(12) STANDARD PATENT (11) Application No. **AU 2016280700 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Glucosyltransferase amino acid sequence motifs for producing linear poly alpha-1,3-glucan

(51) International Patent Classification(s)
C12P 19/04 (2006.01) ***C12P 19/18*** (2006.01)
C12N 9/10 (2006.01)

(21) Application No: **2016280700** (22) Date of Filing: **2016.06.15**

(87) WIPO No: **WO16/205401**

(30) Priority Data

(31) Number	(32) Date	(33) Country
62/180,788	**2015.06.17**	**US**
62/180,779	**2015.06.17**	**US**

(43) Publication Date: **2016.12.22**
(44) Accepted Journal Date: **2020.12.24**

(71) Applicant(s)
E I du Pont de Nemours and Company

(72) Inventor(s)
Payne, Mark S.;Brun, Yefim;Bott, Richard R.

(74) Agent / Attorney
Houlihan² Pty Ltd, PO Box 611, BALWYN NORTH, VIC, 3104, AU

(56) Related Art
US 20140087431 A1
GenBank Accession No. WP_013990740.1 27 May 2013

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 December 2016 (22.12.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/205401 A1

(51) International Patent Classification:
C12P 19/04 (2006.01) *C12N 9/10* (2006.01)
C12P 19/18 (2006.01)

(21) International Application Number:
PCT/US2016/037673

(22) International Filing Date:
15 June 2016 (15.06.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/180,779 17 June 2015 (17.06.2015) US
62/180,788 17 June 2015 (17.06.2015) US

(71) Applicant: **E I DU PONT DE NEMOURS AND COMPANY** [US/US]; Chestnut Run Plaza, 974 Centre Road, P.O. Box 2915, Wilmington, Delaware 19805 (US).

(72) Inventors: **PAYNE, Mark S.**; 4617 Old Linden Hill Road, Wilmington, Delaware 19808 (US). **BRUN, Yefim**; 500 Rockwood Road, Wilmington, Delaware 19802 (US). **BOTT, Richard R.**; 3032 Hillside Drive, Burlingame, California 94010 (US).

(74) Agent: **CHESIRE, Dennis**; E.I. du Pont de Nemours and Company, Legal Patent Records Center, Chestnut Run Plaza 721/2340, 974 Centre Road, PO Box 2915 Wilmington, Delaware 19805 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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,

[Continued on next page]

(54) Title: GLUCOSYLTRANSFERASE AMINO ACID SEQUENCE MOTIFS FOR PRODUCING LINEAR POLY ALPHA-1,3-GLUCAN

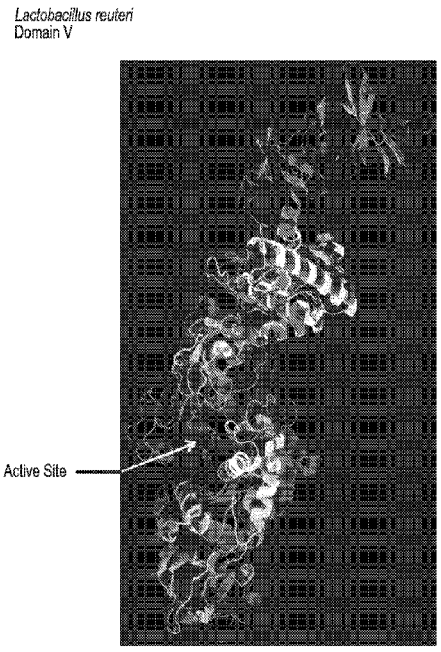

FIG. 1

(57) Abstract: Reactions comprising water, sucrose, and one or more glucosyltransferase enzymes are disclosed herein. Glucosyltransferase enzymes used in these reactions comprise certain motifs allowing production of insoluble poly alpha-1,3-glucan having at least 95% alpha-1,3 glycosidic linkages.

WO 2016/205401 A1

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))*

GLUCOSYLTRANSFERASE AMINO ACID SEQUENCE MOTIFS FOR PRODUCING LINEAR POLY ALPHA-1,3-GLUCAN

This application claims the benefit of U.S. Provisional Application Nos. 62/180,779 (filed June 17, 2015) and 62/180,788 (filed June 17, 2015), which are both incorporated herein by reference in their entirety.

FIELD OF INVENTION

The present disclosure is in the field of enzyme catalysis. For example, the disclosure pertains to producing linear poly alpha-1,3-glucan using a glucosyltransferase having certain amino acid sequence motifs.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

The official copy of the sequence listing is submitted electronically via EFS-Web as an ASCII formatted sequence listing with a file named 20160615_CL6452USNPSequenceListing_ST25_ExtraLinesRemoved.txt created on June 14, 2016, and having a size of 704 kilobytes and is filed concurrently with the specification. The sequence listing contained in this ASCII-formatted document is part of the specification and is herein incorporated by reference in its entirety.

BACKGROUND

Driven by a desire to find new structural polysaccharides using enzymatic syntheses or genetic engineering of microorganisms or plant hosts, researchers have discovered polysaccharides that are biodegradable and can be made economically from renewably sourced feedstocks. One such polysaccharide is poly alpha-1,3-glucan, a glucan polymer characterized by having alpha-1,3-glycosidic linkages. This polymer has been isolated by contacting an aqueous solution of sucrose with a glucosyltransferase (GTF) enzyme isolated from *Streptococcus salivarius* (Simpson et al., *Microbiology* 141:1451-1460, 1995).

U.S. Patent 7000000 disclosed the preparation of a polysaccharide fiber using an *S. salivarius* gtfJ enzyme. At least 50% of the hexose units within the polymer of this fiber were linked via alpha-1,3-glycosidic linkages. *S. salivarius* gtfJ enzyme utilizes sucrose as a substrate in a polymerization reaction producing poly alpha-1,3-glucan and fructose as end-products (Simpson et al., 1995). The disclosed polymer formed a liquid crystalline solution when it was dissolved above a critical concentration in a solvent or in

a mixture comprising a solvent. Continuous, strong, cotton-like fibers were obtained from this solution that could be spun and used in textile applications.

Not all glucosyltransferase enzymes can produce glucan with a molecular weight and percentage of alpha-1,3 glycosidic linkages suitable for use in spinning fibers. For example, most glucosyltransferase enzymes do not produce glucan having at least 50% alpha-1,3 glycosidic linkages and a number average degree of polymerization of at least 100. Therefore, it is desirable to identify glucosyltransferase enzymes that can convert sucrose to glucan polymers having a high percentage of alpha-1,3 glycosidic linkages and high molecular weight.

Reactions are disclosed herein that comprise glucosyltransferase enzymes containing certain amino acid motifs. These enzymes can synthesize high molecular weight, linear alpha-1,3-glucan polymer. Also disclosed are methods for identifying such enzymes.

SUMMARY OF INVENTION

In one embodiment, the disclosure concerns a reaction solution comprising water, sucrose, and a glucosyltransferase enzyme, wherein the glucosyltransferase enzyme comprises a catalytic domain comprising the following three motifs:

(i) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:78,

(ii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:79, and

(iii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:80;

wherein the glucosyltransferase enzyme does not comprise residues 54-957 of SEQ ID NO:65, residues 55-960 of SEQ ID NO:30, residues 55-960 of SEQ ID NO:4, residues 55-960 of SEQ ID NO:28, or residues 55-960 of SEQ ID NO:20; and wherein the glucosyltransferase enzyme produces insoluble poly alpha-1,3-glucan having at least 95% alpha-1,3 glycosidic linkages and a weight average degree of polymerization (DP_w) of at least 100.

In another embodiment, the catalytic domain comprises an amino acid sequence that is at least 90% identical to amino acid positions 54-957 of SEQ ID NO:65.

In another embodiment, the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:78 aligns with amino acid positions 231-243 of SEQ ID

NO:65; the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:79 aligns with amino acid positions 396-425 of SEQ ID NO:65; and/or the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:80 aligns with amino acid positions 549-567 of SEQ ID NO:65.

In another embodiment, motif (i) comprises SEQ ID NO:78, motif (ii) comprises SEQ ID NO:79, and motif (iii) comprises SEQ ID NO:80.

In another embodiment, the glucosyltransferase enzyme synthesizes poly alpha-1,3-glucan having 100% alpha-1,3 glycosidic linkages.

In another embodiment, the glucosyltransferase enzyme synthesizes poly alpha-1,3-glucan having a DP_w of at least 400.

Another embodiment of the disclosure concerns a method of producing insoluble poly alpha-1,3-glucan. This method comprises: (a) contacting at least water, sucrose, and a glucosyltransferase enzyme, wherein the glucosyltransferase enzyme comprises a catalytic domain comprising the following three motifs:

(i) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:78,

(ii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:79, and

(iii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:80,

wherein the glucosyltransferase enzyme does not comprise residues 54-957 of SEQ ID NO:65, residues 55-960 of SEQ ID NO:30, residues 55-960 of SEQ ID NO:4, residues 55-960 of SEQ ID NO:28, or residues 55-960 of SEQ ID NO:20; whereby insoluble poly alpha-1,3-glucan is produced having at least 95% alpha-1,3 glycosidic linkages and a weight average degree of polymerization (DP_w) of at least 100; and b) optionally, isolating the poly alpha-1,3-glucan produced in step (a).

In another embodiment, the catalytic domain comprises an amino acid sequence that is at least 90% identical to amino acid positions 54-957 of SEQ ID NO:65.

In another embodiment, the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:78 aligns with amino acid positions 231-243 of SEQ ID NO:65; the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:79 aligns with amino acid positions 396-425 of SEQ ID NO:65; and/or the position of

the amino acid sequence that is at least 90% identical to SEQ ID NO:80 aligns with amino acid positions 549-567 of SEQ ID NO:65.

In another embodiment, motif (i) comprises SEQ ID NO:78, motif (ii) comprises SEQ ID NO:79, and motif (iii) comprises SEQ ID NO:80.

In another embodiment, the glucosyltransferase enzyme synthesizes poly alpha-1,3-glucan having 100% alpha-1,3 glycosidic linkages.

In another embodiment, the glucosyltransferase enzyme synthesizes poly alpha-1,3-glucan having a DP_w of at least 400.

Another embodiment of the disclosure concerns a method of identifying a glucosyltransferase enzyme. This method comprises detecting the presence of at least one motif in a glucosyltransferase catalytic domain, the at least one motif selected from the group consisting of:

(i) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:78,

(ii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:79, and

(iii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:80;

thereby identifying a glucosyltransferase enzyme that produces insoluble poly alpha-1,3-glucan having at least 95% alpha-1,3 glycosidic linkages and a weight average degree of polymerization (DP_w) of at least 100.

In another embodiment, the detecting step is performed (a) *in silico*, (b) with a method comprising a nucleic acid hybridization step, (c) with a method comprising a protein sequencing step, and/or (d) with a method comprising a protein binding step.

In another embodiment, the detecting step comprises detecting the presence of each of motifs (i), (ii) and (iii) in the catalytic domain.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCES

FIG. 1: Comparison of the main chain tertiary fold of *Lactobacillus reuteri* GTF (gray) and *Streptococcus mutans* GTF (black). The structure of the *L. reuteri* GTF includes a fifth domain (Domain V) that was truncated from the structure of *S. mutans* GTF. The active site is also indicated and is formed by a cavity in the central domains (the so-called A and B domains); this location is based on spatial similarity with similar domains in alpha amylases. The amino acid sequence of the *S. mutans* 3AIE GTF

structure is SEQ ID NO:66, and the amino acid sequence of the *L. reuteri* 3KLK GTF structure is SEQ ID NO:67.

FIG. 2: Alignment of twenty-four GTF sequences with sequences of portions of GTFs from *S. mutans* (3AIE, SEQ ID NO:66) and *L. reuteri* (3KLK, SEQ ID NO:67) for which crystallographic structures are known; single-letter amino acid code is used. GTF amino acid sequences that produced glucan with 100% alpha-1,3 linkages and high molecular weight (DP_w of at least 400 under the tested initial sucrose concentrations, see Table 4) are designated "++". Those GTFs producing glucan with 100% alpha-1,3 linkages and a DP_w of at least 100 are designated "+-". Other GTFs producing glucan with mixed linkages are designated "--".

FIG. 3: The sequence of the tested GTF enzymes in the vicinity of Motifs 1a and 1b. The sequence region of Motifs 1a and 1b along with upstream and downstream flanking reference sequence motifs are shown as boxed regions. Motifs 1a and 1b are located in box labeled "Insertion 1". The alignment in this figure represents a portion of the alignment in FIG. 2.

FIG. 4a and 4b: Visualization of Motif 1a through comparison of a homology model of GTF 7527 (SEQ ID NO:65) based on the reference crystallographic structures of *S. mutans* (3AIE, SEQ ID NO:66) (FIG. 4a) and *L. reuteri* (3KLK, SEQ ID NO:67) (FIG. 4b). The main chain folding of the homology model in each view is shown with dark lines while the main chain folding of the reference structure is shown with lighter lines. The residues forming the catalytic sites in the reference crystallographic structures are shown as Van der Waals spheres for reference. Motif 1a (between the arrows) is presented in both homology models as an open loop (black) extending into the solvent as a consequence of there being no homologous segment to provide means to position with respect to the remainder of the GTF catalytic domain.

FIG. 5: The sequence of the tested GTF enzymes in the vicinity of Motif 2. The sequence region of Motif 2 along with upstream and downstream flanking reference sequence motifs are shown as boxed regions. Motif 2 is located in box labeled "Insertion 2". The alignment in this figure represents a portion of the alignment in FIG. 2.

FIG. 6a and 6b: Visualization of Motif 2 through comparison of a homology model of GTF 7527 (SEQ ID NO:65) based on the reference crystallographic structures of *S. mutans* (3AIE, SEQ ID NO:66) (FIG. 6a) and *L. reuteri* (3KLK, SEQ ID NO:67) (FIG. 6b). The main chain folding of the homology model in each view is shown with dark lines

while the main chain folding of the reference structure is shown with lighter lines. The residues forming the catalytic sites in the reference crystallographic structures are shown as Van der Waals spheres for reference. Motif 2 (between the arrows) is presented in both homology models as an open loop (black) extending into the solvent as a consequence of there being no homologous segment to provide means to position with respect to the remainder of the GTF catalytic domain.

FIG. 7: The sequence of the tested GTF enzymes in the vicinity of Motifs 3a and 3b. The sequence region of Motifs 3a and 3b along with upstream and downstream flanking reference sequence motifs are shown as boxed regions. Motifs 3a and 3b are located in box labeled "Insertion 3". The alignment in this figure represents a portion of the alignment in FIG. 2.

FIG. 8a and 8b: Visualization of Motif 3a through comparison of a homology model of GTF 7527 (SEQ ID NO:65) based on the reference crystallographic structures of *S. mutans* (3AIE, SEQ ID NO:66) (FIG. 8a) and *L. reuteri* (3KLK, SEQ ID NO:67) (FIG. 8b). The main chain folding of the homology model in each view is shown with dark lines while the main chain folding of the reference structure is shown with lighter lines. The residues forming the catalytic sites in the reference crystallographic structures are shown as Van der Waals spheres for reference. Motif 3a (between the arrows) is presented in both homology models as an open loop (black) extending into the solvent as a consequence of there being no homologous segment to provide means to position with respect to the remainder of the GTF catalytic domain.

Table 1. Summary of Nucleic Acid and Protein SEQ ID Numbers

Description	Nucleic acid SEQ ID NO.	Protein SEQ ID NO.
"0874 GTF", *Streptococcus sobrinus*. DNA codon-optimized for expression in *E. coli*. The first 156 amino acids of the protein are deleted compared to GENBANK Identification No. 450874; a start methionine is included.	1	2 (1435 aa)
"6855 GTF", *Streptococcus salivarius* SK126. DNA codon-optimized for expression in *E. coli*. The first 178 amino acids of the protein are deleted compared to GENBANK Identification No. 228476855; a start methionine is included.	3	4 (1341 aa)
"2379 GTF", *Streptococcus salivarius*. DNA codon-optimized for expression in *E. coli*. The first 203 amino acids of the protein are deleted compared to	5	6 (1247 aa)

GENBANK Identification No. 662379; a start methionine is included.		
"7527" or "GTFJ", *Streptococcus salivarius.* DNA codon-optimized for expression in *E. coli*. The first 42 amino acids of the protein are deleted compared to GENBANK Identification No. 47527; a start methionine is included.	7	8 (1477 aa)
"1724 GTF", *Streptococcus downei.* DNA codon-optimized for expression in *E. coli*. The first 162 amino acids of the protein are deleted compared to GENBANK Identification No. 121724; a start methionine is included.	9	10 (1436 aa)
"0544 GTF", *Streptococcus mutans.* DNA codon-optimized for expression in *E. coli*. The first 164 amino acids of the protein are deleted compared to GENBANK Identification No. 290580544; a start methionine is included.	11	12 (1313 aa)
"5926 GTF", *Streptococcus dentirousetti.* DNA codon-optimized for expression in *E. coli*. The first 144 amino acids of the protein are deleted compared to GENBANK Identification No. 167735926; a start methionine is included.	13	14 (1323 aa)
"4297 GTF", *Streptococcus oralis.* DNA codon-optimized for expression in *E. coli*. The first 228 amino acids of the protein are deleted compared to GENBANK Identification No. 7684297; a start methionine is included.	15	16 (1348 aa)
"5618 GTF", *Streptococcus sanguinis.* DNA codon-optimized for expression in *E. coli*. The first 223 amino acids of the protein are deleted compared to GENBANK Identification No. 328945618; a start methionine is included.	17	18 (1348 aa)
"2765 GTF", unknown *Streptococcus* sp. C150. DNA codon-optimized for expression in *E. coli*. The first 193 amino acids of the protein are deleted compared to GENBANK Identification No. 322372765; a start methionine is included.	19	20 (1340 aa)
"4700 GTF", *Leuconostoc mesenteroides.* DNA codon-optimized for expression in *E. coli*. The first 36 amino acids of the protein are deleted compared to GENBANK Identification No. 21654700; a start methionine is included.	21	22 (1492 aa)
"1366 GTF", *Streptococcus criceti.* DNA codon-optimized for expression in *E. coli*. The first 139 amino acids of the protein are deleted compared to GENBANK Identification No. 146741366; a start methionine is included.	23	24 (1323 aa)

"0427 GTF", *Streptococcus sobrinus.* DNA codon-optimized for expression in *E. coli*. The first 156 amino acids of the protein are deleted compared to GENBANK Identification No. 940427; a start methionine is included.	25	26 (1435 aa)
"2919 GTF", *Streptococcus salivarius* PS4. DNA codon-optimized for expression in *E. coli*. The first 92 amino acids of the protein are deleted compared to GENBANK Identification No. 383282919; a start methionine is included.	27	28 (1340 aa)
"2678 GTF", *Streptococcus salivarius* K12. DNA codon-optimized for expression in *E. coli*. The first 188 amino acids of the protein are deleted compared to GENBANK Identification No. 400182678; a start methionine is included.	29	30 (1341 aa)
"2381 GTF", *Streptococcus salivarius.* DNA codon-optimized for expression in *E. coli*. The first 273 amino acids of the protein are deleted compared to GENBANK Identification No. 662381; a start methionine is included.	31	32 (1305 aa)
"3929 GTF", *Streptococcus salivarius* JIM8777. DNA codon-optimized for expression in *E. coli*. The first 178 amino acids of the protein are deleted compared to GENBANK Identification No. 387783929; a start methionine is included.	33	34 (1341 aa)
"6907 GTF", *Streptococcus salivarius* SK126. DNA codon-optimized for expression in *E. coli*. The first 161 amino acids of the protein are deleted compared to GENBANK Identification No. 228476907; a start methionine is included.	35	36 (1331 aa)
"6661 GTF", *Streptococcus salivarius* SK126. DNA codon-optimized for expression in *E. coli*. The first 265 amino acids of the protein are deleted compared to GENBANK Identification No. 228476661; a start methionine is included.	37	38 (1305 aa)
"0339 GTF", *Streptococcus gallolyticus* ATCC 43143. DNA codon-optimized for expression in *E. coli*. The first 213 amino acids of the protein are deleted compared to GENBANK Identification No. 334280339; a start methionine is included.	39	40 (1310 aa)
"0088 GTF", *Streptococcus mutans.* DNA codon-optimized for expression in *E. coli*. The first 189 amino acids of the protein are deleted compared to GENBANK Identification No. 3130088; a start methionine is included.	41	42 (1267 aa)
"9358 GTF", *Streptococcus mutans* UA159. DNA codon-optimized for expression in *E. coli*. The first	43	44 (1287 aa)

