions (FIG. 6D). Treatment with 50 mM DTT resulted in only a slight loss of affinity (EC_{50} 7.95 nM to EC_{50} 11.4 nM). In FIGS. 6B and 6D, the lines, from top to bottom, at scFv conc 10 nM represent 0 mM DTT, 1 mM DTT, 10 mM DTT, and 50 mM DTT.

[0062] FIGS. 7A-7C: UVPD fragment maps of trastuzumab (Herceptin) scFvs produced in *E. coli* strains BL21DE3 (FIG. 7A, SEQ ID NO: 30), T7 Shuffle Express (FIG. 7B, SEQ ID NO: 30) and RTΔA-2X310K (FIG. 7C, SEQ ID NO: 31). For the 193 nm UVPD sequence information shown above, covalently bonded (or potentially bonded) cysteine residues (FIG. 7B) and covalently bonded selenocysteine residues (FIG. 7C) are shaded in gray. (FIG. 7A) Even fragmentation throughout the protein sequence in indicates no formation of disulfide bonds. (FIG. 7B) Lack of fragmentation in the second half of the sequence confirms formation of a disulfide bond in the VH region only. No disulfide bond formed in the VL region despite expression in oxidizing conditions. (FIG. 7C) Lack of fragmentation in both the VL and VH regions confirms formation of two diselenide bonds. Only the diselenide scFv adopted the native and expected covalent architecture.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

[0063] The use of non-canonical amino acids in proteins offers the possibility of polypeptides having greatly expanded functionality that could be exploited for wide range of applications. For example, by incorporation of selenocysteine into polypeptides it may be possible to develop enzymes having enhanced levels of stability or activity and to produce highly active therapeutic polypeptides. However, these approaches have, to date, been hampered by the inability to produce organisms that stably retain translation pathways and that predictable and reliably incorporate selenocysteine into encoded polypeptides. Accordingly, the present disclosure overcomes challenges associated with current

technologies by providing an evolved *E. coli* strain and advanced *E. coli* genes capable of efficient selenocysteine protein production. Particularly, for the first time commercially relevant amounts of selenoproteins showing essentially complete incorporation of encoded selenocysteine residues can be achieved. Bacterial strains and mutant genes that enable this production are also provided. For example, studies presented here have identified several point mutations in genes, including *acrB, adhE, arcB, cysK, dnaE, ftsA*, and *ftsL*, that optimize survival in selenium enriched media and contains pathways to ensure dependence on selenocysteine is maintained. One exemplary bacterial strain is the mutant RTΔA 2X310K *E. coli* strain which contains these core mutations providing enhanced growth rates, greater final cell density and approximately 100% incorporation of selenocysteine in polypeptides produced using this strain. Thus, methods and compositions are provided herein for the efficient incorporation of selenocysteine and for production of polypeptides that incorporate selenocysteine positions.

I. System for Selenocysteine Incorporation

[0064] Embodiments of the present disclosure provide a system for the incorporation of selenocysteine, such as for the production of selenoproteins. In one aspect, the system is a bacterial strain, such as an *E. coli* strain, which comprises one or more amino acid substitutions or deletions within one or more of the genes which enhance cell growth or selenocysteine incorporation. The bacterial strain may be *E. coli* strain K012 MG1655 (GenBank Accession No. U00096.3), *E. coli* stain C321ΔA (US2016/0060301), or *E. coli* strain RTΔA. In certain embodiments the bacterial strain is the RTΔA 2X310K strain.

[0065] In particular aspects, the system comprises evolved bacterial strains which are conditionally dependent on selenocysteine. In one example, the evolved bacterial strain is derived from a strain, such as the RTΔA strain, which comprises deletions of the genomic *selA*, *selB* and *selC* genes (encoding SelA, SelB and $tRNA^{Sec}$, respectively) and which lacks the *prfA* gene encoding release factor 1 (RF1) allowing for efficient incorporation of a range of unnatural amino acids. The deletion of these genes eliminated any crosstalk between a UAG suppressing tRNA (*e.g.*, $tRNA^{SecUx}$) and the endogenous selenocysteine incorporation machinery. In addition, the RTΔA strain contains a reporter protein to selenocysteine, the NMC-A β-lactamase from *Enterobacter cloacae* with an essential selenyl-sulfhydryl or diselenide bond (Δ*aphC*::*nmcA* C69U C238 or C69U C238U) rendering conditional dependence on selenocysteine incorporation for resistance to β-lactam antibiotics.

Conditional dependence on selenocysteine incorporation prevents loss or attenuation of the otherwise toxic selenocysteine biosynthesis and incorporation machinery. The efficiency of selenocysteine incorporation may be enhanced by the expression of an evolved tRNA for site specific incorporation of selenocysteine (*e.g.*, for example, $tRNA^{SecUX}$; U.S. Patent App. Publn. 2017/0166945, incorporated herein by reference). Further enhancements may include expression of various genes including, but not limited to, a *selD, selA* and/or *pstK* gene. For example, a bacterial strain could express *E. coli selD*, *E. coli selA* and *M. jannaschii pstK*.

[0066] In certain embodiments, the selenocysteine-incorporating system encodes for one or more *E. coli* genes with one or more amino acid deletions or substitutions. The genes can include *acrB* (SEQ ID NO:1), *adhE* (SEQ ID NO:2), *arcB* (SEQ ID NO:3), *cysK* (SEQ ID NO:4), *dnaE* (SEQ ID NO:5), *ftsA* (SEQ ID NO:6), *ftsL* (SEQ ID NO:18), *hemA* (SEQ ID NO:7), *mdfA* (SEQ ID NO:8), *ompR* (SEQ ID NO:9), *oxyR* (SEQ ID NO:19), *pcnB* (SEQ ID NO:10), *prfB* (SEQ ID NO:17), *pta* (SEQ ID NO:11), *queE* (SEQ ID NO: 69), and/or *ydiL* (SEQ ID NO:13). Additional genes include *pdxB* (SEQ ID NO:14), *yeeJ* (SEQ ID NO:15), and *yphG* (SEQ ID NO:16). For example, a selenocysteine-incorporating bacterial strain could comprise a T246A mutation in the *prfB* gene in combination with at least one *hemA* mutation. As another example, a selenocysteine-incorporating bacterial strain could comprise a T246A mutation in the *prfB* gene in combination with at least one *hemA* mutation and at least one *yeeJ* mutation. Exemplary amino acid substitutions for these genes are depicted in Table 1.

[0067] Table 1: Mutations enriched in evolved strains.

***E. coli* gene**	**Mutation(s)**
AcrB (SEQ ID NO:1)	A999T, A457T, **L344P**, L591P:A183T, F358L, G23R, M902I, L230R, A889V:R620C, D174G
AdhE (SEQ ID NO:2)	**R702H**
ArcB (SEQ ID NO:3)	**T655A**, R398C, T652A, R264C, V21A
CysK (SEQ ID NO:4)	I45N, F290L, G271D, **T73A**, H153Y, I45V, L284P, T73I, P68L, T69I, R305C, F144S:A281V
DnaE (SEQ ID NO:5)	E916G, Y938H, R860H, **E781G**, G925D, T470M
FtsA (SEQ ID NO:6)	T115A, G386R, **A136V**, Q155R, P98S:E201A, A294T, H159Y, V112I
HemA (SEQ ID NO:7)	A173V, **V183A**, P196L, L180F, D127A:S249A, A5V:P273L, A5V, I176N
MdfA (SEQ ID NO:8)	**M1R**
OmpR (SEQ ID NO:9)	**Y102C**, R15C, Q30R
PcnB (SEQ ID NO:10)	**R105C**
Pta (SEQ ID NO:11)	A119V, S535P, W373R, **P673L**, P673S, S535P,

	A119T, D601A, A119T:P673L, E103K, G31D, T662I, E359K:N519D
QueE (SEQ ID NO:12)	**A69G**
YdiK (SEQ ID NO:13)	**L107S**
PdxB (SEQ ID NO:14)	V212I, D162N, T299A, C220R
YeeJ (SEQ ID NO:15)	S1467P, V1184A, S1730N, G1888D, A352T, G374S, S1423R, D1502E, R285Q, E470K, A939T, L1669F, V2034I, S1713N, D1233N, A704T, M1084I
YphG (SEQ ID NO:16)	S112F, T126A, P978L, E199G, A476T, A735V
PrfB (SEQ ID NO:17)	**T246A**
FtsI (SEQ ID NO:18)	**V545I**, D76G:A293V, A537T, A3V:P311S, M471T, A228V:P311S, A257T
OxyR (SEQ ID NO:19)	M193I, A233T, **L124P**, V300A, C199R

Mutations in bold are core mutations in *E. coli* RTΔA 2X310K.

[0068] In additional aspects, the *E. coli* genes may be further modified by one or more other amino substitutions. For example, amino acid substitutions can be made at one or more positions wherein the substitution is for an amino acid having a similar hydrophilicity. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like. Thus such conservative substitution can be made in selenocysteine-containing polypeptides and will likely only have minor effects on their activity. As detailed in U.S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 ± 1); glutamate (+3.0 ± 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 ± 1); alanine (0.5); histidine -0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4). These values can be used as a guide and thus substitution of amino acids whose hydrophilicity values are within ±2 are preferred, those that are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred. Thus, any of the selenocysteine-containing polypeptides described herein may be modified by the substitution of an amino acid, for different, but homologous amino acid with a similar hydrophilicity value. Amino acids with hydrophilicities within +/- 1.0, or +/- 0.5 points are considered homologous.

II. Production of Recombinant Polypeptides Comprising Selenocysteine

[0069] The selenocysteine-incorporating system provided herein may be used for the production of polypeptides comprising selenocysteine, such as antibodies or disulfide-bonded proteins. In some cases a selenocysteine residue can be substituted for a naturally occurring cysteine, or at any site in which substitution of a selenocysteine residue does not alter the structure and function of the polypeptide, *e.g.*, at a serine residue. In still further aspects, method according to the embodiments can be used to produce polypeptides that naturally comprise one or more selenocysteine residues, such a human selenoprotein (*e.g.*, human TrxR).

[0070] In certain embodiments, a nucleic acid (*e.g.*, plasmid) encoding the polypeptide can be introduced to the system and cultured in selenium-containing growth medium under conditions in which the cell incorporates at least one selenocysteine residue into the polypeptide. The resultant selenocysteine-containing polypeptide can then be isolated and analyzed such as by mass spectrometry or x-ray crystallography.

[0071] The nucleic acids can be introduced into and maintained in the cell in a recombinant vector that is capable of autonomously replicating in the cell according to standard techniques. The method of transformation, and the choice of expression vehicle, will depend on the nature of the polypeptide to be expressed and the host system selected. Transformation methods are described, *e.g.*, in Ausubel *et al.* (eds.) Current Protocols in Molecular Biology (John Wiley & Sons, New York, 1994); expression vehicles can be chosen from those well-known in the art, e.g., in Cloning Vectors: A Laboratory Manual (P. H. Pouwels *et al.*, 1985, Suppl. 1987).To induce expression of the heterologous polypeptide, the cell culture medium typically contains between 1 and 50 ng/ml, preferably 2 to 40 ng/ml, and most preferably 5 to 25 ng/ml of selenium. The selenium may be present as sodium selenite or another soluble, oxidized form of selenium, (*e.g.*, 0.1 to 50 μM, particularly 5 to 25 μM Na_2SeO_3, can be used). By "heterologous" nucleic acid is meant a nucleic acid which is partly or entirely foreign to the cell or animal in which it is introduced, or a nucleic acid which is homologous to an endogenous gene of the cell or animal with the exception that the heterologous protein contains selenocysteine substituted for at least one amino acid.

[0072] In order to obtain expression of the nucleic acid sequences, the sequences may be incorporated in a vector having one or more control sequences operably linked to the

nucleic acid to control its expression. The vectors may include other sequences such as promoters to drive the expression of the inserted nucleic acid, nucleic acid sequences so that the polypeptide or peptide is produced as a fusion and/or nucleic acid encoding secretion signals so that the polypeptide produced in the host cell is secreted from the cell. Polypeptides can then be obtained by transforming the vectors into host cells in which the vector is functional, culturing the host cells so that the polypeptide is produced and recovering the polypeptide from the host cells or the surrounding medium. Prokaryotic cells useful in embodiments of the present invention include *E. coli*, and strains which have T7 RNA polymerase may be preferred for ease of overexpression.

[0073] In some cases the selenocysteine residue can be substituted for a naturally occurring cysteine, or at any site in which substitution of a selenocysteine residue does not alter the structure and function of the polypeptide, *e.g.*, at a serine residue. Such amino acids can be identified by means well known to those skilled in the art, and will usually occur at positions that are not involved in the catalytic or binding activity of the protein (as determined for example by mutational analysis), or at positions considered critical for the structural integrity of the polypeptide. In particular polypeptides, the selenocysteine residues may be incorporated in positions that normally would be occupied by two cysteine residues that form a disulfide bridge; thus, the selenocysteine residues will form a diselenide bond that is identifiable by mass spectrometry. Using standard techniques, the selenocysteine can also be modified to form a selenide, selenoxide, seleninic acid, selenonic acid, selenone, or a seleno-sulfur group.