176 amino acids of the protein are deleted compared to GENBANK Identification No. 24379358; a start methionine is included.		
"8242 GTF", *Streptococcus gallolyticus* ATCC BAA-2069. DNA codon-optimized for expression in *E. coli*. The first 191 amino acids of the protein are deleted compared to GENBANK Identification No. 325978242; a start methionine is included.	45	46 (1355 aa)
"3442 GTF", *Streptococcus sanguinis* SK405. DNA codon-optimized for expression in *E. coli*. The first 228 amino acids of the protein are deleted compared to GENBANK Identification No. 324993442; a start methionine is included.	47	48 (1348 aa)
"7528 GTF", *Streptococcus salivarius*. DNA codon-optimized for expression in *E. coli*. The first 173 amino acids of the protein are deleted compared to GENBANK Identification No. 47528; a start methionine is included.	49	50 (1427 aa)
"3279 GTF", *Streptococcus* sp. C150. DNA codon-optimized for expression in *E. coli*. The first 178 amino acids of the protein are deleted compared to GENBANK Identification No. 322373279; a start methionine is included.	51	52 (1393 aa)
"6491 GTF", *Leuconostoc citreum* KM20. DNA codon-optimized for expression in *E. coli*. The first 244 amino acids of the protein are deleted compared to GENBANK Identification No. 170016491; a start methionine is included.	53	54 (1262 aa)
"6889 GTF", *Streptococcus salivarius* SK126. DNA codon-optimized for expression in *E. coli*. The first 173 amino acids of the protein are deleted compared to GENBANK Identification No. 228476889; a start methionine is included.	55	56 (1427 aa)
"4154 GTF", *Lactobacillus reuteri*. DNA codon-optimized for expression in *E. coli*. The first 38 amino acids of the protein are deleted compared to GENBANK Identification No. 51574154.	57	58 (1735 aa)
"3298 GTF", *Streptococcus* sp. C150. The first 209 amino acids of the protein are deleted compared to GENBANK Identification No. 322373298; a start methionine is included.		59 (1242 aa)
Wild type GTFJ, *Streptococcus salivarius.* GENBANK Identification No. 47527.		60 (1518 aa)
Wild type GTF corresponding to 2678 GTF, *Streptococcus salivarius* K12.		61 (1528 aa)
Wild type GTF corresponding to 6855 GTF,		62 (1518 aa)

Streptococcus salivarius SK126.		
Wild type GTF corresponding to 2919 GTF, *Streptococcus salivarius* PS4.		63 (1431 aa)
Wild type GTF corresponding to 2765 GTF, *Streptococcus* sp. C150.		64 (1532 aa)
Shorter version of 7527, *Streptococcus salivarius*, (also referred to as "7527-NT" herein. The first 178 amino acids of the protein are deleted compared to GENBANK Identification No. 47527; a start methionine is included.		65 (1341 aa)
"3AIE", portion of a GTF from *Streptococcus mutans* as annotated in the Protein Data Bank under pdb entry no. 3AIE.		66 (844 aa)
"3KLK", portion of a GTF from *Lactobacillus reuteri* as annotated in the Protein Data Bank under pdb entry no. 3KLK.		67 (1039 aa)
Catalytic active site motif FDxxRxDAxDNV		68 (12 aa)
Catalytic active site motif ExWxxxDxxY		69 (10 aa)
Catalytic active site motif FxRAHD		70 (6 aa)
Catalytic active site motif IxNGYAF		71 (7 aa)
Motif SxxRxxN upstream of Motifs 1a and 1b		72 (7 aa)
Motif GGxxxLLxNDxDxSNPxVQAExLN downstream of Motifs 1a and 1b		73 (24 aa)
Motif WxxxDxxY upstream of Motif 2		74 (8 aa)
Motif YxFxRAHD downstream of Motif 2		75 (8 aa)
Motif YxxGGQ upstream of Motifs 3a and 3b		76 (6 aa)
Motif VRxG downstream of Motifs 3a and 3b		77 (4 aa)
Motif 1a: D/N-K-S-I/V-L-D-E-Q-S-D-P-N-H (motif i)		78 (13 aa)
Motif 2: N-K-D-G-S-K/T-A-Y-N-E-D-G-T-V/A-K-Q/K-S-T-I-G-K-Y-N-E-K-Y-G-D-A-S (motif ii)		79 (30 aa)
Motif 3a: L-P-T-D-G-K-M-D-N/K-S-D-V-E-L-Y-R-T-N/S-E (motif iii)		80 (19 aa)
Motif 1b: D-S/P-R-F-T-Y/F-N-A/Q/P-N-D-P		81 (11 aa)
Motif 3b: I-G-N-G-E		82 (5 aa)
Wild type GTF corresponding to 5926 GTF,		83 (1466 aa)

Streptococcus dentirousetti.		
"7527-NT-dIS1a", GTF lacking Motif 1a. DNA codon-optimized for expression in *E. coli*.	84	85 (1325 aa)
"7527-NT-dIS2", GTF lacking Motif 2. DNA codon-optimized for expression in *E. coli*.	86	87 (1311 aa)
"7527-NT-dIS3a", GTF lacking Motif 3a. DNA codon-optimized for expression in *E. coli*.	88	89 (1319 aa)
"7527-NT-dIS1a,2", GTF lacking Motifs 1a and 2. DNA codon-optimized for expression in *E. coli*.	90	91 (1295 aa)
"7527-NT-dIS1a,3a", GTF lacking Motifs 1a and 3a. DNA codon-optimized for expression in *E. coli*.	92	93 (1303 aa)
"7527-NT-dIS2,3a", GTF lacking Motifs 2 and 3a. DNA codon-optimized for expression in *E. coli*.	94	95 (1289 aa)
"7527-NT-dIS1a,2,3a", GTF lacking Motifs 1a, 2 and 3a. DNA codon-optimized for expression in *E. coli*.	96	97 (1273 aa)

DETAILED DESCRIPTION

The disclosures of all patent and non-patent literature cited herein are incorporated herein by reference in their entirety.

Unless otherwise disclosed, the terms "a" and "an" as used herein are intended to encompass one or more (i.e., at least one) of a referenced feature.

Where present, all ranges are inclusive and combinable, except as otherwise noted. For example, when a range of "1 to 5" is recited, the recited range should be construed as including ranges "1 to 4", "1 to 3", "1-2", "1-2 & 4-5", "1-3 & 5", and the like.

The terms "poly alpha-1,3-glucan", "alpha-1,3-glucan polymer", "glucan polymer" and the like are used interchangeably herein. Poly alpha-1,3-glucan is a polymer comprising glucose monomeric units linked together by glycosidic linkages, wherein at least about 50% of the glycosidic linkages are alpha-1,3-glycosidic linkages. Poly alpha-1,3-glucan in certain embodiments comprises at least 95% alpha-1,3-glycosidic linkages.

The terms "glycosidic linkage", "glycosidic bond" and the like are used interchangeably herein and refer to the covalent bond that joins a carbohydrate (sugar) molecule to another group such as another carbohydrate. The term "alpha-1,3-glycosidic linkage" as used herein refers to the type of covalent bond that joins alpha-D-glucose molecules to each other through carbons 1 and 3 on adjacent alpha-D-glucose

rings. The glycosidic linkages of an alpha-1,3-glucan herein can also be referred to as "glucosidic linkages". Herein, "alpha-D-glucose" will be referred to as "glucose".

The term "intrinsic viscosity" as used herein refers to a measure of the contribution of a glucan polymer (e.g., branched alpha-glucan) to the viscosity of a liquid (e.g., solution) comprising the glucan polymer. Intrinsic viscosity can be measured, for example using the methodology disclosed in the Examples below, or as disclosed by Weaver et al. (*J. Appl. Polym. Sci.* 35:1631-1637) and Chun and Park (*Macromol. Chem. Phys.* 195:701-711), for example.

The terms "branching index", "branching ratio" and the like (can be denoted as g') are used interchangeably herein, and refer to the ratio of hydrodynamic volume of a branched polymer chain with a given molar mass, to the hydrodynamic volume of a linear polymer chain with the same molar mass. Branched polymer has a smaller size in solution than its linear counterpart with the same molar mass. Thus, the branching ratio is a useful measure of the overall branching frequency in a polydispersed polymer. Branching index can be measured, for example using the methodology disclosed in the Examples below, or as disclosed by Zdunek et al. (*Food Bioprocess Technol.* 7:3525-3535) and Herget et al. (*BMC Struct. Biol.* 8:35).

The term "sucrose" herein refers to a non-reducing disaccharide composed of an alpha-D-glucose molecule and a beta-D-fructose molecule linked by an alpha-1,2-glycosidic bond. Sucrose is known commonly as table sugar.

The terms "glucosyltransferase enzyme", "GTF enzyme", "GTF", "glucansucrase" and the like are used interchangeably herein. The activity of a GTF enzyme herein catalyzes the reaction of the substrate sucrose to make the products poly alpha-1,3-glucan and fructose. Other products (byproducts) of a GTF reaction can include glucose, various soluble gluco-oligosaccharides (DP2-DP7), and leucrose. Wild type forms of GTF enzymes generally contain (in the N-terminal to C-terminal direction) a signal peptide, a variable domain, a catalytic domain, and a glucan-binding domain. A GTF herein is classified under the glycoside hydrolase family 70 (GH70) according to the CAZy (Carbohydrate-Active EnZymes) database (Cantarel et al., *Nucleic Acids Res.* 37:D233-238, 2009).

The term "glucosyltransferase catalytic domain" herein refers to the domain of a glucosyltransferase enzyme that provides poly alpha-1,3-glucan-synthesizing activity to

a glucosyltransferase enzyme. A glucosyltransferase catalytic domain preferably does not require the presence of any other domains to have this activity.

A "reaction solution" as used herein generally refers to a solution comprising sucrose, water, at least one active glucosyltransferase enzyme, and optionally other components. A reaction solution can alternatively be referred to herein as a "glucan synthesis reaction", "glucan reaction", "GTF reaction", or "reaction composition", for example. Other components that can be in a glucan synthesis reaction include fructose, glucose, leucrose, and soluble gluco-oligosaccharides (e.g., DP2-DP7). It would be understood that certain glucan products, such as poly alpha-1,3-glucan with a degree of polymerization (DP) of at least 8 or 9, are water-insoluble and thus are not dissolved in a glucan synthesis reaction, but rather may be present out of solution. It is in a reaction solution where the step of contacting water, sucrose and a glucosyltransferase enzyme is performed. The term "under suitable reaction conditions" as used herein refers to reaction conditions that support conversion of sucrose to poly alpha-1,3-glucan via glucosyltransferase enzyme activity. A reaction solution as claimed herein is not believed to be naturally occurring.

The "percent dry solids" of a reaction solution refers to the wt% of all the sugars in the glucan synthesis reaction. The percent dry solids of a reaction solution can be calculated, for example, based on the amount of sucrose used to prepare the reaction.

The "yield" of poly alpha-1,3-glucan by a reaction solution herein represents the weight of poly alpha-1,3-glucan product expressed as a percentage of the weight of sucrose substrate that is converted in the reaction. For example, if 100 g of sucrose in a reaction solution is converted to products, and 10 g of the products is poly alpha-1,3-glucan, the yield of the poly alpha-1,3-glucan would be 10%. This yield calculation can be considered as a measure of selectivity of the reaction toward poly alpha-1,3-glucan.

The term "motif" herein refers to a distinctive and recurring structural unit, such as within an amino acid sequence. By "recurring" it is meant that a motif occurs in multiple related polypeptides, for example.

The term "motif (i)" as used herein refers to an amino acid sequence comprising a sequence that is at least 90% identical to SEQ ID NO:78 (Motif 1a, Table 1).

The term "motif (ii)" as used herein refers to an amino acid sequence comprising a sequence that is at least 90% identical to SEQ ID NO:79 (Motif 2, Table 1).

The term "motif (iii)" as used herein refers to an amino acid sequence comprising a sequence that is at least 90% identical to SEQ ID NO:80 (Motif 3a, Table 1).

The terms "percent by volume", "volume percent", "vol %", "v/v %" and the like are used interchangeably herein. The percent by volume of a solute in a solution can be determined using the formula: [(volume of solute)/(volume of solution)] x 100%.

The terms "percent by weight", "weight percentage (wt%)", "weight-weight percentage (% w/w)" and the like are used interchangeably herein. Percent by weight refers to the percentage of a material on a mass basis as it is comprised in a composition, mixture, or solution.

The terms "polynucleotide", "polynucleotide sequence", "nucleic acid sequence", "nucleotide sequence" and the like are used interchangeably herein. A polynucleotide may be a polymer of DNA or RNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof. Nucleotides (ribonucleotides or deoxyribonucleotides) can be referred to by a single letter designation as follows: "A" for adenylate or deoxyadenylate (for RNA or DNA, respectively), "C" for cytidylate or deoxycytidylate (for RNA or DNA, respectively), "G" for guanylate or deoxyguanylate (for RNA or DNA, respectively), "U" for uridylate (for RNA), "T" for deoxythymidylate (for DNA), "R" for purines (A or G), "Y" for pyrimidines (C or T), "K" for G or T, "H" for A or C or T, "I" for inosine, "W" for A or T, and "N" for any nucleotide (e.g., N can be A, C, T, or G, if referring to a DNA sequence; N can be A, C, U, or G, if referring to an RNA sequence).

The term "gene" as used herein refers to a DNA polynucleotide sequence that expresses an RNA (RNA is transcribed from the DNA polynucleotide sequence) from a coding region, which RNA can be a messenger RNA (encoding a protein) or a non-protein-coding RNA. A gene may refer to the coding region alone, or may include regulatory sequences upstream and/or downstream to the coding region (e.g., promoters, 5'-untranslated regions, 3'-transcription terminator regions). A coding region encoding a protein can alternatively be referred to herein as an "open reading frame" (ORF). A gene that is "native" or "endogenous" refers to a gene as found in nature with its own regulatory sequences; such a gene is located in its natural location in the genome of a host cell. A "chimeric" gene refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature (i.e.,

the regulatory and coding regions are heterologous with each other). Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. A "foreign" or "heterologous" gene refers to a gene that is introduced into the host organism by gene transfer. Foreign/heterologous genes can comprise native genes inserted into a non-native organism, native genes introduced into a new location within the native host, or chimeric genes. Polynucleotide sequences in certain embodiments herein are heterologous. A "transgene" is a gene that has been introduced into the genome by a gene delivery procedure (e.g., transformation). A "codon-optimized" open reading frame has its frequency of codon usage designed to mimic the frequency of preferred codon usage of the host cell.

A "non-native" amino acid sequence or polynucleotide sequence herein comprised in a cell or organism herein does not occur in a native (natural) counterpart of such cell or organism.

"Regulatory sequences" as used herein refer to nucleotide sequences located upstream of a gene's transcription start site (e.g., promoter), 5' untranslated regions, introns, and 3' non-coding regions, and which may influence the transcription, processing or stability, and/or translation of an RNA transcribed from the gene. Regulatory sequences herein may include promoters, enhancers, silencers, 5' untranslated leader sequences, introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites, stem-loop structures, and other elements involved in regulation of gene expression. One or more regulatory elements herein may be heterologous to a coding region herein.

A "promoter" as used herein refers to a DNA sequence capable of controlling the transcription of RNA from a gene. In general, a promoter sequence is upstream of the transcription start site of a gene. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. Promoters that cause a gene to be expressed in a cell at most times under all circumstances are commonly referred to as "constitutive promoters". One or more promoters herein may be heterologous to a coding region herein.

A "strong promoter" as used herein refers to a promoter that can direct a relatively large number of productive initiations per unit time, and/or is a promoter driving a higher level of gene transcription than the average transcription level of the genes in a cell.

The terms "3' non-coding sequence", "transcription terminator", "terminator" and the like as used herein refer to DNA sequences located downstream of a coding sequence. This includes polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression.

As used herein, a first nucleic acid sequence is "hybridizable" to a second nucleic acid sequence when a single-stranded form of the first nucleic acid sequence can anneal to the second nucleic acid sequence under suitable annealing conditions (e.g., temperature, solution ionic strength). Hybridization and washing conditions are well known and exemplified in Sambrook J, Fritsch EF and Maniatis T, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989), which is incorporated herein by reference, particularly Chapter 11 and Table 11.1.

The term "DNA manipulation technique" refers to any technique in which the sequence of a DNA polynucleotide sequence is modified. Although the DNA polynucleotide sequence being modified can be used as a substrate itself for modification, it does not have to be physically in hand for certain techniques (e.g., a sequence stored in a computer can be used as the basis for the manipulation technique). A DNA manipulation technique can be used to delete and/or mutate one or more DNA sequences in a longer sequence. Examples of a DNA manipulation technique include recombinant DNA techniques (restriction and ligation, molecular cloning), polymerase chain reaction (PCR), and synthetic DNA methods (e.g., oligonucleotide synthesis and ligation). Regarding synthetic DNA techniques, a DNA manipulation technique can entail observing a DNA polynucleotide *in silico*, determining desired modifications (e.g., one or more deletions) of the DNA polynucleotide, and synthesizing a DNA polynucleotide that contains the desired modifications.

The term "*in silico*" herein means in or on an information storage and/or processing device such as a computer; done or produced using computer software or simulation, i.e., virtual reality.

The terms "cassette", "expression cassette", "gene cassette" and the like are used interchangeably herein. A cassette can refer to a promoter operably linked to a DNA

sequence encoding a protein-coding RNA. A cassette may optionally be operably linked to a 3' non-coding sequence. The structure of a cassette herein can optionally be represented by the simple notation system of "X::Y::Z". Specifically, X describes a promoter, Y describes a coding sequence, and Z describes a terminator (optional); X is operably linked to Y, and Y is operably linked to Z.

The terms "upstream" and "downstream" as used herein with respect to polynucleotides refer to "5' of" and "3' of", respectively.

The term "expression" as used herein refers to (i) transcription of RNA (e.g., mRNA or a non-protein-coding RNA) from a coding region, and/or (ii) translation of a polypeptide from mRNA. Expression of a coding region of a polynucleotide sequence can be up-regulated or down-regulated in certain embodiments.

The term "operably linked" as used herein refers to the association of two or more nucleic acid sequences such that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence. That is, the coding sequence is under the transcriptional control of the promoter. A coding sequence can be operably linked to one (e.g., promoter) or more (e.g., promoter and terminator) regulatory sequences, for example.

The term "recombinant" when used herein to characterize a DNA sequence such as a plasmid, vector, or construct refers to an artificial combination of two otherwise separated segments of sequence, e.g., by chemical synthesis and/or by manipulation of isolated segments of nucleic acids by genetic engineering techniques. Methods for preparing recombinant constructs/vectors herein can follow standard recombinant DNA and molecular cloning techniques as described by J. Sambrook and D. Russell (Molecular Cloning: A Laboratory Manual, 3rd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001); T.J. Silhavy et al. (Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1984); and F.M. Ausubel et al. (Short Protocols in Molecular Biology, 5th Ed. Current Protocols, John Wiley and Sons, Inc., NY, 2002), for example.

The term "transformation" as used herein refers to the transfer of a nucleic acid molecule into a host organism or host cell by any method. A nucleic acid molecule that has been transformed into an organism/cell may be one that replicates autonomously in the organism/cell, or that integrates into the genome of the organism/cell, or that exists

transiently in the cell without replicating or integrating. Non-limiting examples of nucleic acid molecules suitable for transformation are disclosed herein, such as plasmids and linear DNA molecules. Host organisms/cells herein containing a transforming nucleic acid sequence can be referred to as "transgenic", "recombinant", "transformed", "engineered", as a "transformant", and/or as being "modified for exogenous gene expression", for example.