[0074] In another embodiment, a selenocysteine residue is incorporated at a site in a polypeptide in which the substitution is known (or predicted) to alter the structure and/or function of the polypeptide. This may be to improve the function or characteristics of the polypeptide, to help determine the biological function of the polypeptide, to determine the structure of the polypeptide or for purification of the polypeptide, *e.g.*, to aid in identification of a polypeptide domain, an active site, or a binding site for a drug or another polypeptide.

A. Antibodies with Selenocysteine

[0075] The recombinant polypeptide comprising selenocysteine can encode for an antibody, such as a monoclonal antibody. In some aspects, the selenocysteine residue can be substituted at a position that would be cysteine in the wild type antibody. The antibody may be an IgG, IgM, IgA or an antigen binding fragment thereof. In certain aspects, the antibody

is a Fab', a F(ab')2, a F(ab')3, a monovalent scFv, a bivalent scFv, or a single domain antibody. The antibody may be a non-human antibody, a murine antibody, a human antibody, humanized antibody or de-immunized antibody. In some cases the antibody may be conjugated to an imaging agent, a chemotherapeutic agent, a toxin or a radionuclide. Also provided herein is a composition comprising an antibody of the embodiments and aspects described herein in a pharmaceutically acceptable carrier.

[0076] As used herein, the term "antibody" is intended to refer broadly to any immunologic binding agent, such as IgG, IgM, IgA, IgD, IgE, and genetically modified IgG as well as polypeptides comprising antibody CDR domains that retain antigen binding activity. The antibody may be selected from the group consisting of a chimeric antibody, an affinity matured antibody, a polyclonal antibody, a monoclonal antibody, a humanized antibody, a human antibody, or an antigen-binding antibody fragment or a natural or synthetic ligand. Thus, by known means and as described herein, polyclonal or monoclonal antibodies, antibody fragments, and binding domains and CDRs (including engineered forms of any of the foregoing) may be produced that contain at least one selenocysteine residue. In some aspects, the antibody comprises at least two selenocysteine residues that can form a diselenide bond.

[0077] A monoclonal antibody is a single species of antibody wherein every antibody molecule recognizes the same epitope because all antibody producing cells are derived from a single B-lymphocyte cell line. The methods for generating monoclonal antibodies (MAbs) generally begin along the same lines as those for preparing polyclonal antibodies. In some embodiments, rodents such as mice and rats are used in generating monoclonal antibodies. In some embodiments, rabbit, sheep, or frog cells are used in generating monoclonal antibodies. The use of rats is well known and may provide certain advantages. Mice (*e.g.*, BALB/c mice) are routinely used and generally give a high percentage of stable fusions.

[0078] In one embodiment, the antibody is a chimeric antibody, for example, an antibody comprising antigen binding sequences from a non-human donor grafted to a heterologous non-human, human, or humanized sequence (e.g., framework and/or constant domain sequences). Methods have been developed to replace light and heavy chain constant domains of the monoclonal antibody with analogous domains of human origin, leaving the variable regions of the foreign antibody intact. Alternatively, "fully human" monoclonal antibodies are produced in mice transgenic for human immunoglobulin genes. Methods have

also been developed to convert variable domains of monoclonal antibodies to more human form by recombinantly constructing antibody variable domains having both rodent, for example, mouse, and human amino acid sequences. In "humanized" monoclonal antibodies, only the hypervariable CDR is derived from mouse monoclonal antibodies, and the framework and constant regions are derived from human amino acid sequences (see U.S. Pat. Nos. 5,091,513 and 6,881,557). It is thought that replacing amino acid sequences in the antibody that are characteristic of rodents with amino acid sequences found in the corresponding position of human antibodies will reduce the likelihood of adverse immune reaction during therapeutic use. A hybridoma or other cell producing an antibody may also be subject to genetic mutation or other changes, which may or may not alter the binding specificity of antibodies produced by the hybridoma.

[0079] Examples of antibody fragments suitable for the present embodiments include, without limitation: (i) the Fab fragment, consisting of V_L, V_H, C_L, and C_{H1} domains; (ii) the "Fd" fragment consisting of the V_H and C_{H1} domains; (iii) the "Fv" fragment consisting of the V_L and V_H domains of a single antibody; (iv) the "dAb" fragment, which consists of a V_H domain; (v) isolated CDR regions; (vi) F(ab')2 fragments, a bivalent fragment comprising two linked Fab fragments; (vii) single chain Fv molecules ("scFv"), wherein a V_H domain and a V_L domain are linked by a peptide linker that allows the two domains to associate to form a binding domain; (viii) bi-specific single chain Fv dimers (see U.S. Pat. No. 5,091,513); and (ix) diabodies, multivalent or multispecific fragments constructed by gene fusion (US Patent App. Pub. 20050214860). Fv, scFv, or diabody molecules may be stabilized by the incorporation of disulphide bridges linking the V_H and V_L domains. Minibodies comprising a scFv joined to a CH3 domain may also be made (Hu *et al.*, 1996).

[0080] Antibody-like binding peptidomimetics are also contemplated in embodiments. Liu *et al.* (2003) describe "antibody like binding peptidomimetics" (ABiPs), which are peptides that act as pared-down antibodies and have certain advantages of longer serum half-life as well as less cumbersome synthesis methods.

[0081] Antibodies may be produced from any animal source, including birds and mammals. Preferably, the antibodies are bovine, ovine, murine, rat, rabbit, goat, guinea pig, camel, horse, or chicken. In addition, newer technology permits the development of and screening for human antibodies from human combinatorial antibody libraries. For example, bacteriophage antibody expression technology allows specific antibodies to be produced in

the absence of animal immunization, as described in U.S. Pat. No. 6,946,546, which is incorporated herein by reference. These techniques are further described in: Marks (1992); Stemmer (1994); Gram et al. (1992); Barbas et al. (1994); and Schier et al. (1996).

[0082] In some aspects an antibody having a selenocysteine residue can be an antibody for use as a therapeutic. For example, an antibody can comprise the CDR sequences of a commercial antibody therapeutic such as Cetuximab.

III. Deposit Information

[0083] A representative frozen deposit of *E. coli* strain RTΔA 2X310K has been made with the National Collections of Industrial, Food and Marine Bacteria (NCIMB), 23 St. Machar Drive, Aberdeen AB2 1RY, Scotland, United Kingdom on June 22, 2016. Those deposited cells have been assigned Accession No. NCIMB 42595.

[0084] The foregoing deposit was made in accordance with the terms and provisions of the Budapest Treaty relating to deposit of microorganisms and were made for a term of at least thirty (30) years and at least five (05) years after the most recent request for the furnishing of a sample of the deposits is received by the depository, or for the effective term of the patent, whichever is longer, and will be replaced if it becomes non-viable during that period.

IV. Examples

[0085] The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

Example 1 – Production and Characterization of RTΔA 2X310K

[0086] Several bacterial strains were produced for selenocysteine incorporation and then evolved. A previously described *E. coli* RTΔA stain (Thyer et al., 2015; incorporated

herein by reference) was used for the derivation of more efficient selenocysteine incorporating strains. Briefly, the RTΔA strain comprises deletions of the *selA*, *selB* and *selC* genes (encoding SelA, SelB and $tRNA^{Sec}$, respectively) in strain C321.ΔA (Lajole *et al.*, 2013) which lacks the *prfA* gene encoding release factor 1 (RF1) allowing for efficient incorporation of a range of unnatural amino acids. The deletion of these genes eliminated any crosstalk between the new $tRNA^{Sec}$ library and the endogenous selenocysteine incorporation machinery.

[0087] In addition, the RTΔA strain contains a selenocysteine dependent reporter protein, the NMC-A β-lactamase from *Enterobacter cloacae* with an essential selenyl-sulfhydryl or diselenide bond (Δ*aphC*::*nmcA* C69U C238 or C69U C238U) rendering conditional dependence on selenocysteine. This enzyme has high sequence similarity to the SME-1 β-lactamase from *Serratia marcescens*, an enzyme that has previously been shown to require a disulfide bond adjacent to the active site serine residue for activity, but that confers a significant fitness cost on *E. coli*. It was observed that an *E. coli* strain containing an integrated β-lactamase with the native disulfide bond did not require selenium supplementation to survive in the presence of the β-lactam antibiotic carbenicillin. *E. coli* strains with either one or two essential selenocysteine residues which form either a selenyl-sulfhydryl or diselenide bond respectively required selenium for resistance to β-lactam antibiotics (FIGS. 1A-1C). Conditional dependence on selenocysteine incorporation prevents loss or attenuation of the otherwise toxic selenocysteine biosynthesis and incorporation machinery.

[0088] Two plasmids were introduced which provide the machinery necessary for selenocysteine incorporation. The first plasmid contained a CloDF13 origin of replication and expressed the *E. coli selD* gene and an evolved tRNA for site specific incorporation of selenocysteine ($tRNA^{SecUX}$; U.S. Patent App. Publn. 2017/0166945). The second plasmid (containing a RSF1030 origin of replication) expressed the *E. coli selA* gene and the *M. jannaschii pstK* gene. Following transformation with these two plasmids, the NMC-A beta lactamase containing either zero, one or two selenocysteine residues in place of the essential disulfide bond forming cysteine residues was integrated into the genome at the *aphC* locus. This conferred selenocysteine dependent resistance to some beta-lactam antibiotics and served to set a minimum threshold for selenocysteine incorporation in the cells.

[0089] The resulting strains (in triplicate, along with control strains) were evolved for more than 2500 generations (205 passages to confluence) under two different conditions: increasing antibiotic concentration or increasing temperature. Following evolution, whole genome sequencing was performed on all strains. Strains contained between one and three hundred nonsynonymous mutations within coding regions. A subset of these mutations which were highly enriched (present in multiple independent lines) were introduced to the parental strain for characterization. These mutations (see, Table 1 above) conferred increased growth rate, viability, selenite resistance or other beneficial characteristics. All evolved strains showed dramatically improved growth compared to the parent strains in a variety of different conditions.

[0090] A single clone from one of the lines (designated 2X310K) was isolated and used to benchmark the potential of these evolved strains for recombinant selenoprotein production. This strain is capable of producing diselenide containing proteins with significant yields. While the parental RTΔA strain containing an integrated NMC-A beta lactamase with two selenocysteine residues showed moderate resistance to the β-lactam antibiotic carbenicillin, the mutant 2X310K strain showed a greater resistance to carbenicillin (FIGS. 2A-2B). This improved growth makes the mutant strain a superior host for the production of recombinant selenoproteins. In addition, while the parental strain showed an extended lag phase and poor growth in defined media, the mutant 2X310K strain showed a higher growth rate and final cell density (FIGS. 3A-3B).

[0091] To monitor the efficiency of selenocysteine incorporation and demonstrate the possibilities for protein engineering, *E. coli* dihydrofolate reductase (DHFR) was produced containing an engineered non-essential selenyl-sulfhydryl bond. Top down mass spectrometry showed close to 100% selenocysteine incorporation with no detectable background corresponding to DHFR containing serine. It was observed that the two selenocysteine residues were incorporated with approximately 100% efficiency at positions 39 and 85 (FIGS. 4A-4B). The analysis also confirmed the presence of a diselenide bond.

[0092] The mutant 2X310K strain was used to produce diselenide stabilized anti-MS2 antibody fragments in the bacterial cytoplasm. Unlike the *E. coli* strains developed for the expression of proteins containing disulfide bonds, this strain does not require an elevated cytoplasmic redox potential. The experimentally determined monoisotopic mass was consistent with a recombinant anti-MS2 scFv containing four selenocysteine residues which

formed two diselenide bonds (FIGS. 5A-5C). Thus, the strain can efficiently produce stabilized selenoproteins. Further, the strain was engineered to contain the T7 RNA polymerase which will further increase the yield of recombinant selenoproteins.

[0093] The mutant 2X310K strain was used to produce seleno anti-ricin A chain scFv. The absence of fragmentation between the selenocysteine residues in the UVPD sequence information, shown in FIG. 6C, confirms formation of the two diselenide bonds. The bond connectivity of the seleno anti-ricin A chain scFv is identical to the wild-type anti-ricin A chain scFv, shown in FIG. 6A. However, treatment of the wild-type anti-ricin A chain scFv with DTT leads to significant loss of activity (FIG. 6B), while the selenocysteine containing scFv is strongly resistant to reducing conditions (FIG. 6D).

[0094] *E. coli* strains BL21DE3, T7 Shuffle Express, and RTΔA-2X310K were used to produce trastuzumab (Herceptin) scFvs. In the BL21DE3 strain, no disulfide bonds were formed, as can be seen by the even fragmentation through the protein sequence in FIG. 7A. In the T7 Shuffle Express strain, a disulfide bond formed in the VH region only, as can be seen by the lack of fragmentation in the second half of the sequence in FIG. 7B. However, in the RTΔA-2X310K strain, two diselenide bonds formed, as can be seen by the lack of fragmentation in both the VL and VH regions in FIG. 7C. Thus, only the diselenide scFv adopted the native and expected covalent architecture.

* * *

[0095] All of the methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

U.S. Pat. No. 4,554,101
U.S. Pat. No. 5,091,513
U.S. Pat. No. 6,946,546
U.S. Pat. No. 6,881,557
U.S. Patent App. Publn. 2005/0214860
U.S. Patent App. Publn. 2017/0166945

Ausubel *et al.* (eds.) Current Protocols in Molecular Biology (John Wiley & Sons, New York, 1994).

Barbas *et al.*, *Proc. Natl. Acad. Sci., USA*, 91:3809-13, 1994.

Cloning Vectors: A Laboratory Manual (P. H. Pouwels *et al.*, 1985, Suppl. 1987).

Gram *et al.*, *Proc. Natl. Acad. Sci., USA*, 89:3576-80, 1992.

Hu *et al.*, *Cancer Res.*, 56:3055-61, 1996.

Kyte and Doolittle, *J. Mol. Biol.*, 157(1):105-32, 1982.

Liu *et al.*, *Cell Mol. Biol.*, 49(2):209-16, 2003.

Marks *et al.*, *J. Biol. Chem.*, 267:16007-10 1992.

Schier *et al.*, *J. Mol. Biol.*, 263:551-67, 1996.

Stemmer, *Nature*, 370:389-91, 1994.

Thyer *et al.*, *J. Am. Chem. Soc.*, 137(1):46-49, 2015.

WHAT IS CLAIMED IS:

1. A composition comprising purified recombinant polypeptides, said polypeptides comprising at least one selenocysteine residue at a selected position not found a wild type version of the polypeptide, wherein at least 80% of the recombinant polypeptides in the composition comprise the selenocysteine residue at the selected position.

2. The composition of claim 1, wherein 80%-99.9% of the recombinant polypeptides in the composition comprise the selenocysteine residue at the selected position.

3. The composition of claim 1, wherein 90%-99% of the recombinant polypeptides in the composition comprise the selenocysteine residue at the selected position.

4. The composition of claim 1, wherein at least 95% of the recombinant polypeptides in the composition comprise the selenocysteine residue at the selected position.

5. The composition of claim 1, wherein at least 99% of the recombinant polypeptides in the composition comprise the selenocysteine residue at the selected position.

6. The composition of any one of claims 1-5, wherein the recombinant polypeptides are at least 90% identical to a human polypeptide.

7. The composition of claim 6, wherein the human polypeptide is a polypeptide involved in a disease.

8. The composition of any one of claims 1-5, wherein the recombinant polypeptides comprise an antibody or an enzyme.

9. The composition of any one of claims 1-5, wherein the polypeptides comprise at least two selenocysteine residues at selected positions.

10. The composition of claim 9, wherein the two selenocysteine residues at the selected positions form a diselenide bond.

11. The composition of claim 10, wherein the polypeptides comprise an enzyme or an antibody.

12. The composition of claim 11, wherein the antibody is an aglycosylated antibody.

13. The composition of any one of claims 1-5, comprising at least 50 µg of the purified recombinant polypeptides.

14. The composition of any one of claims 1-5, wherein purified recombinant polypeptides are 97% to 99.9% pure.

15. The composition of any one of claims 1-11, wherein the polypeptide comprises 2, 3, 4, 5, 6, 7, 8, 9 or 10 selenocysteine residues at selected positions.

16. A pharmaceutical composition comprising a composition according to any one of claims 1-11.

17. A method of treating a subject comprising administering an effective amount of a pharmaceutical composition in accordance with claim 16 to the subject.

18. A nucleic acid molecule encoding a polypeptide comprising an amino acid sequence at least 90% identical to:

(a) SEQ ID NO:1 and having an amino acid substitution or deletion at a position corresponding to position 344 of SEQ ID NO:1;

(b) SEQ ID NO:2 and having an amino acid substitution or deletion at a position corresponding to position 702 of SEQ ID NO:2;

(c) SEQ ID NO:3 and having an amino acid substitution or deletion at a position corresponding to position 655 of SEQ ID NO:3;

(d) SEQ ID NO:4 and having an amino acid substitution or deletion at a position corresponding to position 73 of SEQ ID NO:4;

(e) SEQ ID NO:5 and having an amino acid substitution or deletion at a position corresponding to position 781 of SEQ ID NO:5;

(f) SEQ ID NO:6 and having an amino acid substitution or deletion at a position corresponding to position 136 of SEQ ID NO:6;

(g) SEQ ID NO:7 and having an amino acid substitution or deletion at a position corresponding to position 183 of SEQ ID NO:7;

(h) SEQ ID NO:8 and having an amino acid substitution or deletion at a position corresponding to position 1 of SEQ ID NO:8;

(i) SEQ ID NO:9 and having an amino acid substitution or deletion at a position corresponding to position 102 of SEQ ID NO:9;

(j) SEQ ID NO:10 and having an amino acid substitution or deletion at a position corresponding to position 105 of SEQ ID NO:10;

(k) SEQ ID NO:11 and having an amino acid substitution or deletion at a position corresponding to position 673 of SEQ ID NO:11;

(l) SEQ ID NO:12 and having an amino acid substitution or deletion at a position corresponding to position 69 of SEQ ID NO:12;

(m) SEQ ID NO:13 and having an amino acid substitution or deletion at a position corresponding to position 107 of SEQ ID NO:13;

(n) SEQ ID NO:17 and having an amino acid substitution or deletion at a position corresponding to position 246 of SEQ ID NO:17;

(o) SEQ ID NO:18 and having an amino acid substitution or deletion at a position corresponding to position 545 of SEQ ID NO:18; and/or

(p) SEQ ID NO:19 and having an amino acid substitution or deletion at a position corresponding to position 124 of SEQ ID NO:19.

19. A recombinant polypeptide comprising an amino acid sequence at least 90% identical to:

(a) SEQ ID NO:1 and having an amino acid substitution or deletion at a position corresponding to position 344 of SEQ ID NO:1;

(b) SEQ ID NO:2 and having an amino acid substitution or deletion at a position corresponding to position 702 of SEQ ID NO:2;

(c) SEQ ID NO:3 and having an amino acid substitution or deletion at a position corresponding to position 655 of SEQ ID NO:3;

(d) SEQ ID NO:4 and having an amino acid substitution or deletion at a position corresponding to position 73 of SEQ ID NO:4;

(e) SEQ ID NO:5 and having an amino acid substitution or deletion at a position corresponding to position 781 of SEQ ID NO:5;

(f) SEQ ID NO:6 and having an amino acid substitution or deletion at a position corresponding to position 136 of SEQ ID NO:6;

(g) SEQ ID NO:7 and having an amino acid substitution or deletion at a position corresponding to position 183 of SEQ ID NO:7;

(h) SEQ ID NO:8 and having an amino acid substitution or deletion at a position corresponding to position 1 of SEQ ID NO:8;

(i) SEQ ID NO:9 and having an amino acid substitution or deletion at a position corresponding to position 102 of SEQ ID NO:9;

(j) SEQ ID NO:10 and having an amino acid substitution or deletion at a position corresponding to position 105 of SEQ ID NO:10;

(k) SEQ ID NO:11 and having an amino acid substitution or deletion at a position corresponding to position 673 of SEQ ID NO:11;

(l) SEQ ID NO:12 and having an amino acid substitution or deletion at a position corresponding to position 69 of SEQ ID NO:12;

(m) SEQ ID NO:13 and having an amino acid substitution or deletion at a position corresponding to position 107 of SEQ ID NO:13;

(n) SEQ ID NO:17 and having an amino acid substitution or deletion at a position corresponding to position 246 of SEQ ID NO:17;

(o) SEQ ID NO:18 and having an amino acid substitution or deletion at a position corresponding to position 545 of SEQ ID NO:18; or

(p) SEQ ID NO:19 and having an amino acid substitution or deletion at a position corresponding to position 124 of SEQ ID NO:19.

20. The recombinant polypeptide of claim 19, wherein the polypeptide comprises a Pro substitution at a position corresponding to position 344 of SEQ ID NO:1.

21. The recombinant polypeptide of claim 19, wherein the polypeptide comprises a His substitution at a position corresponding to position 702 of SEQ ID NO:2.

22. The recombinant polypeptide of claim 19, wherein the polypeptide comprises an Ala substitution at a position corresponding to position 655 of SEQ ID NO:3.

23. The recombinant polypeptide of claim 19, wherein the polypeptide comprises an Ala substitution at a position corresponding to position 73 of SEQ ID NO:4.

24. The recombinant polypeptide of claim 19, wherein the polypeptide comprises a Gly substitution at a position corresponding to position 781 of SEQ ID NO:5.

25. The recombinant polypeptide of claim 19, wherein the polypeptide comprises a Val substitution at a position corresponding to position 136 of SEQ ID NO:6.

26. The recombinant polypeptide of claim 19, wherein the polypeptide comprises an Ala substitution at a position corresponding to position 183 of SEQ ID NO:7.

27. The recombinant polypeptide of claim 19, wherein the polypeptide comprises an Arg substitution at a position corresponding to position 1 of SEQ ID NO:8.

28. The recombinant polypeptide of claim 19, wherein the polypeptide comprises a Cys substitution at a position corresponding to position 102 of SEQ ID NO:9.

29. The recombinant polypeptide of claim 19, wherein the polypeptide comprises a Cys substitution at a position corresponding to position 105 of SEQ ID NO:10.

30. The recombinant polypeptide of claim 19, wherein the polypeptide comprises a Leu substitution at a position corresponding to position 673 of SEQ ID NO:11.

31. The recombinant polypeptide of claim 19, wherein the polypeptide comprises a Gly substitution at a position corresponding to position 69 of SEQ ID NO:12.

32. The recombinant polypeptide of claim 19, wherein the polypeptide comprises a Ser substitution at a position corresponding to position 107 of SEQ ID NO:13.

33. The recombinant polypeptide of claim 19, wherein the polypeptide comprises an Ala substitution at a position corresponding to position 246 of SEQ ID NO:17.

34. The recombinant polypeptide of claim 19, wherein the polypeptide comprises an Ile substitution at a position corresponding to position 545 of SEQ ID NO:18.

35. The recombinant polypeptide of claim 19, wherein the polypeptide comprises a Pro substitution at a position corresponding to position 124 of SEQ ID NO:19.

36. The recombinant polypeptide of claim 19, wherein the polypeptide further comprises a Thr substitution at a position corresponding to position 999, a Thr substitution at a position corresponding to position 457, a Pro substitution at a position corresponding to position 591, a Thr substitution at a position corresponding to position 183, a Leu substitution at a position corresponding to position 358, a Arg substitution at a position corresponding to position 23, a Ile substitution at a position corresponding to position 902, a Val substitution at a position corresponding to position 889, a Cys substitution at a position corresponding to position 620, and/or a Gly substitution at a position corresponding to position 174 of SEQ ID NO:1.

37. The recombinant polypeptide of claim 19, wherein the polypeptide further comprises a Cys substitution at a position corresponding to position 398, an Ala substitution at a position corresponding to position 652, a Cys substitution at a position corresponding to position 264, and/or an Ala substitution at a position corresponding to position 21 of SEQ ID NO:3.

38. The recombinant polypeptide of claim 19, wherein the polypeptide further comprises a Asn substitution at a position corresponding to position 45, a Leu substitution at a position corresponding to position 290, a Asp substitution at a position corresponding to position 271, a Tyr substitution at a position corresponding to position 153, a Val substitution at a position corresponding to position 45, a Pro substitution at a position corresponding to position 284, an Ile substitution at a position corresponding to position 73, a Leu substitution at a position corresponding to position 68, an Ile substitution at a position corresponding to position 69, a Cys substitution at a position corresponding to position 305, a Ser substitution at a position corresponding to position 144, and/or a Val substitution at a position corresponding to position 281 of SEQ ID NO:4.

39. The recombinant polypeptide of claim 19, wherein the polypeptide further comprises a Gly substitution at a position corresponding to position 916, a His substitution at a position corresponding to position 938, a His substitution at a position corresponding to position 860, a Asp substitution at a position corresponding to position 925, and/or a Met substitution at a position corresponding to position 470 of SEQ ID NO:5.

40. The recombinant polypeptide of claim 19, wherein the polypeptide further comprises an Ala substitution at a position corresponding to position 115, an Arg substitution at a position corresponding to position 386, an Arg substitution at a position corresponding to position 155, a Ser substitution at a position corresponding to position 98, an Ala substitution at a position corresponding to position 201, a Thr substitution at a position corresponding to position 294, a Tyr substitution at a position corresponding to position 159, and/or an Ile substitution at a position corresponding to position 112 of SEQ ID NO:6.