The terms "sequence identity" or "identity" as used herein with respect to polynucleotide or polypeptide sequences refer to the nucleic acid bases or amino acid residues in two sequences that are the same when aligned for maximum correspondence over a specified comparison window. Thus, "percentage of sequence identity" or "percent identity" refers to the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the results by 100 to yield the percentage of sequence identity. It would be understood that, when calculating sequence identity between a DNA sequence and an RNA sequence, T residues of the DNA sequence align with, and can be considered "identical" with, U residues of the RNA sequence. For purposes of determining "percent complementarity" of first and second polynucleotides, one can obtain this by determining (i) the percent identity between the first polynucleotide and the complement sequence of the second polynucleotide (or vice versa), for example, and/or (ii) the percentage of bases between the first and second polynucleotides that would create canonical Watson and Crick base pairs.

The Basic Local Alignment Search Tool (BLAST) algorithm, which is available online at the National Center for Biotechnology Information (NCBI) website, may be used, for example, to measure percent identity between or among two or more of the polynucleotide sequences (BLASTN algorithm) or polypeptide sequences (BLASTP algorithm) disclosed herein. Alternatively, percent identity between sequences may be performed using a Clustal algorithm (e.g., ClustalW, ClustalV, or Clustal-Omega). For

multiple alignments using a Clustal method of alignment, the default values may correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using a Clustal method may be KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids, these parameters may be KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. Alternatively still, percent identity between sequences may be performed using an EMBOSS algorithm (e.g., needle) with parameters such as GAP OPEN=10, GAP EXTEND=0.5, END GAP PENALTY=false, END GAP OPEN=10, END GAP EXTEND=0.5 using a BLOSUM matrix (e.g., BLOSUM62).

Various polypeptide amino acid sequences and polynucleotide sequences are disclosed herein as features of certain embodiments. Variants of these sequences that are at least about 70-85%, 85-90%, or 90%-95% identical to the sequences disclosed herein can be used or referenced. Alternatively, a variant amino acid sequence or polynucleotide sequence can have at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity with a sequence disclosed herein. The variant amino acid sequence or polynucleotide sequence has the same function/activity of the disclosed sequence, or at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% of the function/activity of the disclosed sequence. Any polypeptide amino acid sequence disclosed herein not beginning with a methionine can typically further comprise at least a start-methionine at the N-terminus of the amino acid sequence.

All the amino acid residues at each amino acid position of the proteins disclosed herein are examples. Given that certain amino acids share similar structural and/or charge features with each other (i.e., conserved), the amino acid at each position of a protein herein can be as provided in the disclosed sequences or substituted with a conserved amino acid residue ("conservative amino acid substitution") as follows:

1. The following small aliphatic, nonpolar or slightly polar residues can substitute for each other: Ala (A), Ser (S), Thr (T), Pro (P), Gly (G);
2. The following polar, negatively charged residues and their amides can substitute for each other: Asp (D), Asn (N), Glu (E), Gln (Q);

3. The following polar, positively charged residues can substitute for each other: His (H), Arg (R), Lys (K);
4. The following aliphatic, nonpolar residues can substitute for each other: Ala (A), Leu (L), Ile (I), Val (V), Cys (C), Met (M); and
5. The following large aromatic residues can substitute for each other: Phe (F), Tyr (Y), Trp (W).

The term "isolated" as used herein refers to a polynucleotide or polypeptide molecule that has been completely or partially purified from its native source. In some instances, the isolated polynucleotide or polypeptide molecule is part of a greater composition, buffer system or reagent mix. For example, the isolated polynucleotide or polypeptide molecule can be comprised within a cell or organism in a heterologous manner. "Isolated" herein can also characterize embodiments that are synthetic/man-made, and/or have properties that are not naturally occurring.

The term "increased" as used herein can refer to a quantity or activity that is at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 50%, 100%, or 200% more than the quantity or activity for which the increased quantity or activity is being compared. The terms "increased", "elevated", "enhanced", "greater than", "improved" and the like are used interchangeably herein.

Glucosyltransferase enzymes that can synthesize high molecular weight, linear alpha-1,3-glucan polymer are sought after. Thus, some embodiments disclosed herein concern a reaction solution comprising water, sucrose, and a glucosyltransferase enzyme, wherein the glucosyltransferase enzyme comprises a catalytic domain comprising the following three motifs:

(i) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:78,

(ii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:79, and

(iii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:80;

wherein the glucosyltransferase enzyme does not comprise SEQ ID NO:4, 20, 28, 30, 65, residues 54-957 of SEQ ID NO:65, residues 55-960 of SEQ ID NO:30, residues 55-

960 of SEQ ID NO:4, residues 55-960 of SEQ ID NO:28, or residues 55-960 of SEQ ID NO:20. Significantly, the glucosyltransferase enzyme(s) in such reaction solutions produces insoluble poly alpha-1,3-glucan having at least 95% alpha-1,3 glycosidic linkages and a weight average degree of polymerization (DP_w) of at least 100. Such glucan, which is mostly or completely linear, is suitable for use in spinning fibers and in other industrial applications.

The molecular weight of poly alpha-1,3-glucan produced by glucosyltransferase enzymes herein can be measured as DP_w (weight average degree of polymerization) or DP_n (number average degree of polymerization). Alternatively, the molecular weight of poly alpha-1,3-glucan herein can be measured as number-average molecular weight (M_n) or as weight-average molecular weight (M_w). Alternatively still, molecular weight can be measured in terms of Daltons or grams/mole.

Poly alpha-1,3-glucan in certain embodiments can have a molecular weight in DP_w or DP_n of at least about 100. For example, the molecular weight can be at least about 400 DP_w or DP_n. DP_w or DP_n in still another embodiment can be at least about 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, or 1000 (or any integer between 100 and 1000).

The molecular weight of poly alpha-1,3-glucan can be measured using any of several means known in the art. For example, glucan polymer molecular weight can be measured using high-pressure liquid chromatography (HPLC), size exclusion chromatography (SEC), or gel permeation chromatography (GPC).

Poly alpha-1,3-glucan in certain embodiments has at least about 95%, 96%, 97%, 98%, 99%, or 100% alpha-1,3 glycosidic linkages. In some embodiments, accordingly, poly alpha-1,3-glucan has less than about 5%, 4%, 3%, 2%, 1%, or 0% of glycosidic linkages that are not alpha-1,3. It should be understood that the higher the percentage of alpha-1,3-glycosidic linkages present in poly alpha-1,3-glucan, the greater the probability that the poly alpha-1,3-glucan is linear, since there are lower occurrences of certain glycosidic linkages forming branch points in the polymer. Thus, poly alpha-1,3-glucan with 100% alpha-1,3 glycosidic linkages is completely linear. In certain embodiments, poly alpha-1,3-glucan has no branch points or less than about 5%, 4%, 3%, 2%, or 1% branch points as a percent of the glycosidic linkages in the polymer. Examples of branch points include alpha-1,6 branch points.

The glycosidic linkage profile of poly alpha-1,3-glucan herein can be determined using any method known in the art. For example, the linkage profile can be determined using methods that use nuclear magnetic resonance (NMR) spectroscopy (e.g., ^{13}C NMR or ^{1}H NMR). These and other methods that can be used are disclosed in Food Carbohydrates: Chemistry, Physical Properties, and Applications (S. W. Cui, Ed., Chapter 3, S. W. Cui, Structural Analysis of Polysaccharides, Taylor & Francis Group LLC, Boca Raton, FL, 2005), which is incorporated herein by reference.

Poly alpha-1,3-glucan produced by a glucosyltransferase herein is typically insoluble in most aqueous systems. In general, the solubility of a glucan polymer in an aqueous systems is related to its linkage type, molecular weight and/or degree of branching. Poly alpha-1,3-glucan is generally insoluble at a DP_w of 8 and above in aqueous (or mostly aqueous) liquids at 20 °C. A glucosyltransferase enzyme herein can produce poly alpha-1,3-glucan as presently disclosed.

A glucosyltransferase enzyme in certain embodiments further comprises a glucosyltransferase catalytic domain comprising an amino acid sequence that is at least 90% identical to amino acid positions 54-957 of SEQ ID NO:65, and have glucosyltransferase activity. Alternatively, a glucosyltransferase catalytic domain can comprise an amino acid sequence that is, for example, at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 98.5%, 99%, or 99.5% (but not 100%) identical to amino acid positions 54-957 of SEQ ID NO:65, and have glucosyltransferase activity.

SEQ ID NOs:65 (GTF 7527), 30 (GTF 2678), 4 (GTF 6855), 28 (GTF 2919), and 20 (GTF 2765) each represent a glucosyltransferase that, compared to its respective wild type counterpart, lacks the signal peptide domain and all or a substantial portion of the variable domain. Thus, each of these glucosyltransferase enzymes has a catalytic domain followed by a glucan-binding domain. The approximate location of catalytic domain sequences in these enzymes is as follows: 7527 (residues 54-957 of SEQ ID NO:65), 2678 (residues 55-960 of SEQ ID NO:30), 6855 (residues 55-960 of SEQ ID NO:4), 2919 (residues 55-960 of SEQ ID NO:28), 2765 (residues 55-960 of SEQ ID NO:20). The amino acid sequences of catalytic domains of GTFs 2678, 6855, 2919 and 2765 have about 94.9%, 99.0%, 95.5% and 96.4% identity, respectively, with a catalytic domain sequence of 7527 (i.e., amino acids 54-957 of SEQ ID NO:65) (Table 4). These particular glucosyltransferase enzymes can produce poly alpha-1,3-glucan with 100%

alpha-1,3 linkages and a DP_w of at least 400 (Table 4). Thus, a glucosyltransferase enzyme in certain embodiments can comprise, or consist of, a glucosyltransferase catalytic domain that is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 98.5%, 99%, or 99.5% (but not 100%) identical to the amino acid sequence of a catalytic domain of GTF 2678, 6855, 2919, or 2765. In some embodiments, a glucosyltransferase catalytic domain sequence does not comprise residues 54-957 of SEQ ID NO:65, residues 55-960 of SEQ ID NO:30, residues 55-960 of SEQ ID NO:4, residues 55-960 of SEQ ID NO:28, or residues 55-960 of SEQ ID NO:20.

Amino acid positions 54-957 of SEQ ID NO:65 represent, approximately, a catalytic domain sequence of the glucosyltransferase identified in GENBANK under GI number 47527 (SEQ ID NO:60). SEQ ID NO:65 generally represents the catalytic domain and glucan-binding domain of SEQ ID NO:60; the signal peptide and variable domains are missing from SEQ ID NO:65. As shown in Example 14, a catalytic domain sequence of SEQ ID NO:65 (residues 54-957) was able to catalyze the production of poly alpha-1,3-glucan. Example 14 also shows that a catalytic domain sequence of SEQ ID NO:14 (residues 57-906 of SEQ ID NO:14 [GTF 5926]) was able to catalyze production of poly alpha-1,3-glucan. The molecular weight of poly alpha-1,3-glucan produced by each of these catalytic domain sequences generally corresponded with the molecular weight of the product produced by their enzyme counterparts containing both the catalytic domain and glucan binding domain (refer to activity of SEQ ID NOs:65 and 14 in Table 4, DP_w150). Thus, it is believed that a catalytic domain sequence herein is an important structural component for a glucosyltransferase enzyme to be capable of producing poly alpha-1,3-glucan.

Although it is believed that a glucosyltransferase enzyme herein need only have a catalytic domain sequence, such as one comprising an amino acid sequence that is at least 90% identical to amino acid positions 54-957 of SEQ ID NO:65 (or positions 55-960 of SEQ ID NO:30, positions 55-960 of SEQ ID NO:4, positions 55-960 of SEQ ID NO:28, or positions 55-960 of SEQ ID NO:20), the glucosyltransferase enzyme can be comprised within a larger amino acid sequence. For example, the catalytic domain may be linked at its C-terminus to a glucan-binding domain, and/or linked at its N-terminus to a variable domain and/or signal peptide.

Still further examples of glucosyltransferase enzymes can be any as disclosed herein and that include 1-300 (or any integer there between [e.g., 10, 15, 20, 25, 30, 35,

40, 45, or 50]) residues on the N-terminus and/or C-terminus. Such additional residues may be from a corresponding wild type sequence from which the glucosyltransferase enzyme is derived, or may be a heterologous sequence such as an epitope tag (at either N- or C-terminus) or a heterologous signal peptide (at N-terminus), for example.

A glucosyltransferase enzyme herein typically lacks an N-terminal signal peptide. An expression system for producing a glucosyltransferase enzyme herein may employ an enzyme-encoding polynucleotide that further comprises sequence encoding an N-terminal signal peptide to direct extra-cellular secretion, if desired. The signal peptide in such embodiments is cleaved from the enzyme during the secretion process. The signal peptide may either be native or heterologous to the glucosyltransferase. An example of a signal peptide useful herein is one from a bacterial (e.g., a *Bacillus* species such as *B. subtilis*) or fungal species. An example of a bacterial signal peptide is an aprE signal peptide, such as one from *Bacillus* (e.g., *B. subtilis*, see Vogtentanz et al., *Protein Expr. Purif.* 55:40-52, which is incorporated herein by reference).

FIG. 2 shows that a catalytic domain sequence of GTF 7527 (residues 54-957 of SEQ ID NO:65) aligns with catalytic domain sequences of several other glucosyltransferase enzymes, with several regions showing complete conservation across all the sequences (residues with dark background). The dark background residues in FIG. 2 visually map out the catalytic domain of each sequence, indicating their length to be about 850 to 900 amino acid residues long. Thus, the catalytic domain of the glucosyltransferase enzyme can be about 800-950 (or any integer between 800 and 950) amino acid residues long, for example.

Certain of the conserved regions in FIG. 2 include catalytic active site motifs SEQ ID NOs:68, 69, 70, and 71 (refer to Example 3). Thus, a catalytic domain sequence of a glucosyltransferase enzyme in some aspects can contain one or more of SEQ ID NOs:68, 69, 70, and 71 in alignment, respectively, with SEQ ID NOs:68, 69, 70, and 71 as present in amino acids 54-957 of SEQ ID NO:65. Other conserved regions in FIG. 2 include SEQ ID NOs:72, 73, 74, 75, 76 and 77 (refer to Example 4). Thus, a catalytic domain sequence of a glucosyltransferase enzyme in some aspects can contain one or more of SEQ ID NOs:72, 73, 74, 75, 76 and 77 in alignment, respectively, with SEQ ID NOs:72, 73, 74, 75, 76 and 77 as present in amino acids 54-957 of SEQ ID NO:65.

The catalytic domain of a glucosyltransferase enzyme herein can have activity as exhibited by a catalytic domain of a glucosyltransferase classified under the glycoside

hydrolase family 70 (GH70). Such a GH70 glucosyltransferase may be found in the CAZy (Carbohydrate-Active EnZymes) database (Cantarel et al., *Nucleic Acids Res.* 37:D233-238, 2009), for example.

A glucosyltransferase enzyme herein can comprise a glucosyltransferase catalytic domain comprising the following three motifs:

(i) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:78,

(ii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:79, and

(iii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:80.

Motif (i) corresponds with "Motif 1a" (FIG. 3). Motif (ii) corresponds with "Motif 2" (FIG. 5). Motif (iii) corresponds with "Motif 3a" (FIG. 7).

Motif (i) can be at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO:78. Motif (ii) can be at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO:79. Motif (iii) can be at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO:80. Thus, it can be seen that in certain embodiments, motif (i) can comprise SEQ ID NO:78, motif (ii) can comprise SEQ ID NO:79, and motif (iii) can comprise SEQ ID NO:80.

Regarding motif (i) in certain embodiments, the first residue of SEQ ID NO:78 (D/N-K-S-I/V-L-D-E-Q-S-D-P-N-H) can be an aspartate (D) and the fourth residue can be an isoleucine (I). Alternatively, the first residue can be an aspartate (D) and the fourth residue can be a valine (V), or the first residue can be an asparagine (N) and the fourth residue can be an isoleucine (I), or the first residue can be an asparagine (N) and the fourth residue can be a valine (V).

Regarding motif (ii) in certain embodiments, the sixth residue of SEQ ID NO:79 (N-K-D-G-S-K/T-A-Y-N-E-D-G-T-V/A-K-Q/K-S-T-I-G-K-Y-N-E-K-Y-G-D-A-S) can be a lysine (K), the fourteenth residue can be a valine (V), and the sixteenth residue can be a glutamine (Q). Alternatively, the sixth residue can be a lysine (K), the fourteenth residue can be an alanine (A), and the sixteenth residue can be a glutamine (Q); or the sixth residue can be a lysine (K), the fourteenth residue can be an valine (V), and the

sixteenth residue can be a lysine (K). Additional examples include where the sixth residue can be a threonine (T).

Regarding motif (iii) in certain embodiments, the ninth residue of SEQ ID NO:80 (L-P-T-D-G-K-M-D-N/K-S-D-V-E-L-Y-R-T-N/S-E) can be an asparagine (N) and the eighteenth residue can be an asparagine (N). Alternatively, the ninth residue can be an asparagine (N) and the eighteenth residue can be a serine (S), or the ninth residue can be a lysine (K) and the eighteenth residue can be an asparagine (N), or the ninth residue can be a lysine (K) and the eighteenth residue can be a serine (S).

The relative positions of motif (i) (SEQ ID NO:78), motif (ii) (SEQ ID NO:79) and motif (iii) (SEQ ID NO:80) align with residues 231-243, 396-425 and 549-567, respectively, of the GTF 7527 sequence (SEQ ID NO:65) shown in FIG. 2. In certain embodiments herein,

(A) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:78 in the glucosyltransferase catalytic domain aligns with amino acid positions 231-243 of SEQ ID NO:65;

(B) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:79 in the glucosyltransferase catalytic domain aligns with amino acid positions 396-425 of SEQ ID NO:65; and/or

(C) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:80 in the glucosyltransferase catalytic domain aligns with amino acid positions 549-567 of SEQ ID NO:65.

The term “aligns with” can be used interchangeably with “corresponds to”, “corresponds with”, and the like. The relative positions of motifs (i), (ii) and/or (iii) in a glucosyltransferase catalytic domain can thus be determined with reference to the above amino acid positions in SEQ ID NO:65. For example, the sequence of a glucosyltransferase catalytic domain can be aligned with SEQ ID NO:65 using any means known in the art, such as through use of an alignment algorithm or software as described above (e.g., BLASTP, ClustalW, ClustalV, EMBOSS).

The relative positions of motifs (i), (ii) and (iii) in a glucosyltransferase catalytic domain can be determined with reference to certain conserved sequences, namely SEQ ID NOs:72, 73, 74, 75, 76 and 77, if desired.

Motif 1a (SEQ ID NO:78) is flanked by upstream and downstream conserved sequences as shown in FIG. 3. Preceding Motif 1a is the sequence SxxRxxN (SEQ ID

NO:72), and following this motif is the sequence GGxxxLLxNDxDxSNPxVQAExLN (SEQ ID NO:73). Thus, the position of motif (i) can be located between SEQ ID NOs:72 and 73. SEQ ID NO:72 can be directly adjacent (upstream) to motif (i), or 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 (or 1-15) amino acid residues upstream motif (i). SEQ ID NO:73 can be directly adjacent (downstream) to motif (i), or 1, 2, 3, 4, or 5 (or 1-5) amino acid residues downstream motif (i).

Motif 2 (SEQ ID NO:79) is flanked by upstream and downstream conserved sequences as shown in FIG. 5. Specifically, preceding Motif 2 is the sequence WxxxDxxY (SEQ ID NO:74), and following this motif is the sequence YxFxRAHD (SEQ ID NO:75). Thus, the position of motif (ii) can be located between SEQ ID NOs:74 and 75. SEQ ID NO:74 can be directly adjacent (upstream) to motif (ii), or 1-65 (or any integer between 1 and 65) amino acid residues upstream motif (ii). SEQ ID NO:75 can be directly adjacent (downstream) to motif (ii), or 1, 2, 3, 4, or 5 (or 1-5) amino acid residues downstream motif (ii).

Motif 3a (SEQ ID NO:80) is flanked by upstream and downstream conserved sequences as shown in FIG. 7. Specifically, preceding Motif 3a is the sequence YxxGGQ (SEQ ID NO:76), and following this motif is the sequence VRxG (SEQ ID NO:77). Thus, the position of motif (iii) can be located between SEQ ID NOs:76 and 77. SEQ ID NO:76 can be directly adjacent (upstream) to motif (iii), or 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or 11 (or 1-11) amino acid residues upstream motif (iii). SEQ ID NO:77 can be directly adjacent (downstream) to motif (iii), or 1, 2, 3, 4, 5, 6, 7, 8, or 9 (or 1-9) amino acid residues downstream motif (iii).