41. The recombinant polypeptide of claim 19, wherein the polypeptide further comprises a Gly substitution at a position corresponding to position 76, a Val substitution at a position corresponding to position 293, a Thr substitution at a position corresponding to position 637, a Val substitution at a position corresponding to position 3, a Ser substitution at a position corresponding to position 311, a Thr substitution at a position corresponding to position 471, a Val substitution at a position corresponding to position 228, a Ser substitution at a position corresponding to position 311, and/or a Thr substitution at a position corresponding to position 257 of SEQ ID NO: 18.

42. The recombinant polypeptide of claim 19, wherein the polypeptide further comprises a Val substitution at a position corresponding to position 173, a Leu substitution at a position corresponding to position 196, a Phe substitution at a position corresponding to position 180, an Ala substitution at a position corresponding to position 249, a Val substitution at a position corresponding to position 5, a Leu substitution at a position corresponding to position 273, and/or an Asn substitution at a position corresponding to position 176 of SEQ ID NO:7.

43. The recombinant polypeptide of claim 19, wherein the polypeptide further comprises a Cys substitution at a position corresponding to position 15 and/or a substitution at a position corresponding to position 30 of SEQ ID NO:9.

44. The recombinant polypeptide of claim 19, wherein the polypeptide further comprises a Ile substitution at a position corresponding to position 193, a Thr substitution at a position

corresponding to position 233, an Ala substitution at a position corresponding to position 300, and/or an Arg substitution at a position corresponding to position 199 of SEQ ID NO:19.

45. The recombinant polypeptide of claim 19, wherein the polypeptide further comprises an Ala substitution at a position corresponding to position 246 of SEQ ID NO:17.

46. The recombinant polypeptide of claim 19, wherein the polypeptide further comprises a Val substitution at a position corresponding to position 119, a Pro substitution at a position corresponding to position 535, an Arg substitution at a position corresponding to position 373, a Ser substitution at a position corresponding to position 535, a Thr substitution at a position corresponding to position 119, an Ala substitution at a position corresponding to position 601, a Lys substitution at a position corresponding to position 103, an Asp substitution at a position corresponding to position 31, an Ile substitution at a position corresponding to position 662, a Lys substitution at a position corresponding to position 359, and/or an Asp substitution at a position corresponding to position 519 of SEQ ID NO:11.

47. A nucleic acid molecule encoding a polypeptide comprising: (i) an amino acid sequence at least 90% identical to SEQ ID NO:14, and (ii) an Ile substitution at a position corresponding to position 212, an Asn substitution at a position corresponding to position 162, an Ala substitution at a position corresponding to position 299, and/or a Arg substitution at a position corresponding to position 220.

48. A recombinant polypeptide comprising: (i) an amino acid sequence at least 90% identical to SEQ ID NO:14, and (ii) an Ile substitution at a position corresponding to position 212, an Asn substitution at a position corresponding to position 162, an Ala substitution at a position corresponding to position 299, and/or a Arg substitution at a position corresponding to position 220.

49. A nucleic acid molecule encoding a polypeptide comprising: (i) an amino acid sequence at least 90% identical to SEQ ID NO:15, and (ii) an Ala substitution at a position corresponding to position 184, an Asn substitution at a position corresponding to position 1730, an Asp substitution at a position corresponding to position 1888, a Thr substitution at a position corresponding to position 352, a Ser substitution at a position corresponding to position 374, an Arg substitution at a position corresponding to position 1423, a Glu substitution at a position corresponding to position 1502, a Gln substitution at a position corresponding to position 285, a Lys substitution at a position corresponding to position 470,

a Thr substitution at a position corresponding to position 939, a Phe substitution at a position corresponding to position 1669, an Ile substitution at a position corresponding to position 2034, an Asn substitution at a position corresponding to position 1713, a Thr substitution at a position corresponding to position 704, and/or an Ile substitution at a position corresponding to position 1084.

50. A recombinant polypeptide comprising: (i) an amino acid sequence at least 90% identical to SEQ ID NO:15, and (ii) an Ala substitution at a position corresponding to position 184, an Asn substitution at a position corresponding to position 1730, an Asp substitution at a position corresponding to position 1888, a Thr substitution at a position corresponding to position 352, a Ser substitution at a position corresponding to position 374, an Arg substitution at a position corresponding to position 1423, a Glu substitution at a position corresponding to position 1502, a Gln substitution at a position corresponding to position 285, a Lys substitution at a position corresponding to position 470, a Thr substitution at a position corresponding to position 939, a Phe substitution at a position corresponding to position 1669, an Ile substitution at a position corresponding to position 2034, an Asn substitution at a position corresponding to position 1713, a Thr substitution at a position corresponding to position 704, and/or an Ile substitution at a position corresponding to position 1084.

51. A nucleic acid molecule encoding a polypeptide comprising: (i) an amino acid sequence at least 90% identical to SEQ ID NO:16, and (ii) a Phe substitution at a position corresponding to position 112, an Ala substitution at a position corresponding to position 126, an Leu substitution at a position corresponding to position 978, a Gly substitution at a position corresponding to position 199, a Thr substitution at a position corresponding to position 476, and/or a Val substitution at a position corresponding to position 735.

52. A recombinant polypeptide comprising: (i) an amino acid sequence at least 90% identical to SEQ ID NO:16, and (ii) a Phe substitution at a position corresponding to position 112, an Ala substitution at a position corresponding to position 126, an Leu substitution at a position corresponding to position 978, a Gly substitution at a position corresponding to position 199, a Thr substitution at a position corresponding to position 476, and/or a Val substitution at a position corresponding to position 735.

53. A bacterial strain comprising at least one nucleic acid molecule selected from those of claims 18, 47, 49 or 51.

54. The bacterial strain of claim 53, wherein the bacterial strain expresses the polypeptide encoded by at least one nucleic acid molecule selected from those of claims 18, 47, 49 or 51.

55. The bacterial strain of claim 53, comprising at least 2, 3, 4, 5, 6, 7, 8, 9 or 10 of said nucleic acid molecules.

56. The bacterial strain of claim 55, wherein the bacterial strain expresses the polypeptide encoded by at least 2, 3, 4, 5, 6, 7, 8, 9 or 10 of said nucleic acid molecules.

57. The bacterial strain of claim 55, wherein the bacterial strain is an *E. coli* strain.

58. The bacterial strain of claim 53, comprising a nucleic acid that encodes a tRNA that is at least 90% identical to SEQ ID NO: 20, SEQ ID NO: 21, or SEQ ID NO: 22 and comprising one or more of the following features:

(i) a G or C at a position corresponding to position 7;

(ii) a T at a position corresponding to position 49;

(iii) a A or C at a position corresponding to position 50;

(iv) a T at a position corresponding to position 64;

(v) a G or A at a position corresponding to position 65; and/or

(vi) a G, T or C at a position corresponding to position 66.

59. The bacterial strain of claim 58, wherein the molecule encodes a tRNA comprising the sequence at least about 90% identical to SEQ ID NO: 20; and comprises one or more of the following features:

(i) a G at a position corresponding to position 7;

(ii) a T at a position corresponding to position 49;

(iii) a C at a position corresponding to position 50;

(iv) a T at a position corresponding to position 64;

(v) a G at a position corresponding to position 65; and/or

(vi) a C at a position corresponding to position 66.

60. The bacterial strain of claim 59, wherein the molecule comprises 2, 3, 4, 5 or 6 of said features (i)-(vi).

61. The bacterial strain of claim 58, wherein the molecule encodes a tRNA comprising a sequence at least 95% identical to SEQ ID NO: 20.

62. The bacterial strain of claim 61, wherein the molecule encodes a tRNA comprising the sequence of SEQ ID NO: 20.

63. The bacterial strain of claim 58, wherein the molecule encodes a tRNA comprising the sequence at least about 90% identical to SEQ ID NO: 21.

64. The bacterial strain of claim 63, wherein the molecule encodes a tRNA comprising the sequence of SEQ ID NO: 21.

65. The bacterial strain of claim 58, wherein the molecule encodes a tRNA comprising the sequence at least about 90% identical to SEQ ID NO: 22.

66. The bacterial strain of claim 65, wherein the molecule encode a tRNA comprising the sequence of SEQ ID NO: 22.

67. The bacterial strain of claim 53, further comprising an expressible nucleic sequence encoding a polypeptide of interest having at least one position in the coding sequence with a TAG codon for selenocysteine incorporation.

68. The bacterial strain of claim 67, wherein the expressible nucleic sequence encodes a human polypeptide.

69. The bacterial strain of claim 67, wherein the expressible nucleic sequence encodes an enzyme or an antibody.

70. The bacterial strain of claim 67, wherein the expressible nucleic sequence comprises a T7 RNA polymerase promoter.

71. The bacterial strain of claim 53, further comprising a nucleic acid sequence encoding T7 RNA polymerase.

72. An *E. coli* bacterial strain deposited at the NCIMB under deposit Accession No. 42595.

73. A culture of bacteria comprising a bacterial strain in accordance with any one of claims 53-72.

74. A culture of bacteria comprising said culture comprising an expressed recombinant polypeptide in an amount of 1 to 100 mg/L of the culture, said expressed recombinant polypeptide comprising at least one selenocysteine residue.

75. The culture of claim 74, comprising an expressed recombinant polypeptide in an amount of 10 to 40 mg/L of the culture, said expressed recombinant polypeptide comprising at least one selenocysteine residue.

76. The culture of claim 74 or 75, wherein the expressed recombinant polypeptide is a polypeptide in accordance with any one of claims

77. The culture of claim 74 or 75, wherein the expressed recombinant polypeptide comprises at least one selenocysteine residue at a selected position not found in a wild type version of the polypeptide.

78. The culture of claim 77, wherein at least 80% of the expressed recombinant polypeptides in the culture comprise the selenocysteine residue at the selected position.

79. The culture of claim 77, wherein 80%-99.9% of the recombinant polypeptides in the culture comprise the selenocysteine residue at the selected position.

80. The culture of claim 77, wherein 90%-99% of the expressed recombinant polypeptides in the culture comprise the selenocysteine residue at the selected position.

81. The culture of claim 77, wherein at least 95% of the expressed recombinant polypeptides in the culture comprise the selenocysteine residue at the selected position.

82. The culture of claim 77, wherein at least 99% of the expressed recombinant polypeptides in the culture comprise the selenocysteine residue at the selected position.

83. The culture of claim 74 or 75, wherein the expressed recombinant polypeptide is at least 90% identical to a human polypeptide.

84. The culture of claim 83, wherein the human polypeptide is a polypeptide involved in a disease.

85. The culture of claim 74 or 75, wherein the expressed recombinant polypeptide comprises an antibody or an enzyme.

86. The culture of claim 74 or 75, wherein the polypeptide comprises at least two selenocysteine residues at selected positions.

87. The culture of claim 86, wherein the two selenocysteine residues at the selected positions form a diselenide bond.

88. The culture of claim 86, wherein the polypeptide comprises 2, 3, 4, 5, 6, 7, 8, 9 or 10 selenocysteine residues at selected positions.

89. A polypeptide comprising at least a first selenocysteine residue purified from a culture according to any one of claims 74-88.

90. A method of expressing a polypeptide comprising at least one selenocysteine residue comprising:

(a) expressing a nucleic acid encoding the polypeptide in a bacterial strain according to any one of claims 53-72 and in the presence of selenium source; and

(b) purifying the recombinant polypeptide from the bacteria.

91. A recombinant polypeptide comprising at least one selenocysteine residue at a selected position produced by a method comprising:

(a) expressing a nucleic acid encoding the polypeptide in a bacterial strain according to any one of claims 53-72 and in the presence of selenium source; and

(b) purifying the recombinant polypeptide from the bacteria.

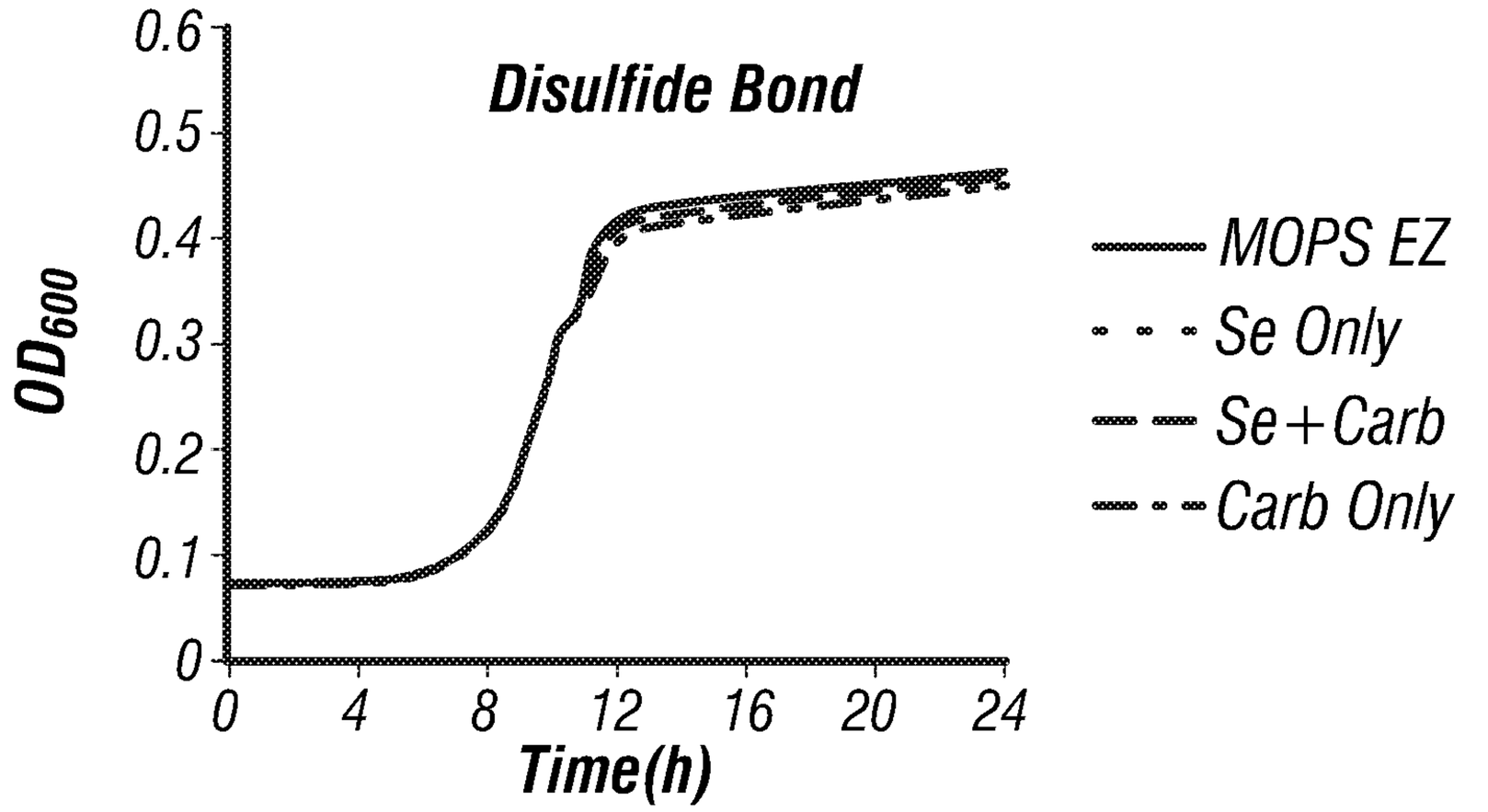

FIG. 1A

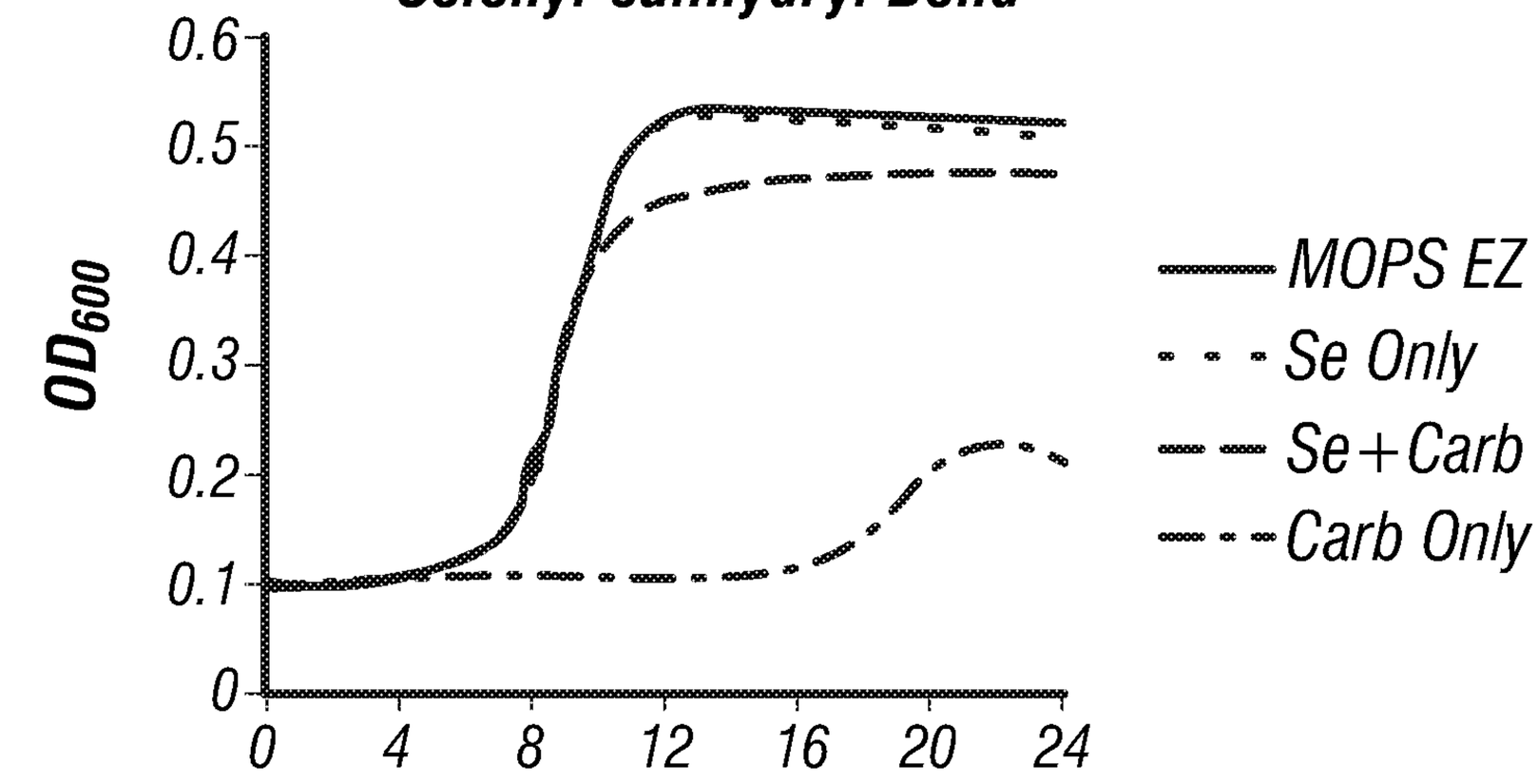

FIG. 1B

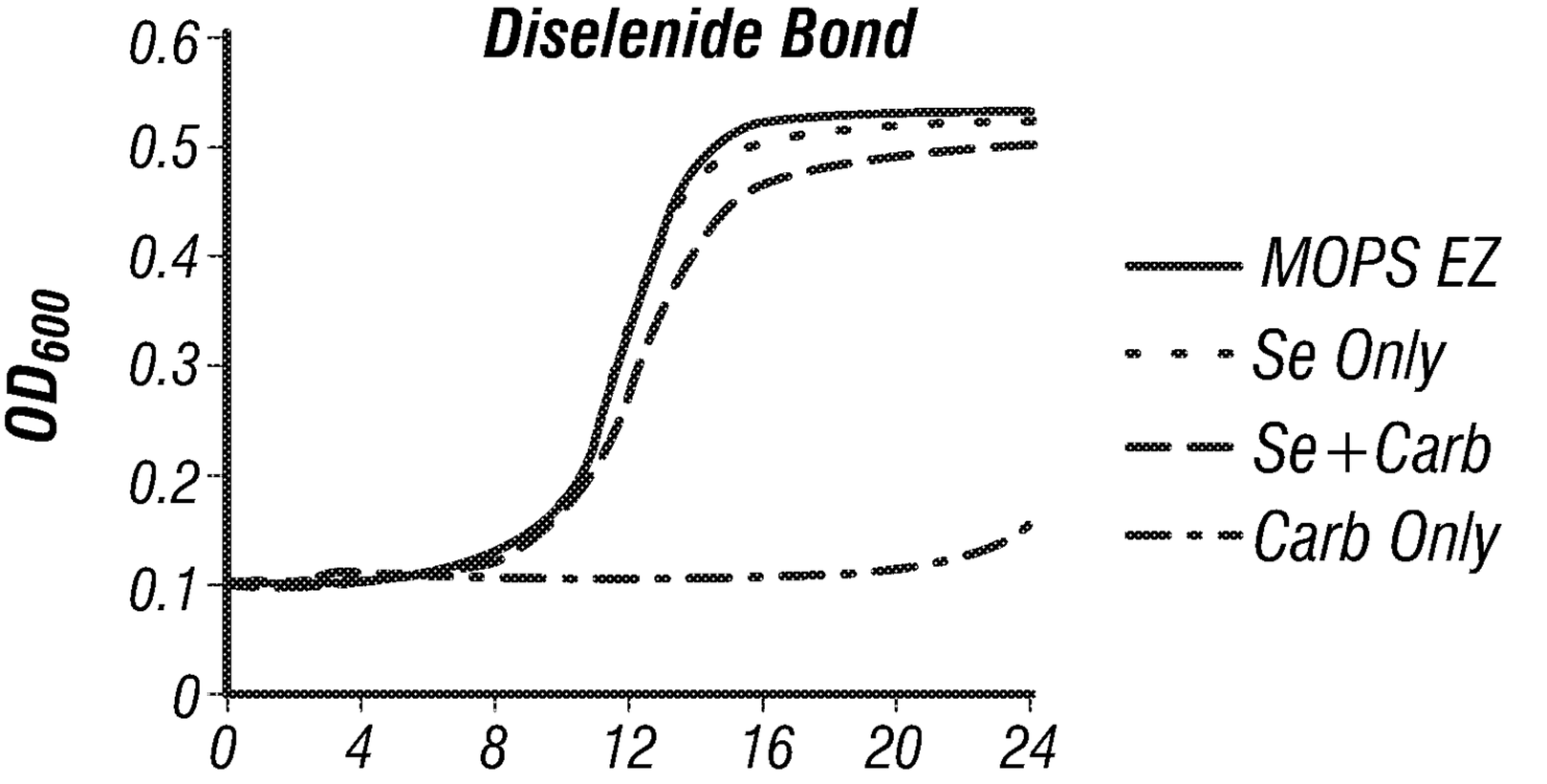

FIG. 1C

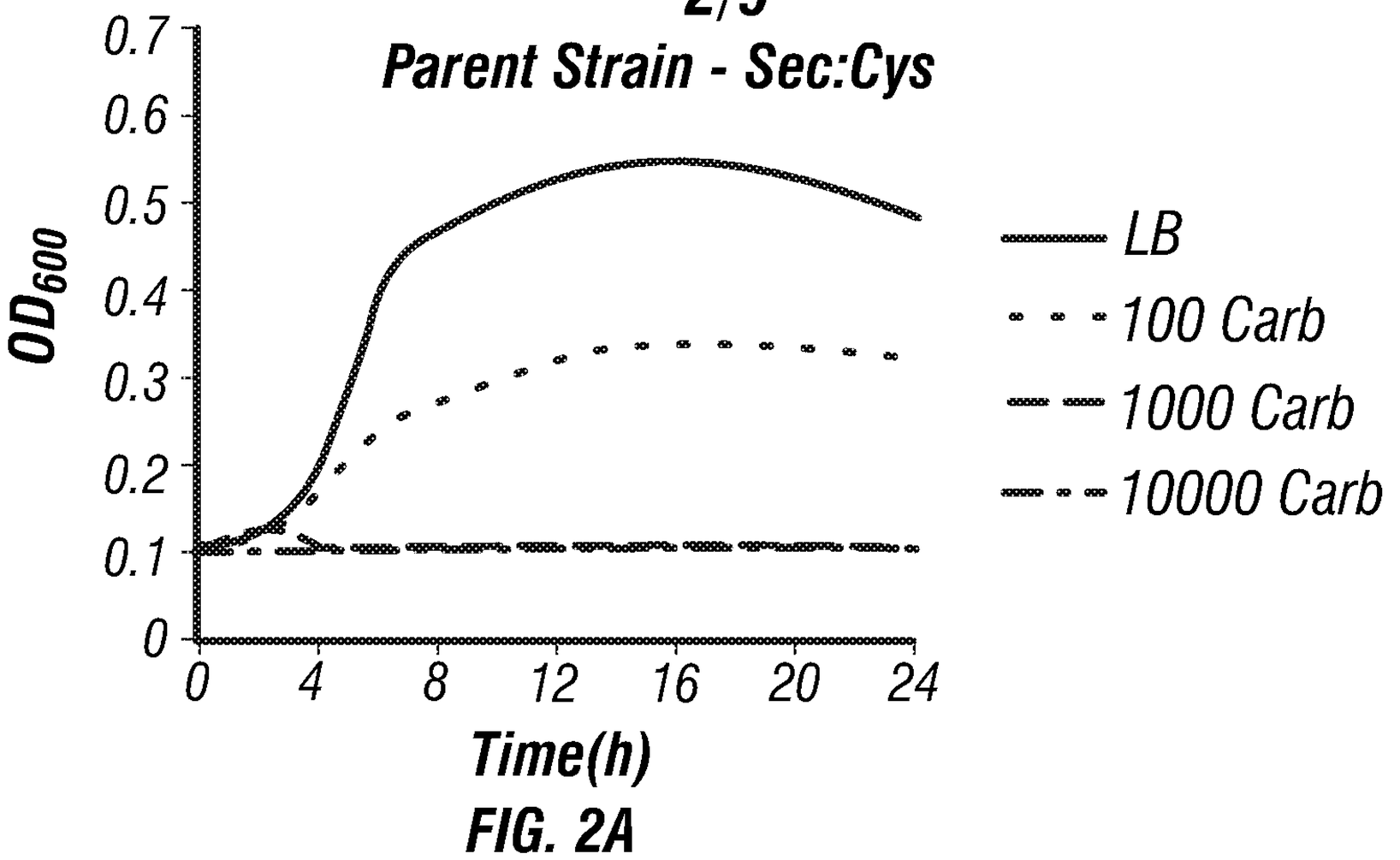

FIG. 2A

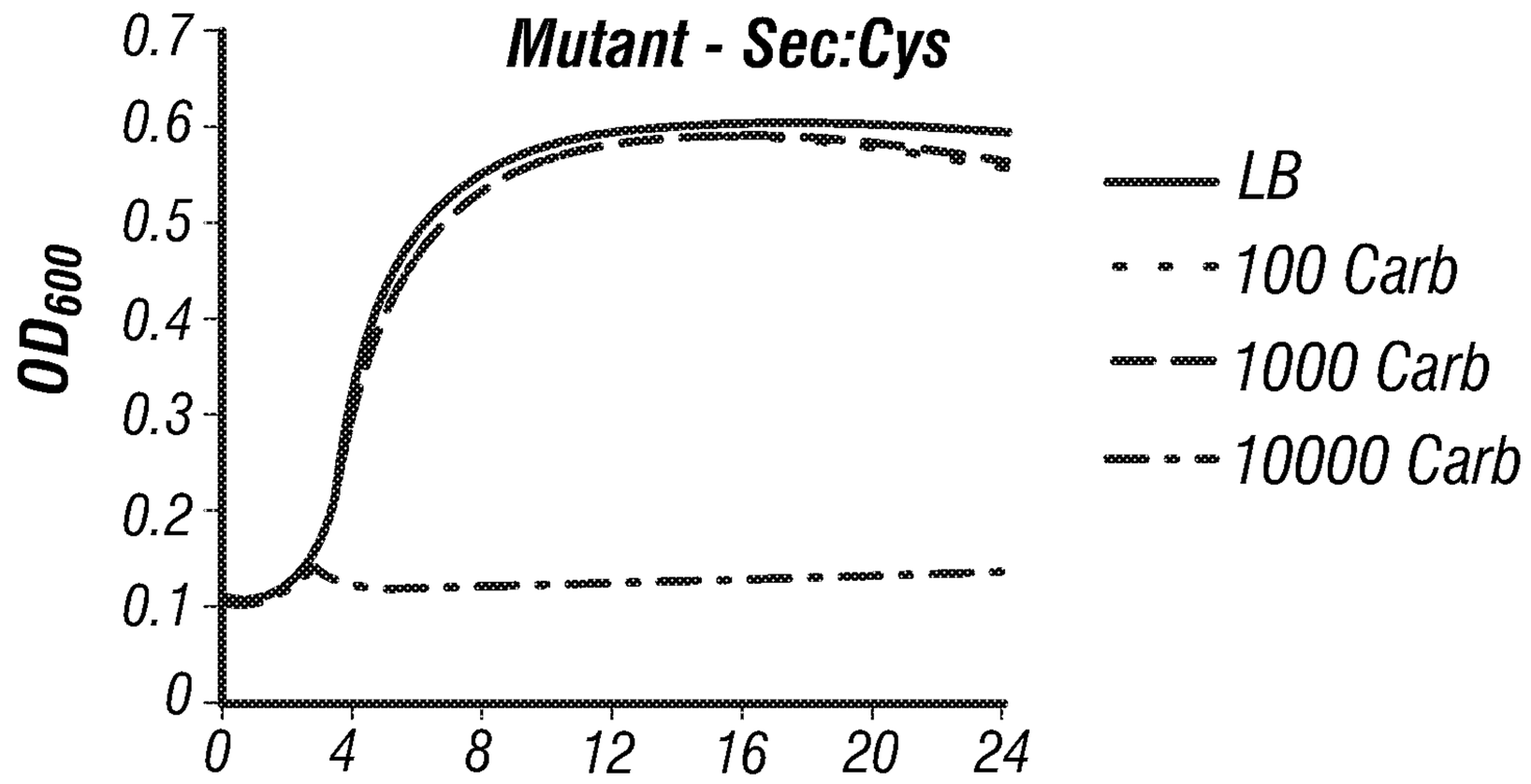

FIG. 2B

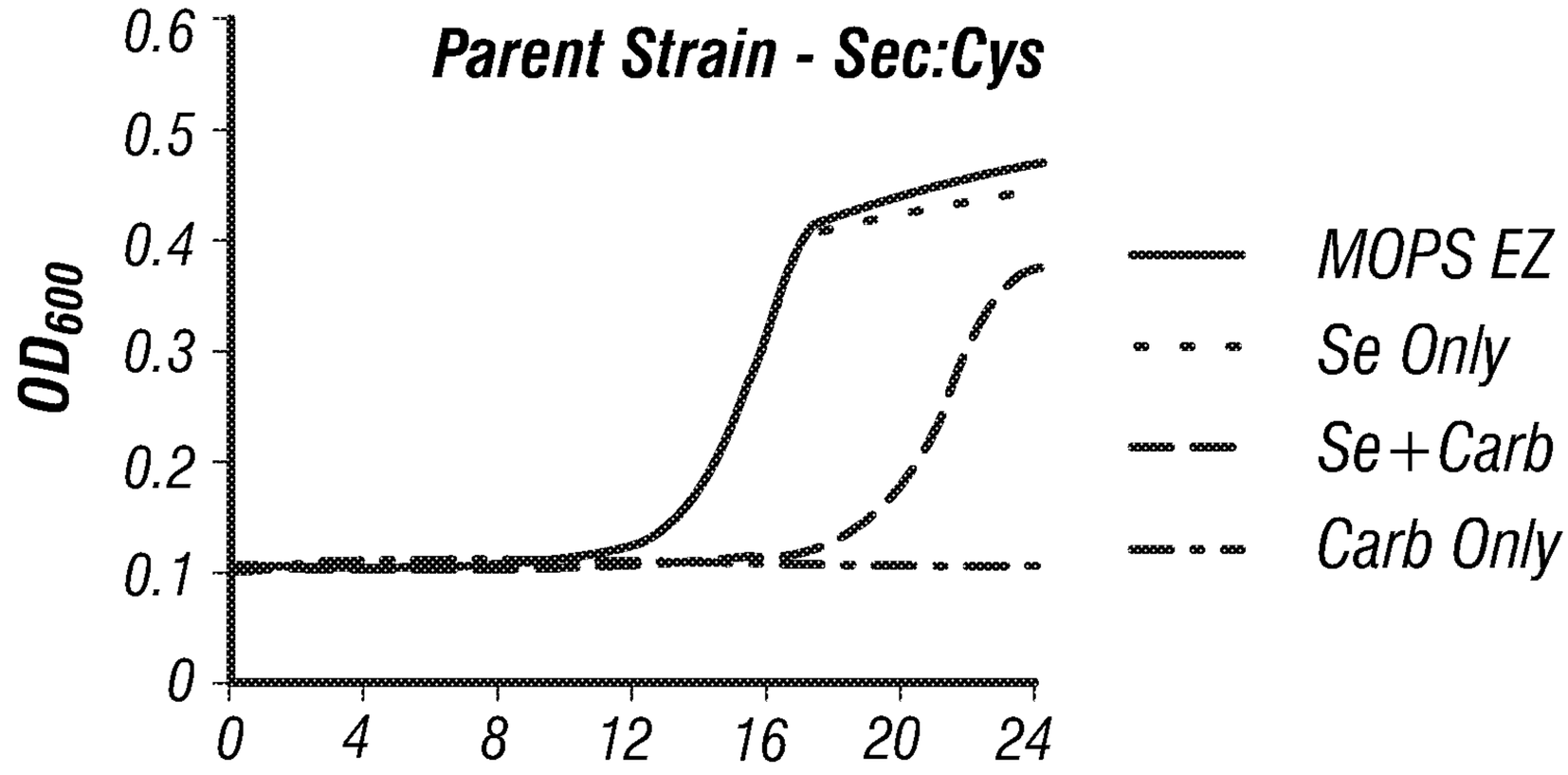

FIG. 3A

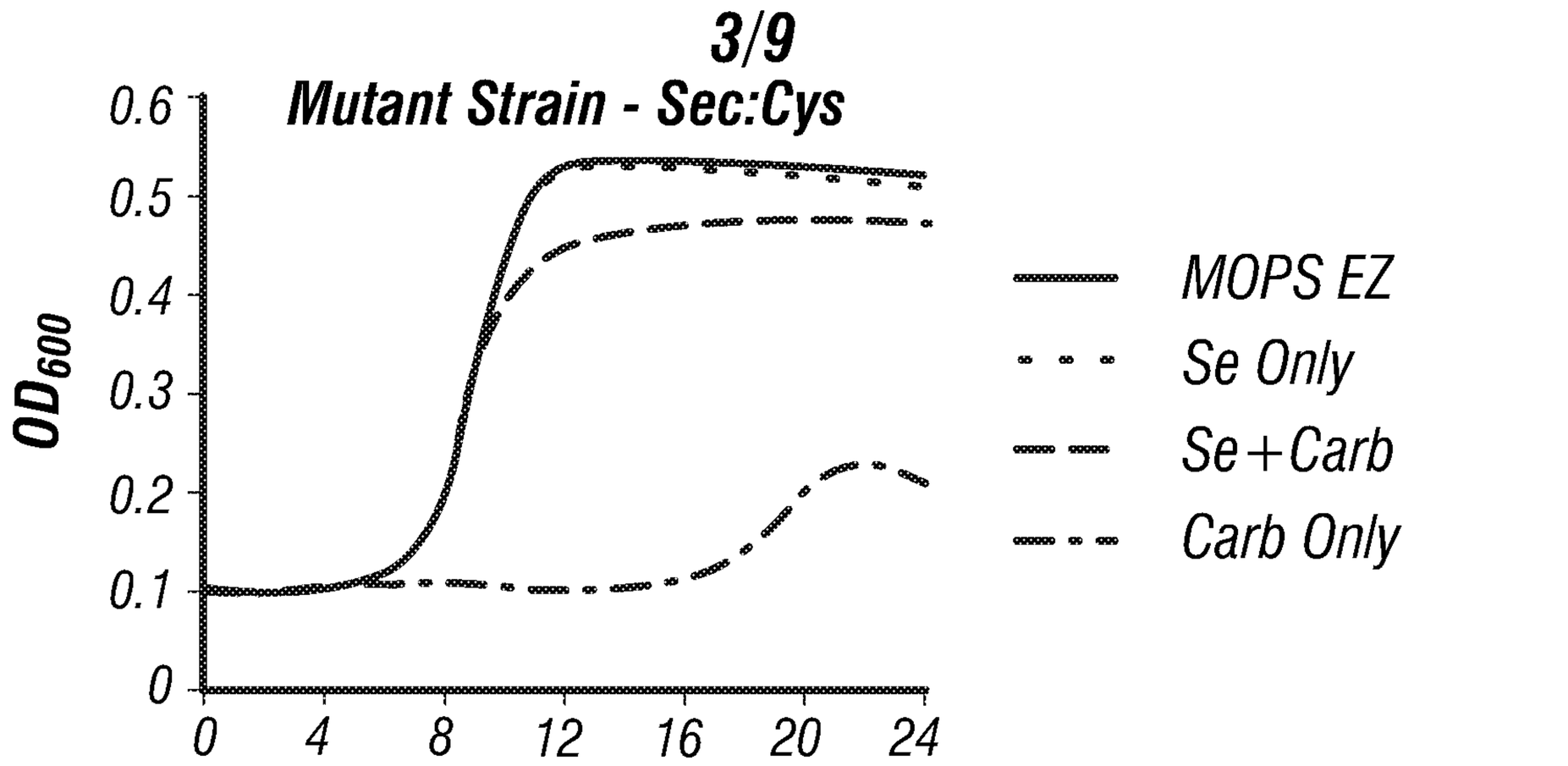

FIG. 3B

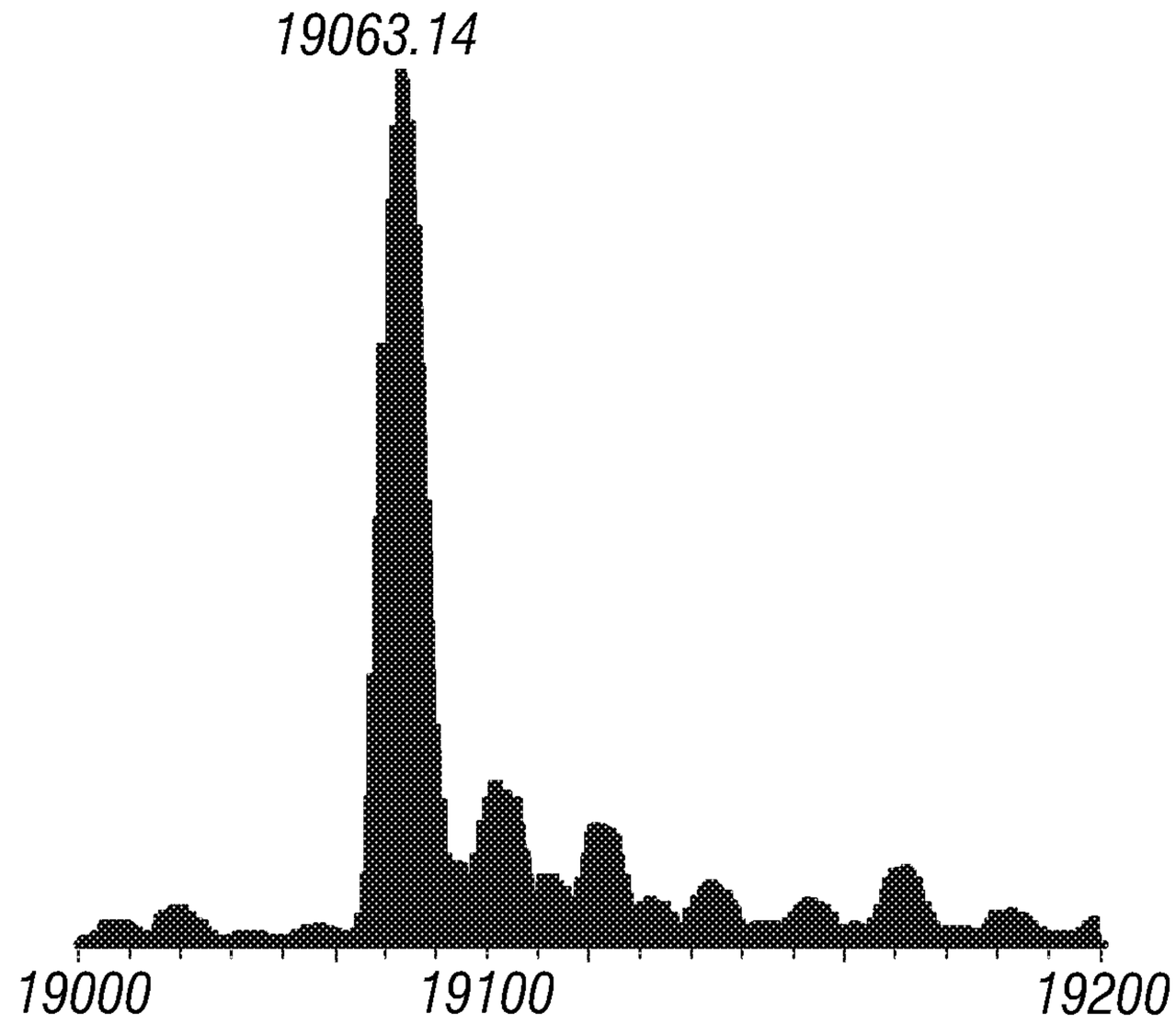

FIG. 4A

N	M	I	S	L	I	A	A	L	A	V	D	R	V	I	G	M	E	N	A	M	P	W	N	L	P	25
26	A	D	L	A	W	F	K	R	N	T	L	N	K	U	V	I	M	G	R	H	T	W	E	S	I	50
51	G	R	P	L	P	G	R	K	N	I	I	L	S	S	Q	P	G	T	D	D	R	V	T	W	V	75
76	K	S	V	D	E	A	I	A	A	U	G	D	V	P	E	I	M	V	I	G	G	G	R	V	Y	100
101	E	Q	F	L	P	K	A	Q	K	L	Y	L	T	H	I	D	A	E	V	E	G	D	T	H	F	125
126	P	D	Y	E	P	D	D	W	E	S	V	F	S	E	F	H	D	A	D	A	Q	N	S	H	S	150
151	Y	C	F	E	I	L	E	R	R	G	S	H	H	H	H	H	H	C								