Certain amino acid positions in the upstream/downstream conserved sequences SEQ ID NOs:72-77 can be any amino acid (indicated by an "x" in each sequence in Table 1). Examples of SEQ ID NOs:72 and 73 are as shown in any of the GTF sequences in FIGs. 2 and 3 at the amino acids of each GTF sequence aligning with positions 214-220 and 245-268, respectively, of SEQ ID NO:65 (GTF 7527). Examples of SEQ ID NOs:74 and 75 are as shown in any of the GTF sequences in FIGs. 2 and 5 at the amino acids of each GTF sequence aligning with positions 334-341 and 428-435, respectively, of SEQ ID NO:65 (GTF 7527). Examples of SEQ ID NOs:76 and 77 are as shown in any of the GTF sequences in FIGs. 2 and 7 at the amino acids of each GTF sequence aligning with positions 537-542 and 572-575, respectively, of SEQ ID NO:65 (GTF 7527).

A glucosyltransferase enzyme herein can be derived from any microbial source, such as a bacteria or fungus. Examples of bacterial glucosyltransferase enzymes are those derived from a *Streptococcus* species, *Leuconostoc* species or *Lactobacillus* species. Examples of *Streptococcus* species include *S. salivarius*, *S. sobrinus*, *S. dentirousetti*, *S. downei*, *S. mutans*, *S. oralis*, *S. gallolyticus* and *S. sanguinis*. Examples of *Leuconostoc* species include *L. mesenteroides*, *L. amelibiosum*, *L. argentinum*, *L. carnosum*, *L. citreum*, *L. cremoris*, *L. dextranicum* and *L. fructosum*. Examples of *Lactobacillus* species include *L. acidophilus*, *L. delbrueckii*, *L. helveticus*, *L. salivarius*, *L. casei*, *L. curvatus*, *L. plantarum*, *L. sakei*, *L. brevis*, *L. buchneri*, *L. fermentum* and *L. reuteri*.

A glucosyltransferase enzyme in some aspects does not comprise SEQ ID NO:4, SEQ ID NO:20, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:65. In certain embodiments, a glucosyltransferase enzyme herein does not comprise positions 2-1341 of SEQ ID NO:4, positions 2-1340 of SEQ ID NO:20, positions 2-1340 of SEQ ID NO:28, positions 2-1341 of SEQ ID NO:30, or positions 2-1341 of SEQ ID NO:65.

A glucosyltransferase enzyme herein can produce poly alpha-1,3-glucan as presently disclosed, such as is the above disclosure.

One or more different glucosyltransferase enzymes may be used in certain aspects. The glucosyltransferase enzyme in certain embodiments does not have, or has very little (less than 1%), dextransucrase, reuteransucrase, alternansucrase activity, or mutansucrase activity. A reaction solution herein may contain one, two, or more glucosyltransferase enzymes, for example.

A glucosyltransferase enzyme for a glucan synthesis reaction herein may be produced by any means known in the art. For example, a glucosyltransferase enzyme may be produced recombinantly in a heterologous expression system, such as a microbial heterologous expression system. Examples of heterologous expression systems include bacterial (e.g., *E. coli* such as TOP10 or MG1655; *Bacillus* sp.) and eukaryotic (e.g., yeasts such as *Pichia* sp. and *Saccharomyces* sp.) expression systems.

In certain embodiments, a heterologous gene expression system may be one that is designed for protein secretion. A glucosyltransferase enzyme typically comprises a signal peptide (signal sequence) in such embodiments. The signal peptide may be either its native signal peptide or a heterologous signal peptide.

A glucosyltransferase enzyme described herein may be used in any purification state (e.g., pure or non-pure). For example, a glucosyltransferase enzyme may be purified and/or isolated prior to its use. Examples of glucosyltransferase enzymes that are non-pure include those in the form of a cell lysate. A cell lysate or extract may be prepared from a bacteria (e.g., *E. coli*) used to heterologously express the enzyme. For example, the bacteria may be subjected to disruption using a French pressure cell. In alternative embodiments, bacteria may be homogenized with a homogenizer (e.g., APV, Rannie, Gaulin). A glucosyltransferase enzyme is typically soluble in these types of preparations. A bacterial cell lysate, extract, or homogenate herein may be used at about 0.15-0.3% (v/v), for example, in a reaction solution for producing poly alpha-1,3-glucan from sucrose.

The activity of a glucosyltransferase enzyme herein can be determined using any method known in the art. For example, glucosyltransferase enzyme activity can be determined by measuring the production of reducing sugars (fructose and glucose) in a reaction solution containing sucrose (50 g/L), dextran T10 (1 mg/mL) and potassium phosphate buffer (pH 6.5, 50 mM), where the solution is held at 22-25 °C for 24-30 hours. The reducing sugars can be measured, for instance, by adding 0.01 mL of the reaction solution to a mixture containing 1 N NaOH and 0.1% triphenyltetrazolium chloride and then monitoring the increase in absorbance at OD_{480nm} for five minutes.

A reaction solution herein refers to a solution comprising at least sucrose, water and an active glucosyltransferase enzyme, and optionally other components. Other components that can be in a glucan synthesis reaction include fructose, glucose, leucrose, soluble oligosaccharides (e.g., DP2-DP7), for example. It would be understood that certain glucan products, such as poly alpha-1,3-glucan with a DP of at least 8 or 9, may be water-insoluble and thus are not dissolved in a glucan synthesis reaction, but rather may be present out of solution. A reaction solution herein may be one that, in addition to producing insoluble glucan product, produces byproducts such as leucrose and/or soluble oligosaccharides.

The temperature of a reaction solution herein can be controlled, if desired. In certain embodiments, the temperature of the reaction is between about 5 °C to about 50 °C. The temperature in certain other embodiments is between about 20 °C to about 40 °C, or about 20 °C to about 30 °C (e.g., about 25 °C).

The initial concentration of sucrose in a reaction solution herein can be about 20 g/L to about 400 g/L, for example. Alternatively, the initial concentration of sucrose can be about 75 g/L to about 175 g/L, or from about 50 g/L to about 150 g/L. Alternatively still, the initial concentration of sucrose can be about 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, or 160 g/L (or any integer value between 40 and 160 g/L), for example. "Initial concentration of sucrose" refers to the sucrose concentration in a GTF reaction solution just after all the reaction solution components have been added (e.g., at least water, sucrose, GTF enzyme).

Sucrose used in a glucan synthesis reaction herein can be highly pure (≥ 99.5%) or be of any other purity or grade. For example, sucrose can have a purity of at least 99.0%, or can be reagent grade sucrose. As another example, incompletely refined sucrose can be used. Incompletely refined sucrose herein refers to sucrose that has not been processed to white refined sucrose. Thus, incompletely refined sucrose can be completely unrefined or partially refined. Examples of unrefined sucrose are "raw sucrose" ("raw sugar") and solutions thereof. Examples of partially refined sucrose have not gone through one, two, three, or more crystallization steps. The ICUMSA (International Commission for Uniform Methods of Sugar Analysis) of incompletely refined sucrose herein can be greater than 150, for example. Sucrose herein may be derived from any renewable sugar source such as sugar cane, sugar beets, cassava, sweet sorghum, or corn. Suitable forms of sucrose useful herein are crystalline form or non-crystalline form (e.g., syrup, cane juice, beet juice), for example.

Methods of determining ICUMSA values for sucrose are well known in the art and disclosed by the International Commission for Uniform Methods of Sugar Analysis in ICUMSA Methods of Sugar Analysis: Official and Tentative Methods Recommended by the International Commission for Uniform Methods of Sugar Analysis (ICUMSA) (Ed. H.C.S. de Whalley, Elsevier Pub. Co., 1964), for example, which is incorporated herein by reference. ICUMSA can be measured, for example, by ICUMSA Method GS1/3-7 as described by R.J. McCowage, R.M. Urquhart and M.L. Burge (Determination of the Solution Colour of Raw Sugars, Brown Sugars and Coloured Syrups at pH 7.0 – Official, Verlag Dr Albert Bartens, 2011 revision), which is incorporated herein by reference.

The pH of a glucan synthesis reaction in certain embodiments can be between about 4.0 to about 8.0. Alternatively, the pH can be about 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, or 8.0. The pH can be adjusted or controlled by the addition or incorporation of a

suitable buffer, including but not limited to: phosphate, tris, citrate, or a combination thereof. Buffer concentration in a glucan synthesis reaction can be from 0 mM to about 100 mM, or about 10, 20, or 50 mM, for example.

Examples of other conditions and components suitable for carrying out a reaction solution herein are disclosed in U.S. Patent No. 7000000, and U.S. Pat. Appl. Publ. Nos. 2013/0244288, 2013/0244287, 2013/0196384, 2013/0157316, and 2014/0087431, all of which are incorporated herein by reference.

The present disclosure also concerns a method for producing insoluble poly alpha-1,3-glucan comprising:

(a) contacting at least water, sucrose, and a glucosyltransferase enzyme, wherein the glucosyltransferase enzyme comprises a catalytic domain comprising the following three motifs:

(i) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:78,

(ii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:79, and

(iii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:80,

and wherein the glucosyltransferase enzyme does not comprise SEQ ID NO:4, 20, 28, 30, 65, residues 54-957 of SEQ ID NO:65, residues 55-960 of SEQ ID NO:30, residues 55-960 of SEQ ID NO:4, residues 55-960 of SEQ ID NO:28, or residues 55-960 of SEQ ID NO:20;

whereby insoluble poly alpha-1,3-glucan is produced having at least 95% alpha-1,3 glycosidic linkages and a weight average degree of polymerization (DP_w) of at least 100; and

b) optionally, isolating the poly alpha-1,3-glucan produced in step (a). Significantly, the poly alpha-1,3-glucan produced in such a method is mostly or completely linear. This method can thus optionally be characterized as a method of producing linear (or mostly linear) poly alpha-1,3-glucan.

A glucan synthesis method as presently disclosed comprises contacting at least water, sucrose, and a glucosyltransferase enzyme as described herein. These and optionally other reagents can be added altogether or added in any order as discussed

below. This step can comprise providing a reaction solution comprising water, sucrose and a glucosyltransferase enzyme. It will be understood that, as the glucosyltransferase enzyme synthesizes poly alpha-1,3-glucan, the reaction solution becomes a reaction mixture given that insoluble poly alpha-1,3-glucan falls out of solution as indicated by clouding of the reaction. The contacting step of the disclosed method can be performed in any number of ways. For example, the desired amount of sucrose can first be dissolved in water (optionally, other components may also be added at this stage of preparation, such as buffer components), followed by the addition of glucosyltransferase enzyme. The solution may be kept still, or agitated via stirring or orbital shaking, for example. Typically, a glucan synthesis reaction is cell-free.

Completion of a reaction in certain embodiments can be determined visually (no more accumulation of insoluble poly alpha-1,3-glucan) and/or by measuring the amount of sucrose left in the solution (residual sucrose), where a percent sucrose consumption of over about 90% can indicate reaction completion, for example. Typically, a reaction of the disclosed process will take about 12, 24, 36, 48, 60, 72, 84, or 96 hours to complete, depending on certain parameters such as the amount of sucrose and glucosyltransferase enzyme used in the reaction.

The percent sucrose consumption of a reaction in certain embodiments is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% of the sucrose initially contacted with water and a glucosyltransferase enzyme. Alternatively, the percent sucrose consumption may be >90% or >95%.

The yield of poly alpha-1,3-glucan produced in some aspects of a glucan synthesis method herein can be at least about 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, or 20%, based on the weight of sucrose converted in the reaction.

Poly alpha-1,3-glucan produced in the disclosed method may optionally be isolated. For example, insoluble poly alpha-1,3-glucan may be separated by centrifugation or filtration. In doing so, poly alpha-1,3-glucan is separated from most of the reaction solution, which may comprise water, fructose and certain byproducts (e.g., leucrose, soluble oligosaccharides DP2-DP7). This solution may also comprise residual sucrose and glucose monomer. Isolation can optionally further comprise washing the poly alpha-1,3-glucan one, two, or more times with water or other aqueous liquid, and/or drying the poly alpha-1,3-glucan.

A glucosyltransferase enzyme in certain embodiments of a glucan synthesis method herein can further comprise a glucosyltransferase catalytic domain comprising an amino acid sequence that is at least 90% identical to amino acid positions 54-957 of SEQ ID NO:65, and have glucosyltransferase activity. Alternatively, a glucosyltransferase catalytic domain can comprise an amino acid sequence that is, for example, at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 98.5%, 99%, or 99.5% (but not 100%) identical to amino acid positions 54-957 of SEQ ID NO:65, and have glucosyltransferase activity.

The above embodiments of poly alpha-1,3-glucan synthesis methods are examples. Any other feature disclosed herein can apply to a glucan synthesis method, accordingly. For example, any of the poly alpha-1,3-glucan product, glucosyltransferase enzyme (e.g., the catalytic domain and its motifs i, ii and iii), and reaction solution condition features disclosed herein can be applied as appropriate.

The present disclosure also concerns a method of identifying a glucosyltransferase enzyme. This method comprises detecting the presence at least one motif in a glucosyltransferase catalytic domain, the at least one motif selected from the group consisting of:

(i) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:78,

(ii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:79, and

(iii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:80;

thereby identifying a glucosyltransferase enzyme that produces insoluble poly alpha-1,3-glucan having at least 95% alpha-1,3 glycosidic linkages and a weight average degree of polymerization (DP_w) of at least 100. Since the poly alpha-1,3-glucan produced in this a method is mostly or completely linear, this method can optionally be characterized as a method of identifying a glucosyltransferase enzyme that produces linear poly alpha-1,3-glucan.

It is contemplated that, although the above method comprises detecting any one of motifs (i), (ii), and (iii) in a glucosyltransferase catalytic domain, the method results in detecting a glucosyltransferase catalytic domain having all three of these motifs. This

being said, a GTF identification method herein can optionally comprise detecting one of, two of, or all three, of motifs (i), (ii) and/or (iii) in a glucosyltransferase catalytic domain.

The detection step in a GTF identification method herein can comprise detecting an isolated amino acid sequence of a glucosyltransferase enzyme having motifs (i), (ii), or (iii). The detecting step can also be performed by detecting an isolated polynucleotide sequence encoding a glucosyltransferase enzyme having motifs (i), (ii), or (iii). The codons used to prepare the isolated polynucleotide sequence in such embodiments optionally are preferred codons for a species (e.g., *E. coli* or *S. cerevisiae*) that may be used to heterologously express the glucosyltransferase enzyme.

The presence of at least one of motifs (i), (ii), or (iii) in the catalytic domain of a glucosyltransferase enzyme can be detected following any means known in the art and/or any procedure described herein. For example, detection can be performed (a) *in silico*, (b) with a method comprising a nucleic acid hybridization step, (c) with a method comprising a protein sequencing step, and/or (d) with a method comprising a protein binding step.

Motifs (i), (ii) and (iii) were identified by *in silico* detection (see Example 4 below). Thus, the amino acid sequences of glucosyltransferase enzymes (and/or nucleotide sequences encoding such glucosyltransferase enzymes) stored in a computer or database (e.g., public databases such as GENBANK, EMBL, REFSEQ, GENEPEPT, SWISS-PROT, PIR, PDB) can be reviewed *in silico* to identify a glucosyltransferase enzyme comprising at least one of motifs (i), (ii) or (iii) in its catalytic domain, for example. Such review could comprise using any means known in the art such as through use of an alignment algorithm or software as described above (e.g., BLASTN, BLASTP, ClustalW, ClustalV, EMBOSS). The sequence of the glucosyltransferase catalytic domain being reviewed could be aligned with a catalytic domain sequence of SEQ ID NO:65 (GTF 7527), which comprises Motifs 1a (SEQ ID NO:78), 2 (SEQ ID NO:79) and 3a (SEQ ID NO:80), to detect the presence or absence of motifs (i), (ii), and/or (iii). Alternatively, the sequence of the glucosyltransferase catalytic domain being reviewed could be aligned with a catalytic domain sequence of SEQ ID NO:30 (GTF 2678), SEQ ID NO:4 (GTF 6855), SEQ ID NO:28 (GTF 2919), and/or SEQ ID NO:20 (GTF 2765), all of which comprise Motifs 1a (SEQ ID NO:78), 2 (SEQ ID NO:79) and 3a (SEQ ID NO:80), to identify the presence or absence of motifs (i), (ii), and/or (iii).

Another *in silico* means for detecting motifs (i), (ii), and/or (iii) in a glucosyltransferase catalytic domain sequence can comprise comparing the predicted three-dimensional structure (tertiary structure) of a glucosyltransferase catalytic domain sequence with a reference structure. The structures of both the catalytic domain being reviewed and the reference can be visually compared using any means known in the art such as with a computer program that provides a structure based on amino acid sequence input (e.g., software package MOE, Chemical Computing Group, Montreal, Canada). For example, if the reference structure lacks motif (i), (ii), and/or (iii), the comparison may detect the presence of motif (i), (ii), and/or (iii) by showing a domain(s) in the structure being reviewed that does not have a corresponding domain in the reference structure. Examples of this type of comparison are shown in FIGs. 4a, 4b, 6a, 6b, 8a and 8b.

Alternatively, detecting a glucosyltransferase enzyme having motifs (i), (ii), and (iii) in its catalytic domain can be through using a method comprising a nucleic acid hybridization step. Such a method can comprise using DNA hybridization (e.g., Southern blot, dot blot), RNA hybridization (e.g., northern blot), or any other method that has a nucleic acid hybridization step (e.g., DNA sequencing, PCR, RT-PCR, all of which may comprise hybridization of an oligonucleotide), for example. As an example, an oligonucleotide that would hybridize to a nucleotide sequence encoding Motif 1a (SEQ ID NO:78), 2 (SEQ ID NO:79), or 3a (SEQ ID NO:80) could be used to detect its presence or absence in a polynucleotide sequence encoding the glucosyltransferase catalytic domain being reviewed. The conditions and parameters for carrying out hybridization methods in general are well known and disclosed, for example, in Sambrook J, Fritsch EF and Maniatis T, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989); Silhavy TJ, Bennan ML and Enquist LW, Experiments with Gene Fusions, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1984); Ausubel FM et al., Current Protocols in Molecular Biology, published by Greene Publishing Assoc. and Wiley-Interscience, Hoboken, NJ (1987); and Innis MA, Gelfand DH, Sninsky JJ and White TJ (Editors), PCR Protocols: A Guide to Methods and Applications, Academic Press, Inc., San Diego, CA (1990).

In another aspect, a glucosyltransferase enzyme that comprises motifs (i), (ii), and (iii) in its catalytic domain can be detected using a method comprising a protein sequencing step. Such a protein sequencing step can comprise one or more

procedures such as N-terminal amino acid analysis, C-terminal amino acid analysis, Edman degradation, or mass spectrometry, for example.

In still another aspect, a glucosyltransferase enzyme that comprises motifs (i), (ii), and (iii) in its catalytic domain can be detected using a method comprising a protein binding step. Such a protein binding step could be performed using an antibody that specifically binds to one of these motifs, for example. Antibodies for identifying the presence or absence of motif (i) can be specific for an amino acid sequence that is at least 90% identical to SEQ ID NO:78. Antibodies for identifying the presence or absence of motif (ii) can be specific for an amino acid sequence that is at least 90% identical to SEQ ID NO:79. Antibodies for identifying the presence or absence of motif (iii) can be specific for an amino acid sequence that is at least 90% identical to SEQ ID NO:80.

Motif (i) can be at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO:78. Motif (ii) can be at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO:79. Motif (iii) can be at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO:80. Thus, it can be seen that in certain embodiments of a detection method herein, motif (i) can comprise SEQ ID NO:78, motif (ii) can comprise SEQ ID NO:79, and motif (iii) can comprise SEQ ID NO:80.

Regarding motif (i) in certain embodiments, the first residue of SEQ ID NO:78 (D/N-K-S-I/V-L-D-E-Q-S-D-P-N-H) can be an aspartate (D) and the fourth residue can be an isoleucine (I). Alternatively, the first residue can be an aspartate (D) and the fourth residue can be a valine (V), or the first residue can be an asparagine (N) and the fourth residue can be an isoleucine (I), or the first residue can be an asparagine (N) and the fourth residue can be a valine (V).