FIG. 4B

	Theoretical Monoisotopic Mass(Da)	Experimental Monoisotopic Mass(Da)	Mass Accuracy (ppm)
anti-MS2 scFv (Ser)	31603.40		
anti-MS2 scFv (2xU,No Crosslink)	31855.10	31851.08	-125.57
anti-MS2 scFv (2xU, Crosslink)	31851.06	31851.08	+0.6

FIG. 5A

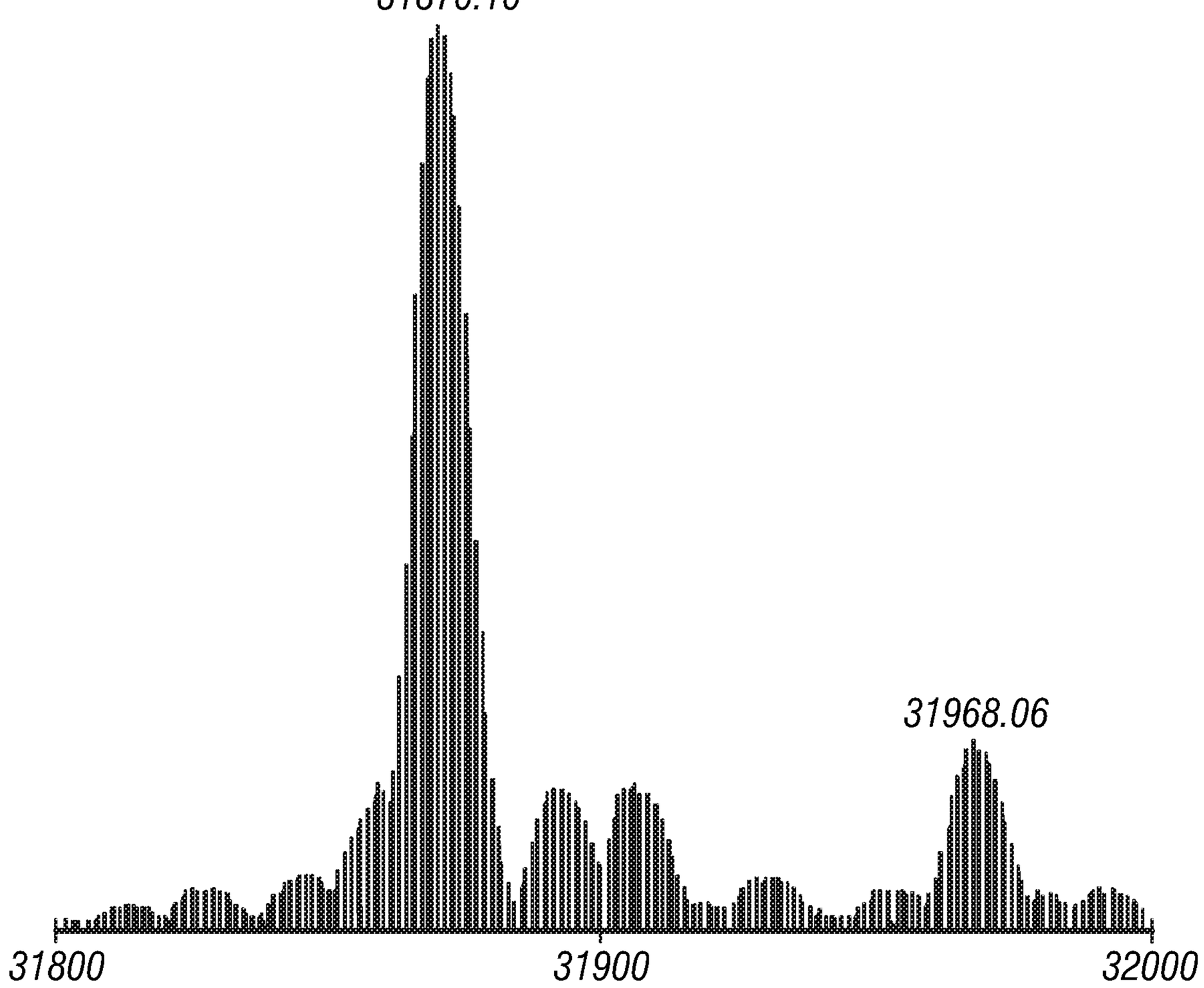

FIG. 5B

N M H H H H H H L V P R G S G K S A M A A E V K L V 25

26 E S G G G G L V K P G G S L K L S U A A S G F T F S 50

51 S Y A M S W V R Q T P E K R L E W V A T I S T G G 75

76 G Y T Y F P D S V K G R F T I S R D N A K N A L Y 100

101 L Q M K S L R S E D T A M Y Y U A R Q G D F G D W 125

126 Y F D V W G A G T T V T V S S G G G G S G G G G S 150

151 G G G G S T D V L M T Q T P L S L P V S L G D Q A 175

176 S I S U R S S Q S L V H S N G N T Y L H W Y L Q K 200

201 P G Q S P K L L I Y K V S N R F S G V P D R F S G 225

226 S G S G T D F T L K I S R V E A E D L G V Y F U S 250

251 Q S T H V P W T F G G G T K L E I K R A A A D Y K 275

276 D H D G D Y K D H D I D Y K D D D D K C

FIG. 5C

```
N M E V Q L L E S G G L A K P G G S L R L S C A A 25
26 S G F T F T D Y Y M D W V R Q A P G K G L E W V S 50
51 R I S P G G D V T W Y A D S V K G R F T I S R D N 75
76 A Q N T L Y L Q M N S L R A E D T A V Y F C A R D 100
101 D I V V S R I F D D W G Q G V L L T V S S G G G G 125
126 S G G G G S G G G G S E L Q M T Q S P S S L S A S 150
151 V G D R V T I T C R A S Q S I R S Y L A W Y Q Q K 175
176 P G K A P K L L I Y D A A H L Q S G V P S R F S G 200
201 S G S G T E F S L T I S S L Q P E D F A V Y Y C Q 225
226 Q R N S Y P L T F G G G T K V E I K R G S H H H H H H 250
251 H H C
```