Regarding motif (ii) in certain embodiments, the sixth residue of SEQ ID NO:79 (N-K-D-G-S-K/T-A-Y-N-E-D-G-T-V/A-K-Q/K-S-T-I-G-K-Y-N-E-K-Y-G-D-A-S) can be a lysine (K), the fourteenth residue can be a valine (V), and the sixteenth residue can be a glutamine (Q). Alternatively, the sixth residue can be a lysine (K), the fourteenth residue can be an alanine (A), and the sixteenth residue can be a glutamine (Q); or the sixth residue can be a lysine (K), the fourteenth residue can be an valine (V), and the sixteenth residue can be a lysine (K). Additional examples include where the sixth residue can be a threonine (T).

Regarding motif (iii) in certain embodiments, the ninth residue of SEQ ID NO:80 (L-P-T-D-G-K-M-D-N/K-S-D-V-E-L-Y-R-T-N/S-E) can be an asparagine (N) and the eighteenth residue can be an asparagine (N). Alternatively, the ninth residue can be an asparagine (N) and the eighteenth residue can be a serine (S), or the ninth residue can be a lysine (K) and the eighteenth residue can be an asparagine (N), or the ninth residue can be a lysine (K) and the eighteenth residue can be a serine (S).

Any of the above features regarding the location of motifs (i), (ii) and (iii) in a glucosyltransferase enzyme catalytic domain sequence can be used appropriately to detect one or more of these motifs. The relative positions of motifs (i) (SEQ ID NO:78), (ii) (SEQ ID NO:79) and (iii) (SEQ ID NO:80) align with residues 231-243, 396-425 and 549-567, respectively, of the GTF 7527 sequence (SEQ ID NO:65) shown in FIG. 2. In certain embodiments herein,

(A) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:78 in the glucosyltransferase catalytic domain aligns with amino acid positions 231-243 of SEQ ID NO:65;

(B) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:79 in the glucosyltransferase catalytic domain aligns with amino acid positions 396-425 of SEQ ID NO:65; and/or

(C) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:80 in the glucosyltransferase catalytic domain aligns with amino acid positions 549-567 of SEQ ID NO:65.

The relative position(s) of the amino acid sequence(s) detected in the glucosyltransferase catalytic domain can thus be determined with reference to the above amino acid positions in SEQ ID NO:65. For example, the sequence of a glucosyltransferase catalytic domain can be aligned with SEQ ID NO:65 using any means known in the art and/or as described above.

Alternatively, motif (i), (ii), and/or (iii) can be detected based on proximity to certain conserved sequences, namely SEQ ID NOs:72, 73, 74, 75, 76 and 77, as described above.

An identification method in some aspects can further comprise detecting a glucosyltransferase catalytic domain as presently disclosed. For example, a glucosyltransferase catalytic domain can be detected that comprises an amino acid

sequence that is at least 90% identical to amino acid positions 54-957 of SEQ ID NO:65, positions 55-960 of SEQ ID NO:30, positions 55-960 of SEQ ID NO:4, positions 55-960 of SEQ ID NO:28, and/or positions 55-960 of SEQ ID NO:20. Alternatively, a glucosyltransferase catalytic domain can be detected that comprises an amino acid sequence that is at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 98.5%, 99%, or 99.5% (but not 100%) identical to any of the foregoing sequences. In some embodiments, an identification method does not detect a glucosyltransferase catalytic domain sequence comprising residues 54-957 of SEQ ID NO:65, residues 55-960 of SEQ ID NO:30, residues 55-960 of SEQ ID NO:4, residues 55-960 of SEQ ID NO:28, or residues 55-960 of SEQ ID NO:20.

Certain of the conserved regions in FIG. 2 include catalytic active site motifs SEQ ID NOs:68, 69, 70, and 71 (refer to Example 3). Thus, a catalytic domain sequence of a glucosyltransferase enzyme in some aspects can be identified based on having one or more of SEQ ID NOs:68, 69, 70, and 71 in alignment, respectively, with SEQ ID NOs:68, 69, 70, and 71 as present in amino acids 54-957 of SEQ ID NO:65. Other conserved regions in FIG. 2 include SEQ ID NOs:72, 73, 74, 75, 76 and 77 (refer to Example 4). Thus, a catalytic domain sequence of a glucosyltransferase enzyme in some aspects can be identified based on having one or more of SEQ ID NOs:72, 73, 74, 75, 76 and 77 in alignment, respectively, with SEQ ID NOs:72, 73, 74, 75, 76 and 77 as present in amino acids 54-957 of SEQ ID NO:65.

Although it is believed that a glucosyltransferase enzyme herein need only have a catalytic domain sequence, such as one comprising an amino acid sequence that is at least 90% identical to amino acid positions 54-957 of SEQ ID NO:65 (or positions 55-960 of SEQ ID NO:30, positions 55-960 of SEQ ID NO:4, positions 55-960 of SEQ ID NO:28, or positions 55-960 of SEQ ID NO:20), a glucosyltransferase enzyme identified in a method herein is typically comprised within a larger amino acid sequence. For example, the catalytic domain may be linked at its C-terminus to a glucan-binding domain, and/or linked at its N-terminus to a variable domain and/or signal peptide.

The catalytic domain of a glucosyltransferase enzyme identified herein can have activity as exhibited by a catalytic domain of a glucosyltransferase classified under the glycoside hydrolase family 70 (GH70). Such a GH70 glucosyltransferase may be found in the CAZy (Carbohydrate-Active EnZymes) database (Cantarel et al., *Nucleic Acids Res.* 37:D233-238, 2009), for example.

A glucosyltransferase enzyme identified herein can synthesize insoluble poly alpha-1,3-glucan having at least 95% alpha-1,3 glycosidic linkages and DP_w of at least 100. In certain embodiments, an identified GTF enzyme can synthesize poly alpha-1,3-glucan in which at least about 95%, 96%, 97%, 98%, 99%, or 100% of the constituent glycosidic linkages are alpha-1,3 linkages. In such embodiments, accordingly, the glucosyltransferase enzyme synthesizes poly alpha-1,3-glucan in which there is less than about 5%, 4%, 3%, 2%, 1%, or 0% of glycosidic linkages that are not alpha-1,3.

In another aspect, a glucosyltransferase enzyme identified herein can synthesize poly alpha-1,3-glucan having no branch points or less than about 5%, 4%, 3%, 2%, or 1% branch points as a percent of the glycosidic linkages in the polymer. Examples of branch points include alpha-1,6 branch points.

In still another aspect, a glucosyltransferase enzyme identified herein can synthesize poly alpha-1,3-glucan having a molecular weight in DP_w or DP_n of at least about 100. Alternatively, the glucosyltransferase enzyme may synthesize poly alpha-1,3-glucan having a molecular weight in DP_n or DP_w of at least about 400. Alternatively still, the glucosyltransferase enzyme may synthesize poly alpha-1,3-glucan having a molecular weight in DP_n or DP_w of at least about 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, or 1000 (or any integer between 100 and 1000).

A glucosyltransferase enzyme identified herein can be further analyzed, if desired. For example, if one or more of motifs i, ii, and/or iii is deleted and/or mutated (such that the one or more motifs are no longer at least 90% identical to SEQ ID NO:78, 79, or 80, respectively) from an identified glucosyltransferase (parent GTF), the modified glucosyltransferase (child GTF) can be expected to produce a branched alpha-glucan polymer. A branched alpha-glucan polymer produced by a child GTF herein can have an intrinsic viscosity and/or branching index that is reduced by at least 30%, for example, compared to the intrinsic viscosity and/or branching index of poly alpha-1,3-glucan synthesized by the corresponding parent GTF. The intrinsic viscosity and/or branching index of an alpha-glucan polymer can be measured by any means known in the art, or as provided in the below Examples.

A glucosyltransferase enzyme identified in a method as presently disclosed can optionally be produced. Such production can be by any means known in the art. For

example, a glucosyltransferase enzyme can be produced recombinantly in a heterologous expression system, such as a microbial heterologous expression system (e.g., U.S. Pat. No. 7000000). Examples of heterologous expression systems include bacterial (e.g., *E. coli* such as TOP10*, Bacillus* sp.) and eukaryotic (e.g., yeasts such as *Pichia* sp. and *Saccharomyces* sp.) expression systems.

Non-limiting examples of compositions and methods disclosed herein include:

1. A reaction solution comprising water, sucrose, and a glucosyltransferase enzyme, wherein the glucosyltransferase enzyme comprises a catalytic domain comprising the following three motifs:
   (i) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:78,
   (ii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:79, and
   (iii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:80;

   wherein the glucosyltransferase enzyme does not comprise SEQ ID NO:4, 20, 28, 30, 65, or residues 54-957 of SEQ ID NO:65, residues 55-960 of SEQ ID NO:30, residues 55-960 of SEQ ID NO:4, residues 55-960 of SEQ ID NO:28, or residues 55-960 of SEQ ID NO:20;

   and wherein the glucosyltransferase enzyme produces insoluble poly alpha-1,3-glucan having at least 95% alpha-1,3 glycosidic linkages and a weight average degree of polymerization (DP_w) of at least 100.
2. The reaction solution of embodiment 1, wherein the catalytic domain comprises an amino acid sequence that is at least 90% identical to amino acid positions 54-957 of SEQ ID NO:65.
3. The reaction solution of embodiment 1 or 2, wherein:
   (A) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:78 aligns with amino acid positions 231-243 of SEQ ID NO:65;
   (B) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:79 aligns with amino acid positions 396-425 of SEQ ID NO:65; and/or

(C) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:80 aligns with amino acid positions 549-567 of SEQ ID NO:65.

4. The reaction solution of embodiment 1, 2, or 3, wherein motif (i) comprises SEQ ID NO:78, motif (ii) comprises SEQ ID NO:79, and motif (iii) comprises SEQ ID NO:80.

5. The reaction solution of embodiment 1, 2, 3, or 4, wherein the glucosyltransferase enzyme synthesizes poly alpha-1,3-glucan having 100% alpha-1,3 glycosidic linkages.

6. The reaction solution of embodiment 1, 2, 3, 4, or 5, wherein the glucosyltransferase enzyme synthesizes poly alpha-1,3-glucan having a DP_w of at least 400.

7. A method of producing insoluble poly alpha-1,3-glucan comprising:

(a) contacting at least water, sucrose, and a glucosyltransferase enzyme, wherein the glucosyltransferase enzyme comprises a catalytic domain comprising the following three motifs:

(i) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:78,

(ii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:79, and

(iii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:80,

and wherein the glucosyltransferase enzyme does not comprise SEQ ID NO:4, 20, 28, 30, 65, residues 54-957 of SEQ ID NO:65, residues 55-960 of SEQ ID NO:30, residues 55-960 of SEQ ID NO:4, residues 55-960 of SEQ ID NO:28, or residues 55-960 of SEQ ID NO:20;

whereby insoluble poly alpha-1,3-glucan is produced having at least 95% alpha-1,3 glycosidic linkages and a weight average degree of polymerization (DP_w) of at least 100; and

b) optionally, isolating the poly alpha-1,3-glucan produced in step (a).

8. The method of embodiment 7, wherein the catalytic domain comprises an amino acid sequence that is at least 90% identical to amino acid positions 54-957 of SEQ ID NO:65.

9. The method of embodiment 7 or 8, wherein:

(1) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:78 aligns with amino acid positions 231-243 of SEQ ID NO:65;

(2) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:79 aligns with amino acid positions 396-425 of SEQ ID NO:65; and/or

(3) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:80 aligns with amino acid positions 549-567 of SEQ ID NO:65.

10. The method of embodiment 7, 8, or 9, wherein motif (i) comprises SEQ ID NO:78, motif (ii) comprises SEQ ID NO:79, and motif (iii) comprises SEQ ID NO:80.

11. The method of embodiment 7, 8, 9, or 10, wherein insoluble poly alpha-1,3-glucan is produced in step (a) having 100% alpha-1,3 glycosidic linkages.

12. The method of embodiment 7, 8, 9, 10, or 11, wherein insoluble poly alpha-1,3-glucan is produced in step (a) having a DP_w of at least 400.

13. A method for identifying a glucosyltransferase enzyme, the method comprising:

detecting the presence of at least one motif in a glucosyltransferase catalytic domain, the at least one motif selected from the group consisting of:

(i) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:78,

(ii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:79, and

(iii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:80;

thereby identifying a glucosyltransferase enzyme that produces insoluble poly alpha-1,3-glucan having at least 95% alpha-1,3 glycosidic linkages and a weight average degree of polymerization (DP_w) of at least 100.

14. The method of embodiment 13, wherein the detecting step is performed:

(a) *in silico*,

(b) with a method comprising a nucleic acid hybridization step,

(c) with a method comprising a protein sequencing step, and/or

(d) with a method comprising a protein binding step.

15. The method of embodiment 13 or 14, wherein the detecting step comprises detecting the presence of each of motifs (i), (ii) and (iii) in the catalytic domain.

EXAMPLES

The present disclosure is further exemplified in the following Examples. It should be understood that these Examples, while indicating certain preferred aspects herein, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of the disclosed embodiments, and without departing from the spirit and scope thereof, can make various changes and modifications to adapt the disclosed embodiments to various uses and conditions.

Abbreviations

The meanings of some of the abbreviations used herein are as follows: "g" means gram(s), "h" means hour(s), "mL" means milliliter(s), "psi" means pound(s) per square inch, "wt%" means weight percentage, "μm" means micrometer(s), "°C" means degrees Celsius, "mg" means milligram(s), "mm" means millimeter(s), "μL" means microliter(s), "mmol" means millimole(s), "min" means minute(s), "mol%" means mole percent, "M" means molar, "rpm" means revolutions per minute, "MPa" means megaPascals, "IV" means intrinsic viscosity, "g'" means branching ratio.

GENERAL METHODS

Preparation of Crude Extracts of Glucosyltransferase (GTF) Enzymes

GTF enzymes were prepared as follows. *E. coli* TOP10® cells (Invitrogen, Carlsbad, CA) were transformed with a pJexpress404®-based construct containing a particular GTF-encoding DNA sequence. Each sequence was codon-optimized to express the GTF enzyme in *E. coli.* Individual *E. coli* strains expressing a particular GTF enzyme were grown in LB (Luria broth) medium (Becton, Dickinson and Company, Franklin Lakes, NJ) with ampicillin (100 μg/mL) at 37 °C with shaking to OD_{600} = 0.4-0.5, at which time IPTG (isopropyl beta-D-1-thiogalactopyranoside, Cat. No. I6758, Sigma-Aldrich, St. Louis, MO) was added to a final concentration of 0.5 mM. The cultures were incubated for 2-4 hours at 37 °C following IPTG induction. Cells were harvested by centrifugation at 5,000 x *g* for 15 minutes and resuspended (20% w/v) in 50 mM phosphate buffer pH 7.0 supplemented with dithiothreitol (DTT, 1.0 mM). Resuspended cells were passed through a French Pressure Cell (SLM Instruments, Rochester, NY) twice to ensure >95% cell lysis. Lysed cells were centrifuged for 30 minutes at 12,000 x *g* at 4 °C. The resulting supernatant was analyzed by the BCA (bicinchoninic acid)

protein assay (Sigma-Aldrich) and SDS-PAGE to confirm expression of the GTF enzyme, and the supernatant was stored at -20 °C.

Determination of GTF Enzymatic Activity

GTF enzyme activity was confirmed by measuring the production of reducing sugars (fructose and glucose) in a GTF reaction solution. A reaction solution was prepared by adding a GTF extract (prepared as above) to a mixture containing sucrose (50 or 150 g/L), potassium phosphate buffer (pH 6.5, 50 mM), and optionally dextran (1 mg/mL, dextran T10, Cat. No. D9260, Sigma-Aldrich); the GTF extract was added to 2.5%-5% by volume. The reaction solution was then incubated at 22-25 °C for 24-30 hours, after which it was centrifuged. Supernatant (0.01 mL) was added to a mixture containing 1 N NaOH and 0.1% triphenyltetrazolium chloride (Sigma-Aldrich). The mixture was incubated for five minutes after which its OD_{480} was determined using an ULTROSPEC spectrophotometer (Pharmacia LKB, New York, NY) to gauge the presence of the reducing sugars fructose and glucose.

Determination of Glycosidic Linkages

Glycosidic linkages in the glucan product synthesized by a GTF enzyme were determined by ^{13}C NMR (nuclear magnetic resonance). Dry glucan polymer (25-30 mg) was dissolved in 1 mL of deuterated dimethyl sulfoxide (DMSO) containing 3% by weight of LiCl with stirring at 50 °C. Using a glass pipet, 0.8 mL of the solution was transferred into a 5-mm NMR tube. A quantitative ^{13}C NMR spectrum was acquired using a Bruker Avance 500-MHz NMR spectrometer (Billerica, MA) equipped with a CPDUL cryoprobe at a spectral frequency of 125.76 MHz, using a spectral window of 26041.7 Hz. An inverse gated decoupling pulse sequence using waltz decoupling was used with an acquisition time of 0.629 second, an inter-pulse delay of 5 seconds, and 6000 pulses. The time domain data was transformed using an exponential multiplication of 2.0 Hz.

Determination of Number Average Degree of Polymerization (DP_n)

The DP_n of a glucan product synthesized by a GTF enzyme was determined by size-exclusion chromatography (SEC). Dry glucan polymer was dissolved at 5 mg/mL in N,N-dimethyl-acetamide (DMAc) and 5% LiCl with overnight shaking at 100 °C. The SEC system used was an Alliance™ 2695 separation module from Waters Corporation (Milford, MA) coupled with three on-line detectors: a differential refractometer 2410 from Waters, a multiangle light scattering photometer Heleos™ 8+ from Wyatt Technologies

(Santa Barbara, CA), and a differential capillary viscometer ViscoStar™ from Wyatt. The columns used for SEC were four styrene-divinyl benzene columns from Shodex (Japan) and two linear KD-806M, KD-802 and KD-801 columns to improve resolution at the low molecular weight region of a polymer distribution. The mobile phase was DMAc with 0.11% LiCl. The chromatographic conditions used were 50 °C in the column and detector compartments, 40 °C in the sample and injector compartment, a flow rate of 0.5 mL/min, and an injection volume of 100 μL. The software packages used for data reduction were Empower™ version 3 from Waters (calibration with broad glucan polymer standard) and Astra® version 6 from Wyatt (triple detection method with column calibration).

Determination of Intrinsic Viscosity

Multidetector size exclusion chromatography (SEC) allowed measurement of molar mass distribution (MMD) using a combination of light scattering (LS) photometer and differential refractometer (DR). Molar mass (M) of the separated fractions across the polymer distribution was measured as a ratio of two detector responses:

M ~ LS/DR, without any column calibration.

In a similar way, an in-line differential viscometer (DV) allowed measurement of intrinsic viscosity (IV) of the separated fractions:

IV ~ DV/DR.

By plotting IV as a function of M in log-log scale, a so-called Mark-Houwink plot was obtained for samples tested.

Determination of Branching Ratio

Mark-Houwink (MH) plots were useful for estimating the degree of branching in polymers through measuring their size as a function of molar mass. Thus, the hydrodynamic size (H) of the macromolecule in dilute solution was determined as H = IV x M, so that using an MH plot, it could be seen how the size of the polymer chain changes with its molar mass. Branched polymer has a smaller size in solution than its linear counterpart with the same molar mass, and the position of the MH-plot indicates the degree of polymer branching.

To quantify the degree of branching, the branching ratio (or branching index) g' was plotted as a function of molar mass. This index is defined as a ratio of hydrodynamic volume of branched polymer chain H_{br} with a given molar mass M, to the

similar volume H_{lin} of the linear chain with the same molar mass; i.e., $g'(M) = H_{br} / H_{lin}$. Since H is defined as a production of IV and M, and M is the same in both numerator and denominator, then g' could be determined for each separated fraction with molar mass M directly from the corresponding MH plots as $g' = IV_{br}/IV_{lin}$. These plots show how the degree of branching changes with the polymer molar mass. The weight-average branching index for each polymer (i.e., $g' = IV_{br,w}/IV_{lin,w}$) was a useful estimation of the overall branching frequency in the polydispersed polymer. A g' value of 1, per this analysis, indicates that a polymer is linear (unbranched), whereas a g' value < 1 indicates that a polymer is branched.

EXAMPLE 1

Production of GTF Enzymes

This Example describes the preparation of N-terminally truncated versions of glucosyltransferase (GTF) enzymes used in this study.

Nucleotide sequences encoding N-terminally truncated versions of GTF enzymes (Table 2, GTF ID) were synthesized using codons optimized for protein expression in *E. coli*. The nucleic acid products (Table 2, nt SEQ ID NO) encoding the GTF enzymes (Table 2, AA SEQ ID NO) were subcloned into pJexpresss404® (DNA2.0, Menlo Park, CA) to generate GTF expression plasmids (Table 2, plasmid ID). The GTF expression plasmids were used to transform *E. coli* TOP10 cells (Invitrogen, Carlsbad, CA) to generate GTF expression strains (Table 2, strain ID). Production of GTF enzymes by bacterial expression and determination of enzymatic activities were performed as described in General Methods.

Table 2

Production of GTF Enzymes

GTF ID	GI No.[a]	nt SEQ ID NO	AA SEQ ID NO	Plasmid ID	Strain ID
0874	450874	1	2	pMP53	TOP10/pMP53
6855	228476855	3	4	pMP66	TOP10/pMP66
2379	662379	5	6	pMP65	TOP10/pMP65
7527	47527	7	8	pMP52	TOP10/pMP52
1724	121724	9	10	pMP55	TOP10/pMP55
0544	290580544	11	12	pMP67	TOP10/pMP67
5926	167735926	13	14	pMP56	TOP10/pMP56
4297	7684297	15	16	pMP70	TOP10/pMP70
5618	328945618	17	18	pMP72	TOP10/pMP72

2765	322372765	19	20	pMP85	TOP10/pMP85
4700	21654700	21	22	pMP83	TOP10/pMP83
1366	146741366	23	24	pMP86	TOP10/pMP86
0427	940427	25	26	pMP87	TOP10/pMP87
2919	383282919	27	28	pMP88	TOP10/pMP88
2678	400182678	29	30	pMP89	TOP10/pMP89
2381	662381	31	32	pMP96	TOP10/pMP96
3929	387783929	33	34	pMP97	TOP10/pMP97
6907	228476907	35	36	pMP57	TOP10/pMP57
6661	228476661	37	38	pMP62	TOP10/pMP62
0339	334280339	39	40	pMP73	TOP10/pMP73
0088	3130088	41	42	pMP69	TOP10/pMP69
9358	24379358	43	44	pMP71	TOP10/pMP71
8242	325978242	45	46	pMP68	TOP10/pMP68
3442	324993442	47	48	pMP75	TOP10/pMP75
7528	47528	49	50	pMP77	TOP10/pMP77
3279	322373279	51	52	pMP79	TOP10/pMP79
6491	170016491	53	54	pMP74	TOP10/pMP74
6889	228476889	55	56	pMP60	TOP10/pMP60
4154	51574154	57	58	pMP80	TOP10/pMP80
3298	322373298		59	pMP98	TOP10/pMP98

[a] GI number as provided for each respective sequence in GENBANK database (NCBI).

EXAMPLE 2

Production of Glucan Polymer using GTF Enzymes

This Example describes using the GTF enzymes prepared in Example 1 to synthesize glucan polymer.

Polymerization reactions were performed with each of the GTF enzymes prepared in Example 1. Reaction solutions were prepared comprising sucrose (50 g/L), potassium phosphate buffer (pH 6.5, 20 mM) and a GTF enzyme (2.5% extract by volume). After 24-30 hours at 22-25 °C, insoluble glucan polymer product was harvested by centrifugation, washed three times with water, washed once with ethanol, and dried at 50 °C for 24-30 hours.

Glycosidic linkages in each insoluble glucan polymer product were determined by ^{13}C NMR, and the DP_n for each insoluble polymer product was determined by SEC, as described in General Methods. These measurements are provided in Table 3 below.

Table 3

Polymer produced by GTF enzymes

GTF ID	SEQ ID NO.	Reducing Sugars	Insoluble Product	Glucan Polymer Linkages		DP_n
				% 1,3	% 1,6	
0874	2	yes	yes	100	0	60
6855	4	yes	yes	100	0	440
2379	6	yes	yes	37	63	310
7527	8	yes	yes	100	0	440
1724	10	yes	yes	100	0	250
0544	12	yes	yes	62	36	980
5926	14	yes	yes	100	0	260
4297	16	yes	yes	31	67	800
5618	18	yes	yes	34	66	1020
2765	20	yes	yes	100	0	280
4700	22	yes	no			
1366	24	yes	no			
0427	26	yes	yes	100	0	120
2919	28	yes	yes	100	0	250
2678	30	yes	yes	100	0	390
2381	32	yes	no			
3929	34	yes	yes	100	0	280
6907	36	yes	no			
6661	38	yes	no			
0339	40	yes	no			
0088	42	yes	no			
9358	44	yes	no			
8242	46	yes	no			
3442	48	yes	no			
7528	50	yes	no			
3279	52	yes	no			
6491	54	yes	no			
6889	56	yes	no			
4154	58	yes	no			
3298	59	yes	no	50	50	
none	na	no	no			

The following GTF enzymes produced glucan polymers comprising at least 50% alpha-1,3-linkages and having a DP_n of at least 100: 6855 (SEQ ID NO:4), 7527 (SEQ ID NO:8), 1724 (SEQ ID NO:10), 0544 (SEQ ID NO:12), 5926 (SEQ ID NO:14), 2765 (SEQ ID NO:20), 0427 (SEQ ID NO:26), 2919 (SEQ ID NO:28), 2678 (SEQ ID NO:30), and 3929 (SEQ ID NO:34) (refer to Table 3). The following GTF enzymes produced glucan polymers comprising 100% alpha-1,3-linkages, indicating linear polymers: 6855 (SEQ ID NO:4), 7527 (SEQ ID NO:8), 1724 (SEQ ID NO:10), 5926 (SEQ ID NO:14),

2765 (SEQ ID NO:20), 0427 (SEQ ID NO:26), 2919 (SEQ ID NO:28), 2678 (SEQ ID NO:30), and 3929 (SEQ ID NO:34). These results clearly indicate that not all GTF enzymes are capable of producing linear alpha-1,3-glucan polymer.

EXAMPLE 3

Structure/Function Relationships Observed in GTF Sequences

This Example describes aligning the amino acid sequences of several GTF enzymes to determine whether they share any structures.

GTF enzymes were evaluated in Example 2 for their ability to produce glucan polymers with a focus on those enzymes that produce glucan with 100% alpha-1,3-linkages. The sequences of several of these enzymes were aligned with three dimensional structures that are formed by certain *S. mutans* and *L. reuteri* GTF sequences (3AIE [SEQ ID NO:66] and 3KLK [SEQ ID NO:67], respectively); the *S. mutans* and *L. reuteri* GTF sequences were aligned to superpose common tertiary structures using the software package MOE (Chemical Computing Group, Montreal, Canada). The sequences for each of the GTF enzymes used in the alignment contain the catalytic and glucan-binding domains of each enzyme, respectively (i.e., the N-terminal signal peptide and variable domains of each GTF are not included in the alignment). FIG. 2 shows the alignment. The sequences of the *S. mutans* and *L. reuteri* GTFs for which crystallographic structures are known were included in the alignment; *S. mutans* GTF is abbreviated as "3AIE" (SEQ ID NO:66) and *L. reuteri* GTF is abbreviated as "3KLK" (SEQ ID NO:67) in FIG. 2.

The alignment in FIG.2 indicates that all the aligned GTF sequences maintain numerous invariant regions (shown with dark background). These invariant sequences are located throughout the catalytic domain of each GTF (based on a homology model as opposed to an experimentally determined structure). The catalytic domains in the aligned GTFs are about 900-950 amino acid residues long and begin after position 1 (artificial start methionine) in each of the sequences shown in FIG.2. The sequence following the catalytic domain in each GTF represents the glucan-binding domain. The aligned GTF sequences share as little as 40% sequence identity with the sequences of the known GTF structures (*S. mutans* 3AIE and *L. reuteri* 3KLK). But the alignment of these sequences in FIG.2 indicates a distributed pattern of conserved sequence motifs and patterns of specific residues that are conserved in all the aligned sequences (residues with dark background in FIG.2). These conserved sequence motifs can be

related to important structural features such as the catalytic site described below and can serve as reference points to identify unique or characteristic features that may be associated with specific performance benefits.

The catalytic site residues may be found in sequence motifs repeated in all the aligned sequences (FIG. 2). Specifically, with reference to the sequence from GTF 7527 (SEQ ID NO:65) in FIG. 2, Arg292 and Asp294 are found in the motif FDxxRxDAxDNV (SEQ ID NO:68) corresponding to Arg475 and Asp477 of *S. mutans* 3AIE GTF and Arg1023 and Asp1025 of *L. reuteri* 3KLK GTF; Glu332 is found in the sequence motif ExWxxxDxxY (SEQ ID NO:69) corresponding to Glu515 in *S. mutans* 3AIE GTF and Glu1063 in *L. reuteri* 3KLK GTF; His434 and Asp435 are found in the sequence motif FxRAHD (SEQ ID NO:70) corresponding to His587 and Asp588 in *S. mutans* 3AIE GTF and His1135 and Asp1136 in *L. reuteri* 3KLK GTF; and Tyr(Y)783 is found in the sequence motif IxNGYAF (SEQ ID NO:71) corresponding to the residues Tyr916 of *S. mutans* 3AIE GTF and Tyr1465 of *L. reuteri* 3KLK GTF.

Thus, the tested GTF enzymes have catalytic domains comprising several highly conserved regions.

EXAMPLE 4

Sequence Motifs in GTF Enzymes that Synthesize High Molecular Weight Alpha-1,3-Glucan

The GTF enzymes whose sequences were aligned in FIG. 2 were further evaluated for their ability to produce glucan polymers with a focus on those enzymes that produce glucan with 100% alpha-1,3-linkages (Table 4).

Table 4

Polymer Produced by Various GTF Enzymes

GTF ID	SEQ ID NO.	Glucan Polymer Features			% Identity[d]	Cat. Domain Region[e]	% Cat. Domain Identity[f]
		% Alpha-1,3 Linkages[a]	DP_w50[b]	DP_w150[b]			
7527[c]	65	100	910	577	100	54-957	100
2678	30	100	740	657	94.1	55-960	94.9
6855	4	100	835	570	98.9	55-960	99.0
2919	28	100	600	414	93.1	55-960	95.5
2765	20	100	670		93.6	55-960	96.4
0088	42	<30			44.7	55-900	50.4
0544	12	62			46.7	55-900	51.2
0427	26	100	260		43.1	55-900	51.8

0874	2	100	105	50	43.3	55-900	52.0
1724	10	100	535	55	42.9	55-900	51.3
5926	14	100	475	68	46.0	55-900	50.9
1366	24	<30			46.1	55-900	50.9
3298	59	<30			44.1	55-910	49.8
2379	6	37			44.5	60-915	50.7
6907	36	<30			55.6	55-885	61.8
5618	18	34			46.2	55-905	51.4
4297	16	31			46.5	55-905	51.2
3442	48	<30			45.8	55-905	51.0
9358	44	<30			49.7	55-915	53.6
6661	38	<30			45.6	55-895	50.5
0339	40	<30			53.7	55-895	57.5
8242	46	<30			54.1	55-910	59.4
7528	50	<30			48.1	55-915	54.2
3279	52	<30			41.8	55-900	48.7

[a] Glucan products having <30% alpha-1,3 linkages were soluble and not further analyzed for DP_w.

[b] DP_w50 and DP_w150 represent, respectively, the DP_w of glucan produced by a GTF in a reaction solution having an initial sucrose concentration of 50 g/L or 150 g/L.

[c] SEQ ID NO:65 is a shorter version of the 7527 GTF of SEQ ID NO:8.

[d] Percent identity of respective GTF with SEQ ID NO:65 (per EMBOSS alignment).

[e] Amino acid position of region within catalytic domain sequence having conservation (FIG. 2) with other listed GTF sequences (approximate location).

[f] Percent identity of catalytic domain region with amino acid residues 54-957 of SEQ ID NO:65 (per EMBOSS alignment).

Nine of the aligned GTF enzymes were found to produce glucan with 100% alpha-1,3-linkages, and five of these nine enzymes produced high molecular weight polymer (DPw > 400, Table 4). Specifically, the five GTF enzymes that displayed the property of producing high molecular weight glucan with 100% alpha-1,3-linkages are 7527 (SEQ ID NO:65), 2678 (SEQ ID NO:30), 6855 (SEQ ID NO:4), 2919 (SEQ ID NO:28) and 2765 (SEQ ID NO:20). The sequences for each of these GTFs are indicated with a "++" in FIG. 2.

Three sequence motifs were found in the amino acid sequences of all five GTF enzymes that produce high molecular weight glucan with 100% alpha-1,3-linkages, and appear as three different "insertions" situated around the catalytic domain of the known GTF structures. Briefly, these sequence motifs are designated as:

Motif 1a (SEQ ID NO:78):

D/N-K-S-I/V-L-D-E-Q-S-D-P-N-H

Motif 2 (SEQ ID NO:79):

N-K-D-G-S-K/T-A-Y-N-E-D-G-T-V/A-K-Q/K-S-T-I-G-K-Y-N-E-K-Y-G-D-A-S

Motif 3a (SEQ ID NO:80):

L-P-T-D-G-K-M-D-N/K-S-D-V-E-L-Y-R-T-N/S-E

The relative positions of motifs 1a, 2 and 3a align with residues 231-243, 396-425 and 549-567, respectively, of the 7527 GTF sequence (SEQ ID NO:65) in FIG. 2. These motifs appear to be conserved among GTF enzymes that synthesize high molecular weight alpha-1,3-glucan.

In the alignment shown in FIG. 2, motif 1a is flanked by upstream and downstream sequences as shown in FIG. 3. Specifically, preceding motif 1a is the sequence SxxRxxN (SEQ ID NO:72), and following motif 1a is the sequence GGxxxLLxNDxDxSNPxVQAExLN (SEQ ID NO:73). Both of these sequences were found in all the aligned GTF sequences and can serve as reference points for identifying motif 1a in other GTF sequences. In the alignment shown in FIG. 2, motif 2 is flanked by upstream and downstream sequences as shown in FIG. 5. Specifically, preceding motif 2 by about 50 amino acids is the sequence WxxxDxxY (SEQ ID NO:74) and following motif 2 is the sequence YxFxRAHD (SEQ ID NO:75). The downstream sequence (SEQ ID NO:75) includes two of the active site residues, His587 and Asp588 (numbered with respect to the *S. mutans* GTF structure, 3AIE). Both of these sequences were found in all the aligned GTF sequences and can serve as reference points for identifying motif 2 in other GTF sequences. In the alignment shown in FIG. 2, motif 3a is flanked by upstream and downstream sequences as shown in FIG. 7. Specifically, preceding motif 3a is sequence YxxGGQ (SEQ ID NO:76) and following motif 3a is the sequence VRxG (SEQ ID NO:77). Both of these sequences were found in all the aligned GTF sequences and can serve as reference points for identifying motif 2 in other GTF sequences.

Identification of motifs 1a (SEQ ID NO:78), 2 (SEQ ID NO:79) and 3a (SEQ ID NO:80) in the catalytic domains of GTF enzymes that synthesize high molecular weight glucan having 100% alpha-1,3-glycosidic linkages indicates that each of these motifs may be useful for identifying other GTFs with similar activity.

EXAMPLE 5

Sequence Motifs in GTF Enzymes that Synthesize Low Molecular Weight Alpha-1,3-Glucan

Four GTF enzymes produced low molecular weight glucan having 100% alpha-1,3-linkages (Table 4). Specifically, these enzymes were 5926 (SEQ ID NO: 14), 0427 (SEQ ID NO: 26), 0874 (SEQ ID NO: 2) and 1724 (SEQ ID NO: 10). The sequences for each of these enzymes are indicated with a "+-" in FIG. 2. Two sequence motifs were found in the amino acid sequences of these GTF enzymes, and appear as two different "insertions" situated around the catalytic domain of the known GTF structures. Briefly, these sequence motifs are designated as:

Motif 1b (SEQ ID NO:81): D-S/P-R-F-T-Y/F-N-A/Q/P-N-D-P

Motif 3b (SEQ ID NO:82): I-G-N-G-E

The relative positions of motifs 1b and 3b align with residues 231-243 and 549-553, respectively, of the 7527 GTF sequence (SEQ ID NO:65) in FIG. 2. Identification of motifs 1b (SEQ ID NO:81) and 3b (SEQ ID NO:82) in the catalytic domains of GTF enzymes that synthesize low molecular weight glucan having 100% alpha-1,3-glycosidic linkages indicates that each of these unique motifs may be useful for identifying other GTFs with similarly activity.

EXAMPLE 6

Production of GTF Enzyme Lacking Sequence Motif 1a

A nucleotide sequence encoding a polypeptide similar to the 7527 GTF of SEQ ID NO:65, but with a deletion of Motif 1a (Example 4), was synthesized using codons optimized for expression in *E. coli* (DNA 2.0, Menlo Park CA). The nucleic acid product (SEQ ID NO:84), encoding GTF protein 7527-NT-dIS1a (SEQ ID NO:85), was subcloned into pJexpress404® (DNA 2.0, Menlo Park CA) to generate the plasmid identified as pMP101. Plasmid pMP101 was used to transform *E. coli* TOP10 cells to generate the strain identified as TOP10/pMP101. It is noted that a GTF catalytic domain sequence is located at amino acid positions 54-941 (approximate) of SEQ ID NO:85.

Production of 7527-NT-dIS1a enzyme (SEQ ID NO:85) with *E. coli* and production of glucan polymer using this enzyme were performed as described above (General Methods). The glucan product is insoluble, and likely comprises only alpha-glycosidic linkages. The intrinsic viscosity and branching of the glucan product (analyzed as described in General Methods) are listed in Table 5 below.

EXAMPLE 7

Production of GTF Enzyme Lacking Sequence Motif 2

A nucleotide sequence encoding a polypeptide similar to the 7527 GTF of SEQ ID NO:65, but with a deletion of Motif 2 (Example 4), was synthesized using codons optimized for expression in *E. coli* (DNA 2.0, Menlo Park CA). The nucleic acid product (SEQ ID NO:86), encoding GTF protein 7527-NT-dIS2 (SEQ ID NO:87), was subcloned into pJexpress404® to generate the plasmid identified as pMP102. Plasmid pMP102 was used to transform *E. coli* TOP10 cells to generate the strain identified as TOP10/pMP102. It is noted that a GTF catalytic domain sequence is located at amino acid positions 54-927 (approximate) of SEQ ID NO:87.

Production of 7527-NT-dIS2 (SEQ ID NO:87) with *E. coli* and production of glucan polymer using this enzyme were performed as described above (General Methods). The glucan product is insoluble, and likely comprises only alpha-glycosidic linkages. The intrinsic viscosity and branching of the glucan product (analyzed as described in General Methods) are listed in Table 5 below.

EXAMPLE 8

Production of GTF Enzyme Lacking Sequence Motif 3a

A nucleotide sequence encoding a polypeptide similar to the 7527 GTF of SEQ ID NO:65, but with a deletion of Motif 3a (Example 4), was synthesized using codons optimized for expression in *E. coli* (DNA 2.0, Menlo Park CA). The nucleic acid product (SEQ ID NO:88), encoding GTF protein 7527-NT-dIS3a (SEQ ID NO:89), was subcloned into pJexpress404® to generate the plasmid identified as pMP103. Plasmid pMP103 was used to transform *E. coli* TOP10 cells to generate the strain identified as TOP10/pMP103. It is noted that a GTF catalytic domain sequence is located at amino acid positions 54-935 (approximate) of SEQ ID NO:89.

Production of 7527-NT-dIS3a (SEQ ID NO:89) with *E. coli* and production of glucan polymer using this enzyme were performed as described above (General Methods). The glucan product is insoluble, and likely comprises only alpha-glycosidic linkages. The intrinsic viscosity and branching of the glucan product (analyzed as described in General Methods) are listed in Table 5 below.