FIG. 6A

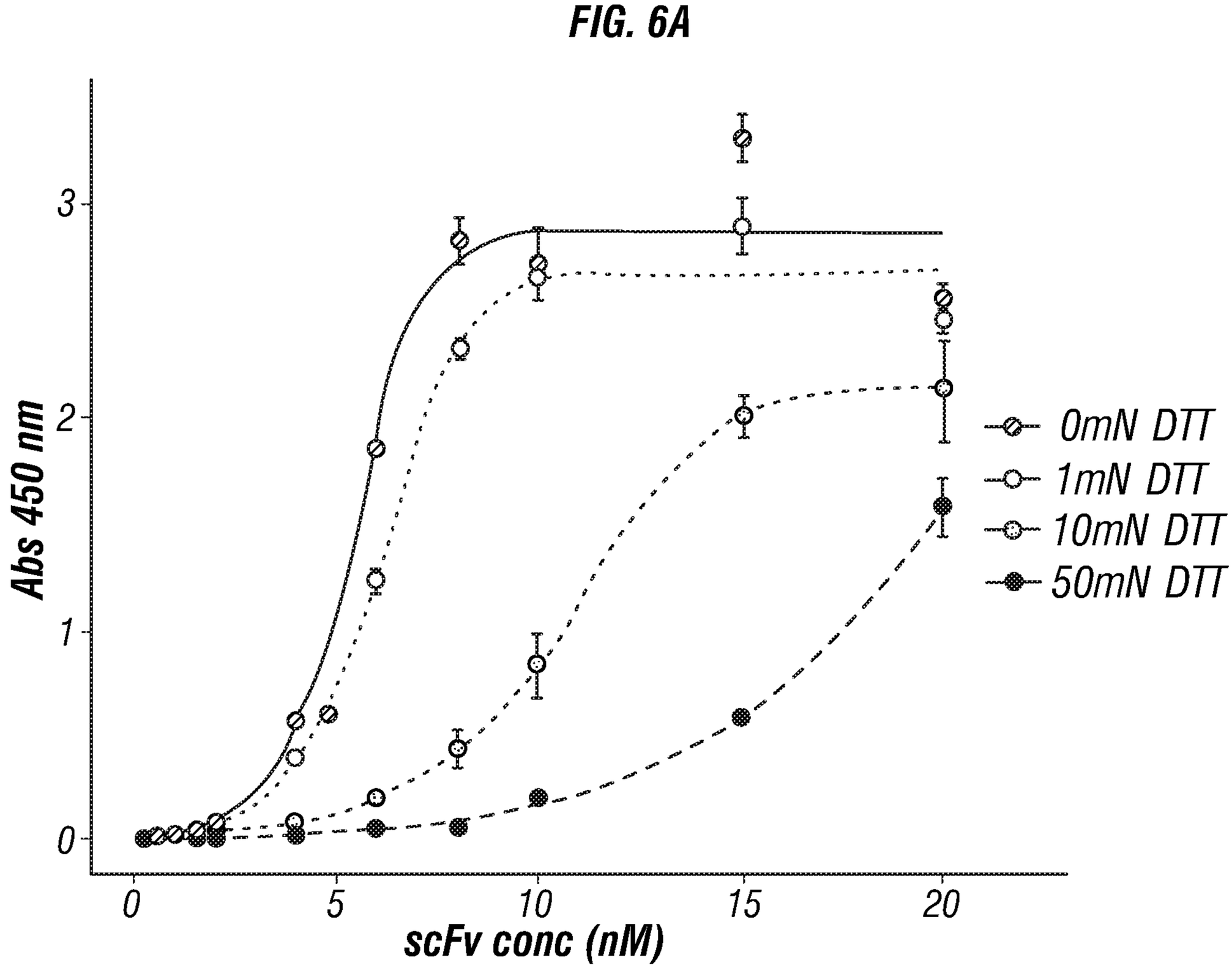

FIG. 6B

```
N M E V Q L L E S G G G L A K P G G S L R L S U A A 25
26 S G F T F T D Y Y M D W V R Q A P G K G L E W V S 50
51 R I S P G G D V T W Y A D S V K G R F T I S R D N 75
76 A Q N T L Y L Q M N S L R A E D T A V Y F U A R D 100
101 D I V V S R I F D D W G Q G V L L T V S S G G G G 125
126 S G G G G S G G G G S E L Q M T Q S P S S L S A S 150
151 V G D R V T I T U R A S Q S I R S Y L A W Y Q Q K 175
176 P G K A P K L L I Y D A A H L Q S G V P S R F S G 200
201 S G S G T E F S L T I S S L Q P E D F A V Y Y U Q 225
226 Q R N S Y P L T F G G G T K V E I K R G S H H H H H 250
251 H H C
```

FIG. 6C

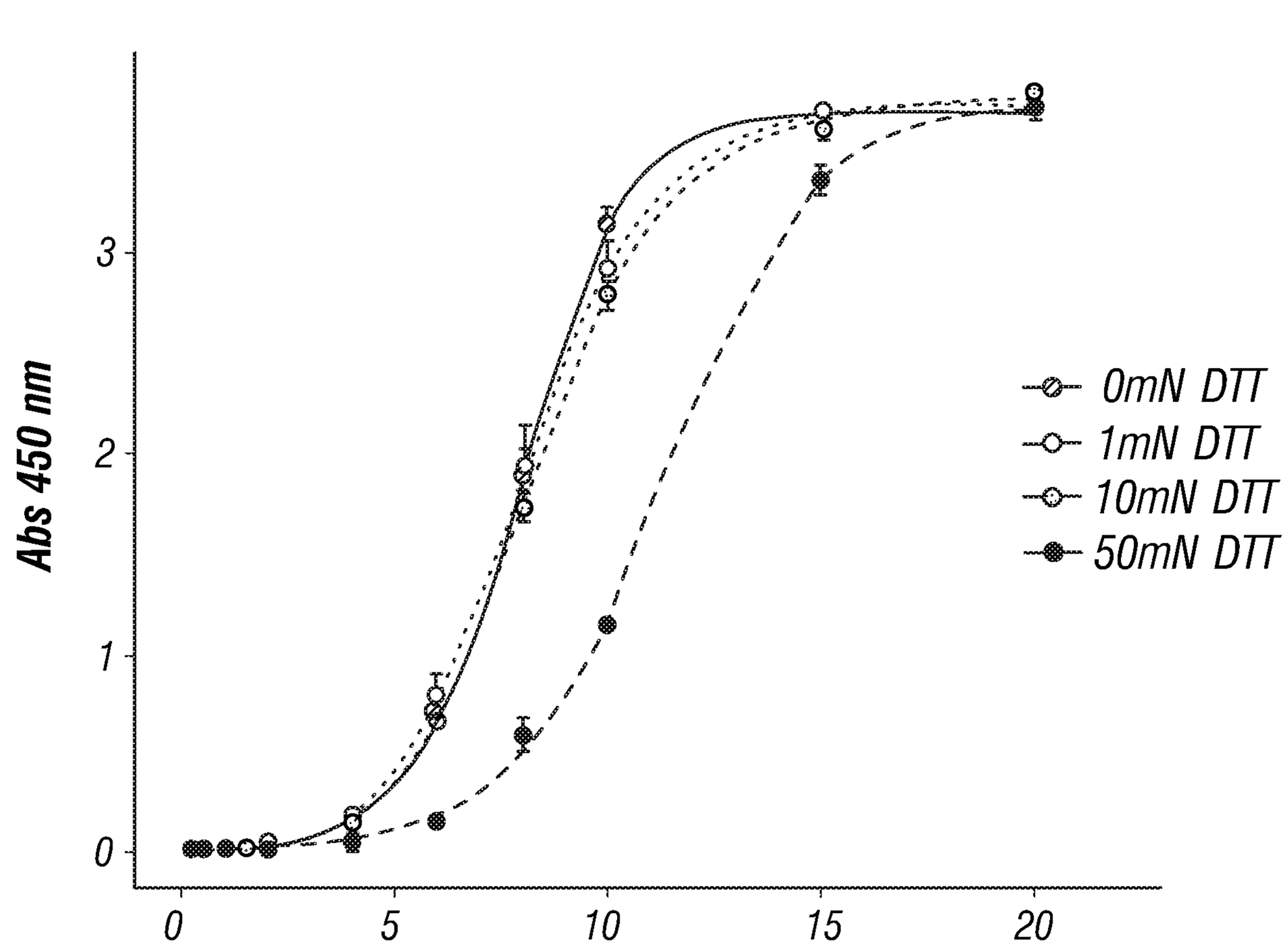

FIG. 6D

```
N M D I Q M T Q S P S S L S A S V G D R V T I T C R 25
26 A S Q D V N T A V A W Y Q Q K P G K A P K L L I Y 50
51 S A S F L Y S G V P S R F S G S R S G T D F T L T
I S S L Q P E D F A T Y Y C Q Q H Y T T P P T F G 100
101 Q G T K V E I K G A S G G G G S G G G G S G G G G
126 S S E V Q L V E S G G G L V Q P G G S L R L S C A 150
151 A S G F N I K D T Y I H W V R Q A P G K G L E W V 175
176 A R I Y P T N G Y T R Y A D S V K G R F T I S A D 200
201 T S K N T A Y L Q M N S L R A E D T A V Y Y C S R 225
226 W G G D G F Y A M D Y W G Q G T L V T V S S G S H 250
251 H H H H H C
```

FIG. 7A

```
N M D I Q M T Q S P S S L S A S V G D R V T I T C R 25
26 A S Q D V N T A V A W Y Q Q K P G K A P K L L I Y 50
51 S A S F L Y S G V P S R F S G S R S G T D F T L T
I S S L Q P E D F A T Y Y C Q Q H Y T T P P T F G 100
101 Q G T K V E I K G A S G G G G S G G G G S G G G G
126 S S E V Q L V E S G G G L V Q P G G S L R L S C A 150
151 A S G F N I K D T Y I H W V R Q A P G K G L E W V 175
176 A R I Y P T N G Y T R Y A D S V K G R F T I S A D 200
201 T S K N T A Y L Q M N S L R A E D T A V Y Y C S R 225
226 W G G D G F Y A M D Y W G Q G T L V T V S S G S H 250
251 H H H H H C
```

FIG. 7B

N M D I Q M T Q S P S S L S A S V G D R V T I T U R 25

26 A S Q D V N T A V A W Y Q Q K P G K A P K L L I Y 50

51 S A S F L Y S G V P S R F S G S R S G T D F T L T 75

76 I S S L Q P E D F A T Y Y U Q Q H Y T T P P T F G 100

101 Q G T K V E I K G A S G G G G S G G G G S G G G G 125

126 S S E V Q L V E S G G G L V Q P G G S L R L S U A 150

151 A S G F N I K D T Y I H W V R Q A P G K G L E W V 175

176 A R I Y P T N G Y T R Y A D S V K G R F T I S A D 200

201 T S K N T A Y L Q M N S L R A E D T A V Y Y U S R 225

226 W G G D G F Y A M D Y W G Q G T L V T V S S G S H 250

251 H H H H H C

FIG. 7C

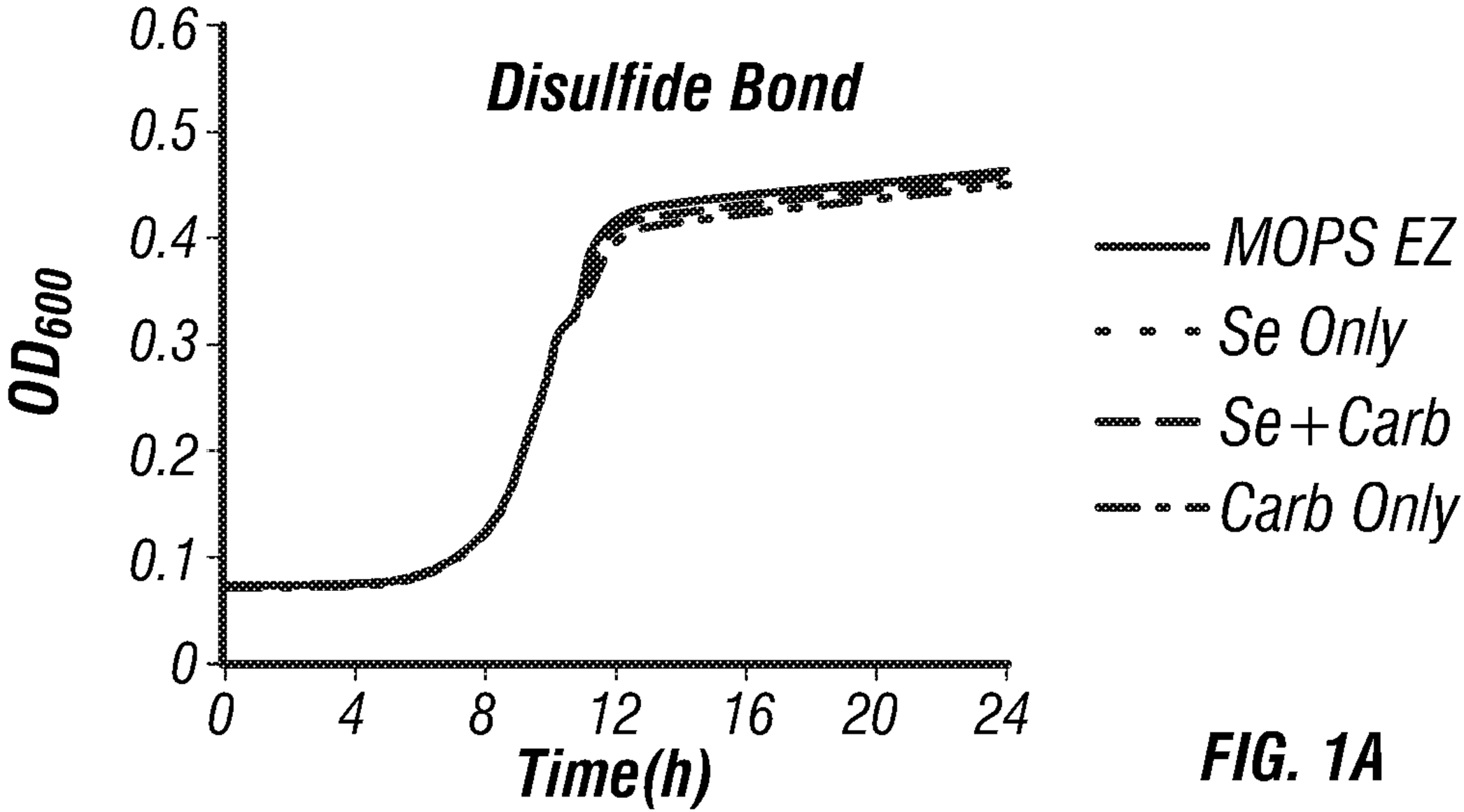
Disulfide Bond
OD_{600}
0.6
0.5
0.4
0.3
0.2
0.1
0
0
4
8
12
16
20
24
Time(h)
MOPS EZ
Se Only
Se+Carb
Carb Only

FIG. 1A