EXAMPLE 9

Production of GTF Enzyme Lacking Sequence Motifs 1a and 2

A nucleotide sequence encoding a polypeptide similar to the 7527 GTF of SEQ ID NO:65, but with deletion of Motifs 1a and 2 (Example 4), was synthesized using codons optimized for expression in *E. coli* (DNA 2.0, Menlo Park CA). The nucleic acid product (SEQ ID NO:90), encoding GTF protein 7527-NT-dIS1a,2 (SEQ ID NO:91), was subcloned into pJexpress404® to generate the plasmid identified as pMP104. Plasmid pMP104 was used to transform *E. coli* TOP10 cells to generate the strain identified as TOP10/pMP104. It is noted that a GTF catalytic domain sequence is located at amino acid positions 54-911 (approximate) of SEQ ID NO:91.

Production of 7527-NT-dIS1a,2 (SEQ ID NO:91) with *E. coli* and production of glucan polymer using this enzyme were performed as described above (General Methods). The glucan product is insoluble, and likely comprises only alpha-glycosidic linkages. The intrinsic viscosity and branching of the glucan product (analyzed as described in General Methods) are listed in Table 5 below.

EXAMPLE 10

Production of GTF Enzyme Lacking Sequence Motifs 1a and 3a

A nucleotide sequence encoding a polypeptide similar to the 7527 GTF of SEQ ID NO:65, but with deletion of Motifs 1a and 3a (Example 4), was synthesized using codons optimized for expression in *E. coli* (DNA 2.0, Menlo Park CA). The nucleic acid product (SEQ ID NO:92), encoding GTF protein 7527-NT-dIS1a,3a (SEQ ID NO:93), was subcloned into pJexpress404® to generate the plasmid identified as pMP105. Plasmid pMP105 was used to transform *E. coli* TOP10 cells to generate the strain identified as TOP10/pMP105. It is noted that a GTF catalytic domain sequence is located at amino acid positions 54-919 (approximate) of SEQ ID NO:93.

Production of 7527-NT-dIS1a,3a (SEQ ID NO:93) with *E. coli* and production of glucan polymer using this enzyme were performed as described above (General Methods). The glucan product is insoluble, and likely comprises only alpha-glycosidic linkages. The intrinsic viscosity and branching of the glucan product (analyzed as described in General Methods) are listed in Table 5 below.

EXAMPLE 11

Production of GTF Enzyme Lacking Sequence Motifs 2 and 3a

A nucleotide sequence encoding a polypeptide similar to the 7527 GTF of SEQ ID NO:65, but with deletion of Motifs 2 and 3a (Example 4), was synthesized using codons optimized for expression in *E. coli* (DNA 2.0, Menlo Park CA). The nucleic acid product (SEQ ID NO:94), encoding GTF protein 7527-NT-dIS2,3a (SEQ ID NO:95), was subcloned into pJexpress404® to generate the plasmid identified as pMP106. Plasmid pMP106 was used to transform *E. coli* TOP10 cells to generate the strain identified as TOP10/pMP106. It is noted that a GTF catalytic domain sequence is located at amino acid positions 54-905 (approximate) of SEQ ID NO:95.

Production of 7527-NT-dIS2,3a (SEQ ID NO:95) with *E. coli* and production of glucan polymer using this enzyme were performed as described above (General Methods). The glucan product is insoluble, and likely comprises only alpha-glycosidic linkages. The intrinsic viscosity and branching of the glucan product (analyzed as described in General Methods) are listed in Table 5 below.

EXAMPLE 12

Production of GTF Enzyme Lacking Sequence Motifs 1a, 2 and 3a

A nucleotide sequence encoding a polypeptide similar to the 7527 GTF of SEQ ID NO:65, but with deletion of Motifs 1a, 2 and 3a (Example 4), was synthesized using codons optimized for expression in *E. coli* (DNA 2.0, Menlo Park CA). The nucleic acid product (SEQ ID NO:96), encoding GTF protein 7527-NT-dIS1a,2,3a (SEQ ID NO:97), was subcloned into pJexpress404® to generate the plasmid identified as pMP107. Plasmid pMP107 was used to transform *E. coli* TOP10 cells to generate the strain identified as TOP10/pMP107. It is noted that a GTF catalytic domain sequence is located at amino acid positions 54-889 (approximate) of SEQ ID NO:97.

Production of 7527-NT-dIS1a,2,3a (SEQ ID NO:97) with *E. coli* and production of glucan polymer using this enzyme were performed as described above (General Methods). The glucan product is insoluble, and likely comprises only alpha-glycosidic linkages. The intrinsic viscosity and branching of the glucan product (analyzed as described in General Methods) are listed in Table 5 below.

EXAMPLE 13

Analysis of Intrinsic Viscosity and Branching of Glucan Products Synthesized by GTF Enzymes

This Example describes measuring the intrinsic viscosity (IV) and branching (g') of glucan polymer synthesized by each of the deletion-containing GTF enzymes prepared in Examples 6-12. These measurements were compared to those obtained with glucan polymer produced by the 7527 GTF of SEQ ID NO:65, which does not have any internal deletions of Motifs 1a, 2 and/or 3a.

It is noted that the glucan polymer synthesized by 7527 GTF, poly alpha-1,3-glucan, has 100% alpha-1,3 linkages and is thus linear (see Table 4, for example).

The intrinsic viscosity and branching of glucan polymer samples produced by deletion-containing versions of 7527 GTF were analyzed as described in the General Methods, and are shown in Table 5 below. Non-deleted 7527 GTF is listed as "7527-NT" in Table 5. Glucan polymer produced by non-deleted 7527 GTF (control), which is listed as "7527-NT" in Table 5, was also analyzed.

Table 5

Intrinsic Viscosity (IV) and Branching Index (g') of Glucan Polymer Produced by Various GTF Enzymes

Enzyme ID	SEQ ID NO	Missing Motif(s)	Glucan Product Measurement	
			IV	g'
7527-NT	65	N/A	206	1.000
7527-NT-dIS1a	85	1a	94	0.410
7527-NT-dIS2	87	2	33	0.231
7527-NT-dIS3a	89	3a	28	0.268
7527-NT-dIS1a,2	91	1a and 2	21	0.261
7527-NT-dIS1a,3a	93	1a and 3a	18	0.215
7527-NT-dIS2,3a	95	2 and 3a	19	0.256
7527-NT-dIS1a,2,3a	97	1a, 2 and 3a	22	0.242

As shown in Table 5, glucan produced by each GTF enzyme missing at least one of Motifs 1a (motif i), 2 (motif ii), or 3a (motif iii) had decreased intrinsic viscosity (IV) and branching index (g'), as compared to glucan produced by the corresponding control GTF (7527-NT) having each of these motifs. Since reductions in either IV and/or g' indicate increased polymer branching, these results demonstrate that each of Motifs 1a, 2 and 3a

may be essential for certain GTF enzymes – ones that naturally contain each of these motifs – to produce linear alpha-1,3-glucan polymer.

This observation was not expected, given that some GTF enzymes that produce linear product do not contain any of Motifs 1a, 2, or 3a. For example, each of GTFs 5926, 0427, 0874, and 1724 produce poly alpha-1,3-glucan with 100% alpha-1,3 linkages (which is linear) (Table 4), despite not having any of these motifs. Indeed, since there appeared to be a correlation between the presence of Motifs 1a, 2 and 3a with increased glucan product molecular weight (see Example 4), it might have been more reasonable to have expected that Motif 1a, 2, and/or 3a removal would reduce glucan product molecular weight (instead of having an effect on branching).

Thus, GTF amino acid Motifs 1a, 2 and 3a play a role in production of linear poly alpha-1,3-glucan by those GTF enzymes that contain these motifs

EXAMPLE 14

GTF Catalytic Domain Activity

This Example describes testing catalytic domain sequences of certain GTFs for the ability to produce insoluble poly alpha-1,3-glucan. Specifically, catalytic domain sequences of GTFs 7527 (SEQ ID NO:65) and 5926 (SEQ ID NO:14) were tested for activity.

A GTF catalytic domain sequence having amino acid residues 54-957 of SEQ ID NO:65 was prepared using the heterologous expression techniques described above. Briefly, a DNA sequence (codon-optimized for expression in *E. coli*) encoding a methionine at the first amino acid position followed by amino acid residues 54-957 of SEQ ID NO:65 was prepared and used to express this catalytic domain sequence. This protein, compared to the amino acid sequence identified in GENBANK under GI number 47527 (SEQ ID NO:60), is truncated by 230 amino acids at the N-terminus and 384 amino acids at the C-terminus.

A GTF catalytic domain sequence having amino acid residues 57-906 of SEQ ID NO:14 was prepared using the heterologous expression techniques described above. Briefly, a DNA sequence (codon-optimized for expression in *E. coli*) encoding a methionine at the first amino acid position followed by amino acid residues 57-906 of SEQ ID NO:14 was prepared and used to express this catalytic domain sequence. This protein, compared to the amino acid sequence identified in GENBANK under GI number

167735926 (SEQ ID NO:83), is truncated by 199 amino acids at the N-terminus and 417 amino acids at the C-terminus.

The above procedures were followed to prepare reaction solutions containing either of these GTF catalytic domain sequences. The reactions were performed at 25 °C and the alpha-1,3-glucan produced in each reaction was analyzed for DP_w. The results are provided in Table 6.

Table 6

Alpha-1,3-Glucan Polymer Produced by Gtf Enzyme Catalytic Domains

Catalytic Domain Sequence	DP_w	Initial sucrose (g/L)	% Sucrose consumption
5926	108	150	100
7527	495	142	94

As shown in Table 6, catalytic domain sequences of GTF 7527 (residues 54-957 of SEQ ID NO:65) and GTF 5926 (residues 57-906 of SEQ ID NO:14) were able to catalyze production of poly alpha-1,3-glucan. The molecular weight of the poly alpha-1,3-glucan produced by each of these catalytic domain sequences generally corresponded with the molecular weight of the product produced by their counterparts containing both the catalytic domain and glucan binding domain (refer to activity of SEQ ID NOs:65 and 14 in Table 4, DP_w150).

Thus, the catalytic domain of a glucosyltransferase enzyme can be used to produce insoluble poly alpha-1,3-glucan in a reaction solution.

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A reaction solution comprising water, sucrose, and a glucosyltransferase enzyme, wherein said glucosyltransferase enzyme comprises a catalytic domain comprising the following three motifs:

(i) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:78,

(ii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:79, and

(iii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:80;

wherein said glucosyltransferase enzyme does not comprise residues 54-957 of SEQ ID NO:65, residues 55-960 of SEQ ID NO:30, residues 55-960 of SEQ ID NO:4, residues 55-960 of SEQ ID NO:28, or residues 55-960 of SEQ ID NO:20;

wherein the catalytic domain comprises an amino acid sequence that is at least 90% identical to amino acid positions 54-957 of SEQ ID NO:65;

wherein the initial concentration of the sucrose in the reaction solution is about 40 g/L to about 80 g/L;

and wherein the glucosyltransferase enzyme produces, from the reaction solution, insoluble poly alpha-1,3-glucan having at least 95% alpha-1,3 glycosidic linkages and a weight average degree of polymerization (DP_w) of at least 700.

2 The reaction solution of claim 1, wherein the initial concentration of the sucrose in the reaction solution is about 40 g/L to about 60 g/L.

3. The reaction solution of claim 1, wherein the initial concentration of the sucrose in the reaction solution is about 50 g/L.

4. The reaction solution of any one of claims 1 to 3, wherein the catalytic domain comprises an amino acid sequence that is at least 95% identical to amino acid positions 54-957 of SEQ ID NO:65.

5. The reaction solution of any one of claims 1 to 4, wherein:
   (A) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:78 aligns with amino acid positions 231-243 of SEQ ID NO:65;
   (B) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:79 aligns with amino acid positions 396-425 of SEQ ID NO:65; and/or
   (C) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:80 aligns with amino acid positions 549-567 of SEQ ID NO:65.

6. The reaction solution of any one of claims 1 to 5, wherein motif (i) comprises SEQ ID NO:78, motif (ii) comprises SEQ ID NO:79, and motif (iii) comprises SEQ ID NO:80.

7. The reaction solution of any one of claims 1 to 6, wherein said glucosyltransferase enzyme produces insoluble poly alpha-1,3-glucan having at least 99% alpha-1,3 glycosidic linkages.

8. The reaction solution of claim 7, wherein said glucosyltransferase enzyme produces insoluble poly alpha-1,3-glucan having 100% alpha-1,3 glycosidic linkages.

9. The reaction solution of any one of claims 1 to 8, wherein said glucosyltransferase enzyme produces insoluble poly alpha-1,3-glucan having a DP_w of at least 800.

10. A method of producing insoluble poly alpha-1,3-glucan comprising:
    contacting at least water, sucrose, and a glucosyltransferase enzyme, wherein said glucosyltransferase enzyme comprises a catalytic domain comprising the following three motifs:
    (i) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:78,

(ii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:79, and

(iii) a motif comprising an amino acid sequence that is at least 90% identical to SEQ ID NO:80;

wherein said glucosyltransferase enzyme does not comprise residues 54-957 of SEQ ID NO:65, residues 55-960 of SEQ ID NO:30, residues 55-960 of SEQ ID NO:4, residues 55-960 of SEQ ID NO:28, or residues 55-960 of SEQ ID NO:20;

wherein the catalytic domain comprises an amino acid sequence that is at least 90% identical to amino acid positions 54-957 of SEQ ID NO:65;

wherein the initial concentration of the sucrose in the contacting step is about 40 g/L to about 80 g/L;

whereby insoluble poly alpha-1,3-glucan is produced having at least 95% alpha-1,3 glycosidic linkages and a weight average degree of polymerization (DP_w) of at least 700.

11. The method of claim 10, wherein the initial concentration of the sucrose in the contacting step is about 40 g/L to about 60 g/L.

12. The method of claim 10, wherein the initial concentration of the sucrose in the contacting step is about 50 g/L.

13. The method of any one of claims 10 to 12, wherein the catalytic domain comprises an amino acid sequence that is at least 95% identical to amino acid positions 54-957 of SEQ ID NO:65.

14. The method of any one of claims 10 to 13, wherein:

(1) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:78 aligns with amino acid positions 231-243 of SEQ ID NO:65;

(2) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:79 aligns with amino acid positions 396-425 of SEQ ID NO:65; and/or

(3) the position of the amino acid sequence that is at least 90% identical to SEQ ID NO:80 aligns with amino acid positions 549-567 of SEQ ID NO:65.

15. The method of any one of claims 10 to 14, wherein motif (i) comprises SEQ ID NO:78, motif (ii) comprises SEQ ID NO:79, and motif (iii) comprises SEQ ID NO:80.

16. The method of any one of claims 10 to 15, wherein insoluble poly alpha-1,3-glucan is produced having a DP_w of at least 800.

17. The method of any one of claims 10 to 16, further comprising isolating the insoluble poly alpha-1,3-glucan.

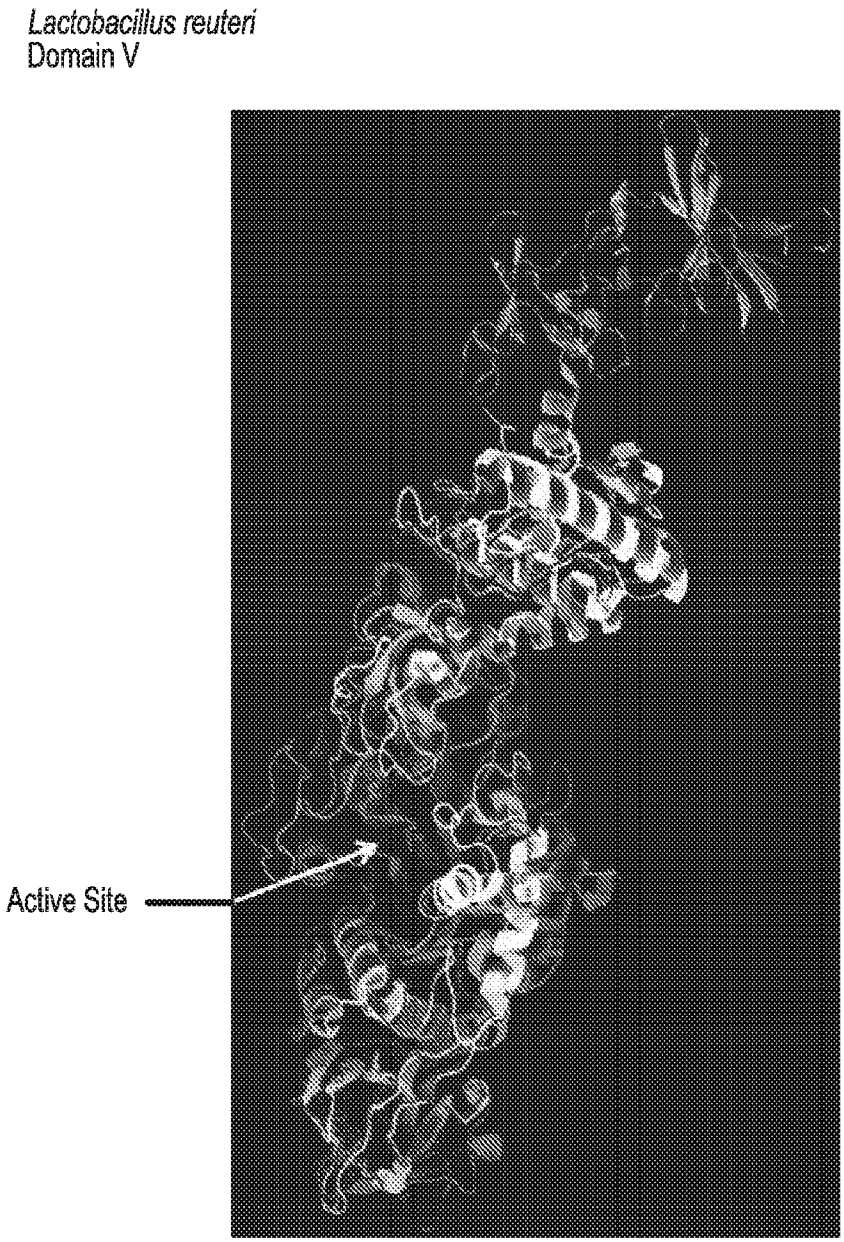

FIG. 1

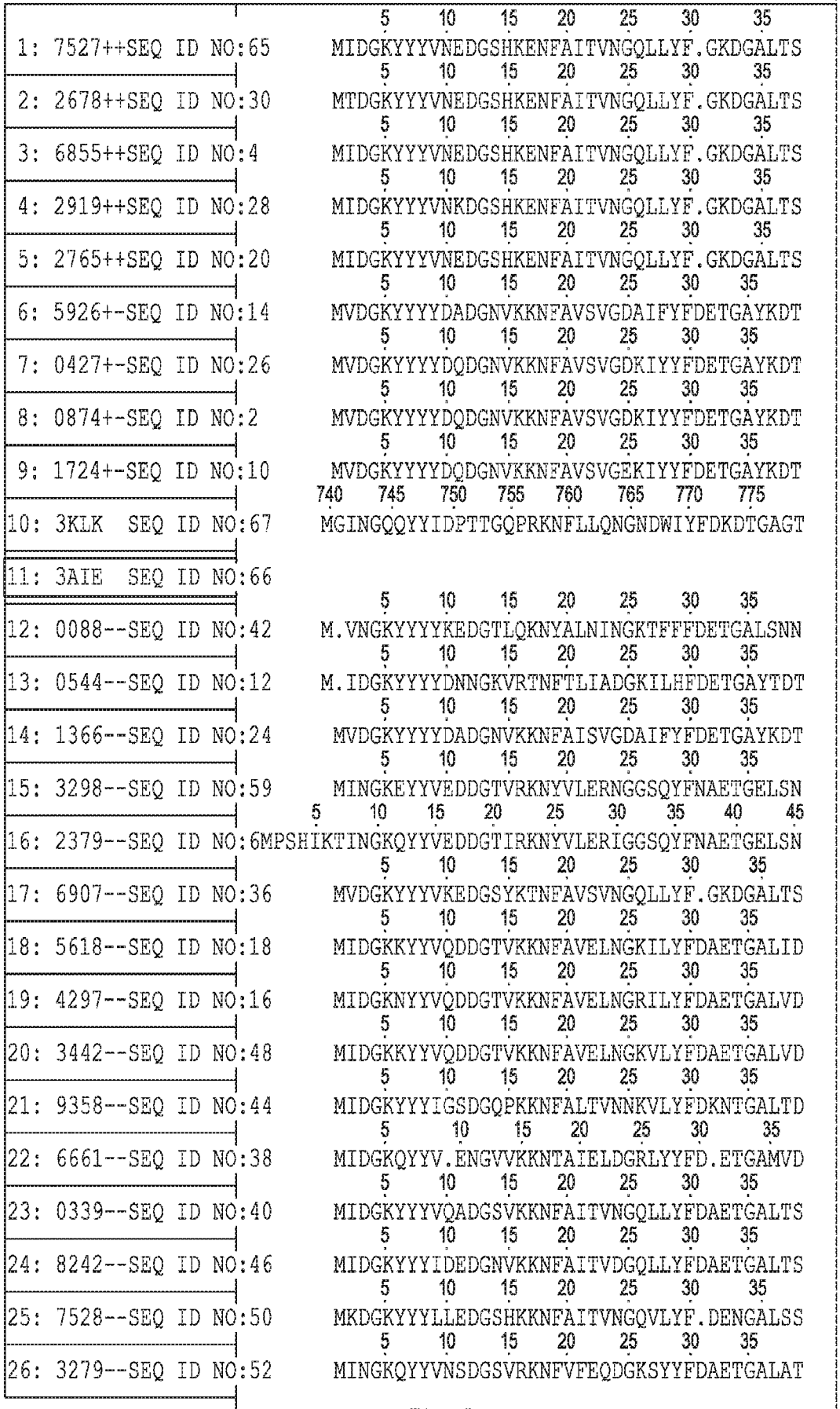
5 10 15 20 25 30 35
1: 7527++SEQ ID NO:65 MIDGKYYYVNEDGSHKENFAITVNGQLLYF.GKDGALTS
2: 2678++SEQ ID NO:30 MTDGKYYYVNEDGSHKENFAITVNGQLLYF.GKDGALTS
3: 6855++SEQ ID NO:4 MIDGKYYYVNEDGSHKENFAITVNGQLLYF.GKDGALTS
4: 2919++SEQ ID NO:28 MIDGKYYYVNKDGSHKENFAITVNGQLLYF.GKDGALTS
5: 2765++SEQ ID NO:20 MIDGKYYYVNEDGSHKENFAITVNGQLLYF.GKDGALTS
6: 5926+-SEQ ID NO:14 MVDGKYYYYDADGNVKKNFAVSVGDAIFYFDETGAYKDT
7: 0427+-SEQ ID NO:26 MVDGKYYYYDQDGNVKKNFAVSVGDKIYYFDETGAYKDT
8: 0874+-SEQ ID NO:2 MVDGKYYYYDQDGNVKKNFAVSVGDKIYYFDETGAYKDT
9: 1724+-SEQ ID NO:10 MVDGKYYYYDQDGNVKKNFAVSVGEKIYYFDETGAYKDT
740 745 750 755 760 765 770 775
10: 3KLK SEQ ID NO:67 MGINGQQYYIDPTTGQPRKNFLLQNGNDWIYFDKDTGAGT
11: 3AIE SEQ ID NO:66
5 10 15 20 25 30 35
12: 0088--SEQ ID NO:42 M.VNGKYYYYKEDGTLQKNYALNINGKTFFFDETGALSNN
13: 0544--SEQ ID NO:12 M.IDGKYYYYDNNGKVRTNFTLIADGKILHFDETGAYTDT
14: 1366--SEQ ID NO:24 MVDGKYYYYDADGNVKKNFAISVGDAIFYFDETGAYKDT
15: 3298--SEQ ID NO:59 MINGKEYYVEDDGTVRKNYVLERNGGSQYFNAETGELSN
5 10 15 20 25 30 35 40 45
16: 2379--SEQ ID NO:6MPSHIKTINGKQYYVEDDGTIRKNYVLERIGGSQYFNAETGELSN
17: 6907--SEQ ID NO:36 MVDGKYYYVKEDGSYKTNFAVSVNGQLLYF.GKDGALTS
18: 5618--SEQ ID NO:18 MIDGKKYYVQDDGTVKKNFAVELNGKILYFDAETGALID
19: 4297--SEQ ID NO:16 MIDGKNYYVQDDGTVKKNFAVELNGRILYFDAETGALVD
20: 3442--SEQ ID NO:48 MIDGKKYYVQDDGTVKKNFAVELNGKVLYFDAETGALVD
21: 9358--SEQ ID NO:44 MIDGKYYYIGSDGQPKKNFALTVNNKVLYFDKNTGALTD
22: 6661--SEQ ID NO:38 MIDGKQYYV.ENGVVKKNTAIELDGRLYYFD.ETGAMVD
23: 0339--SEQ ID NO:40 MIDGKYYYVQADGSVKKNFAITVNGQLLYFDAETGALTS
24: 8242--SEQ ID NO:46 MIDGKYYYIDEDGNVKKNFAITVDGQLLYFDAETGALTS
25: 7528--SEQ ID NO:50 MKDGKYYYLLEDGSHKKNFAITVNGQVLYF.DENGALSS
26: 3279--SEQ ID NO:52 MINGKQYYVNSDGSVRKNFVFEQDGKSYYFDAETGALAT

FIG. 2

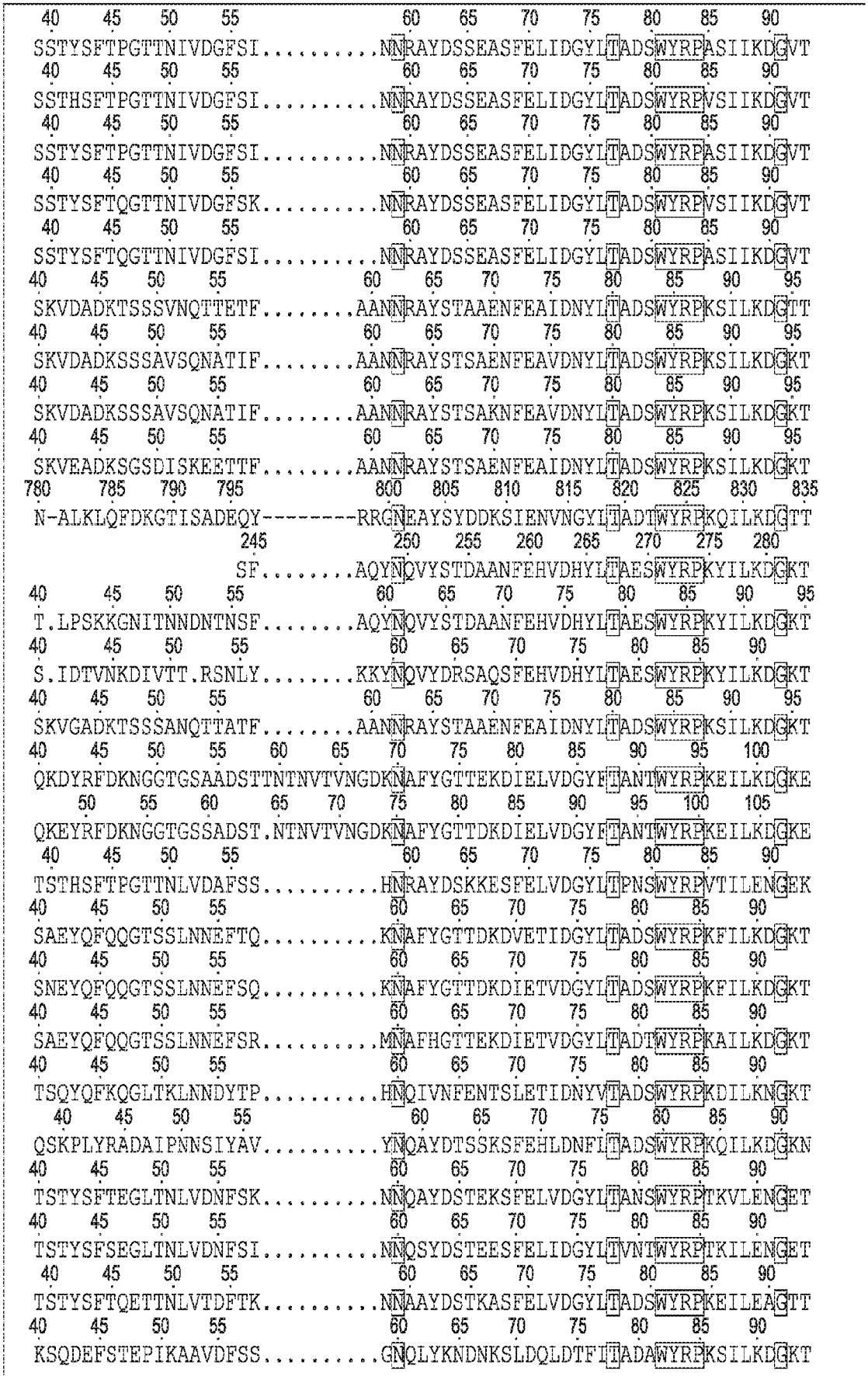

FIG. 2 Continued

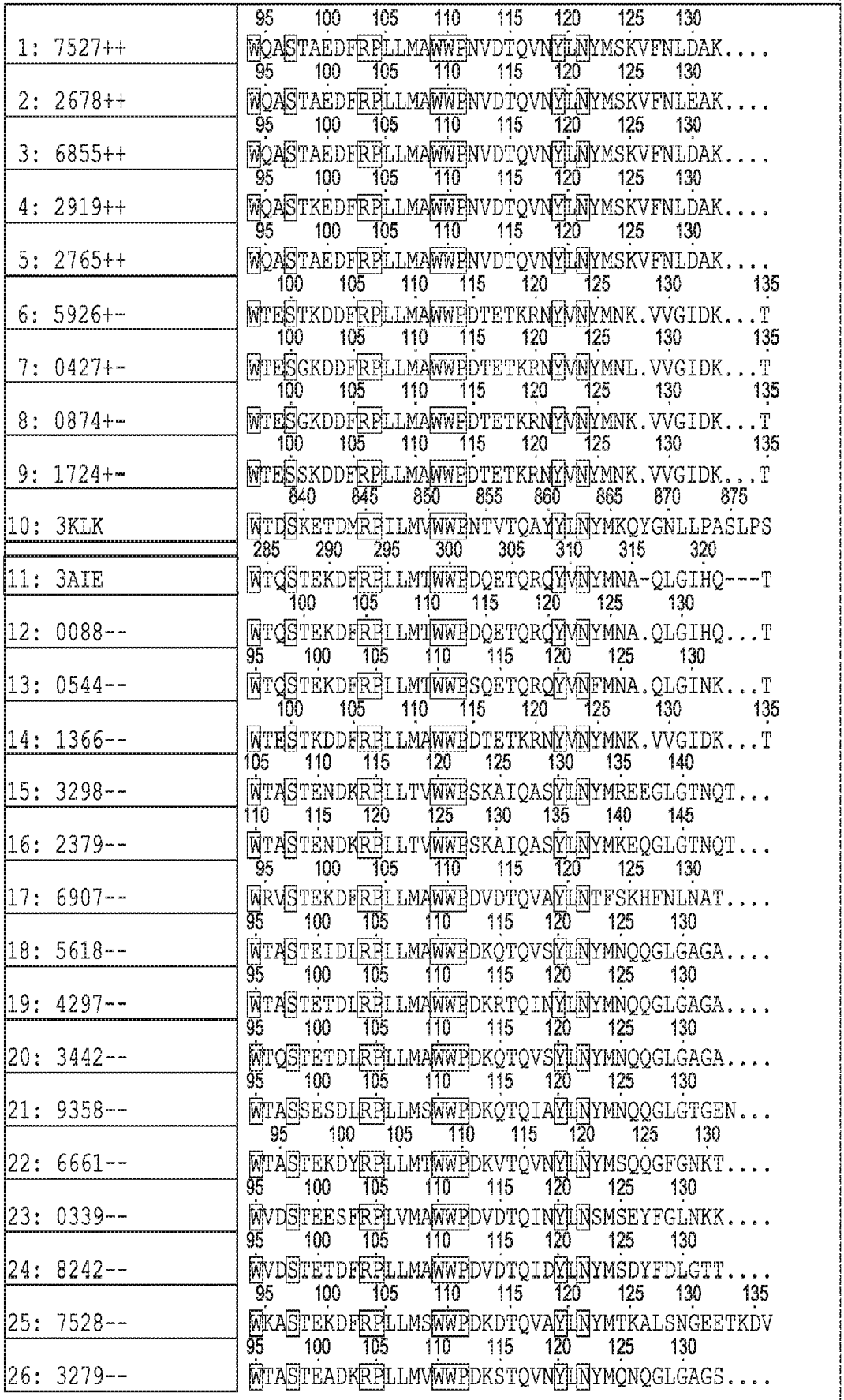

FIG. 2 Continued

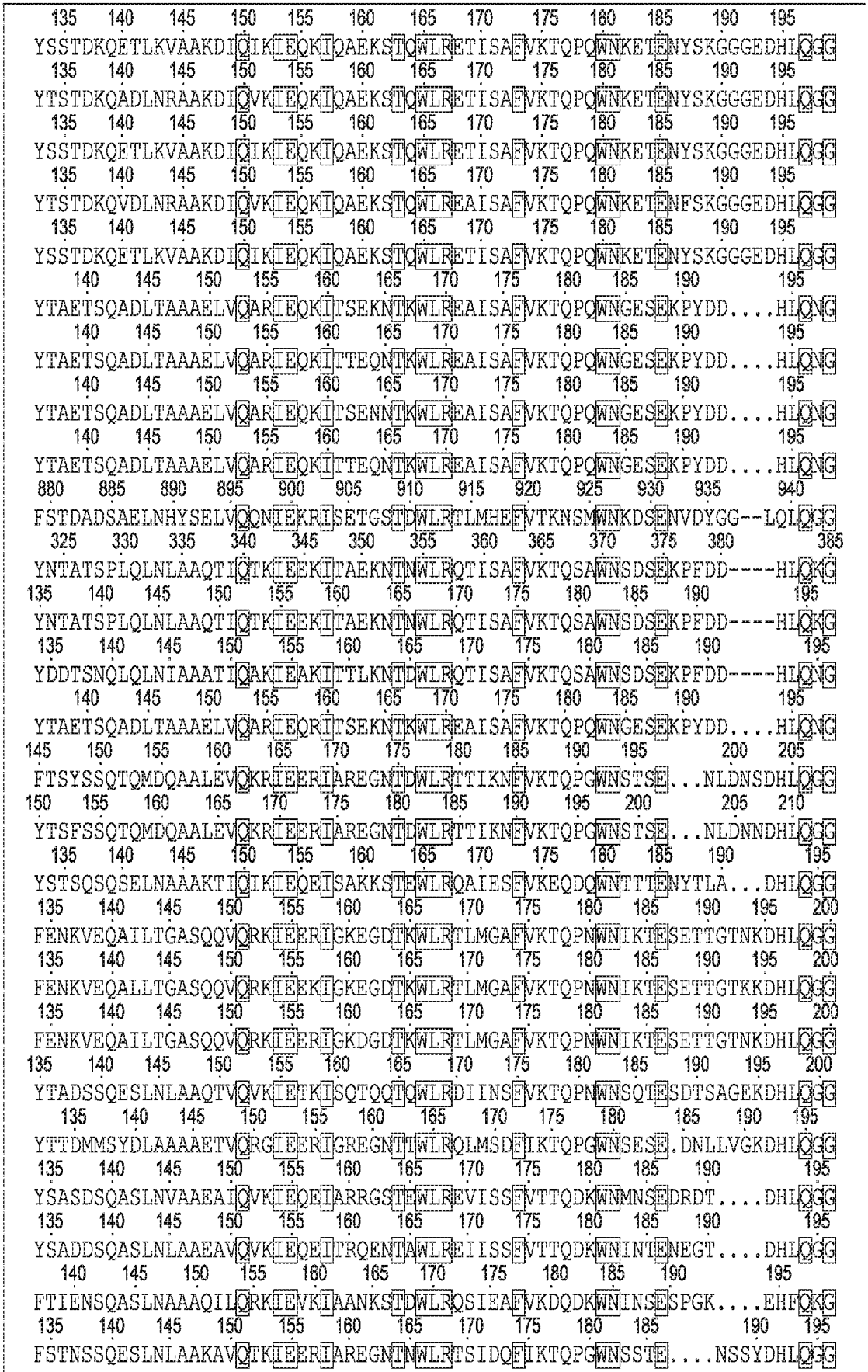

FIG. 2 Continued

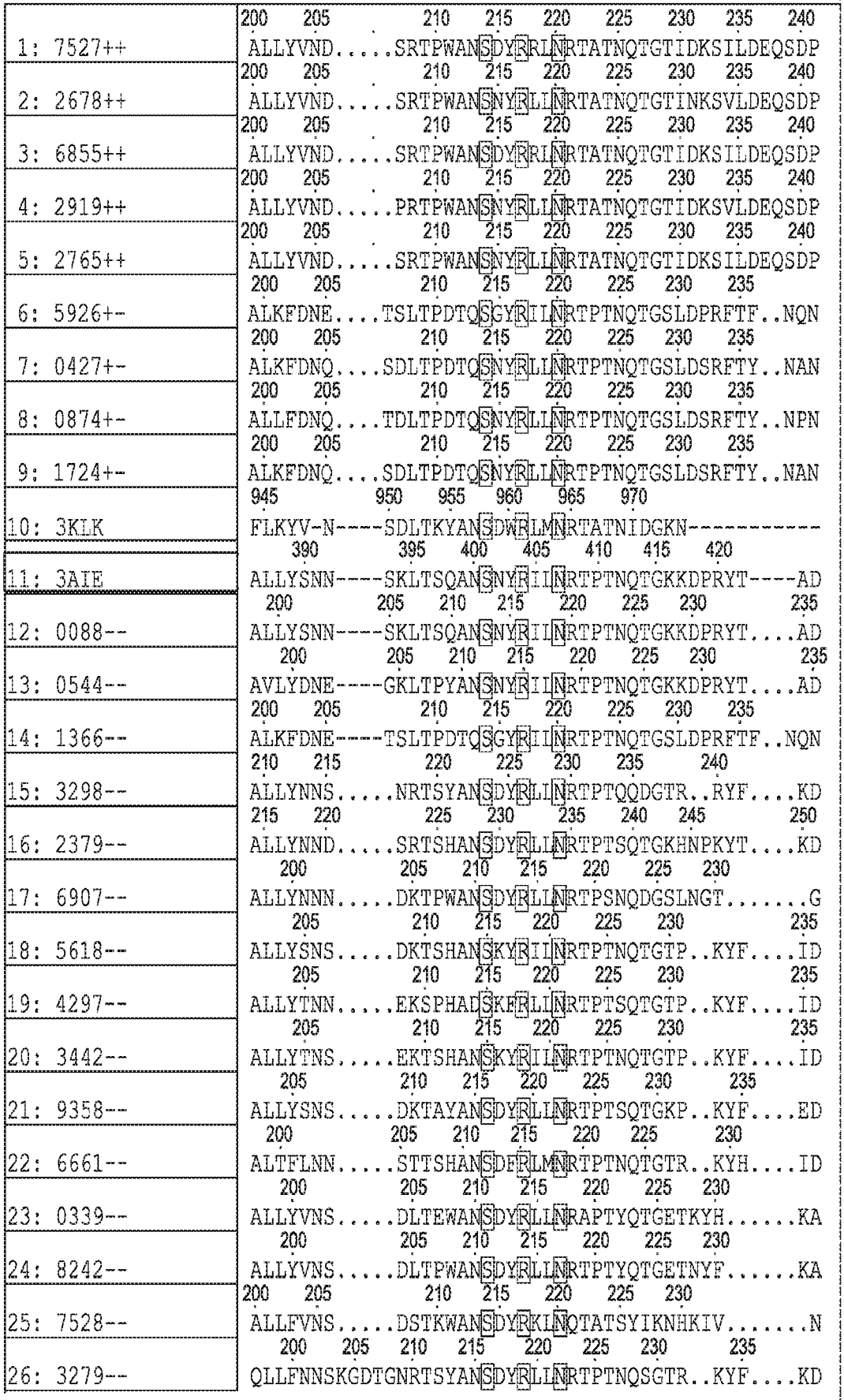

FIG. 2 Continued

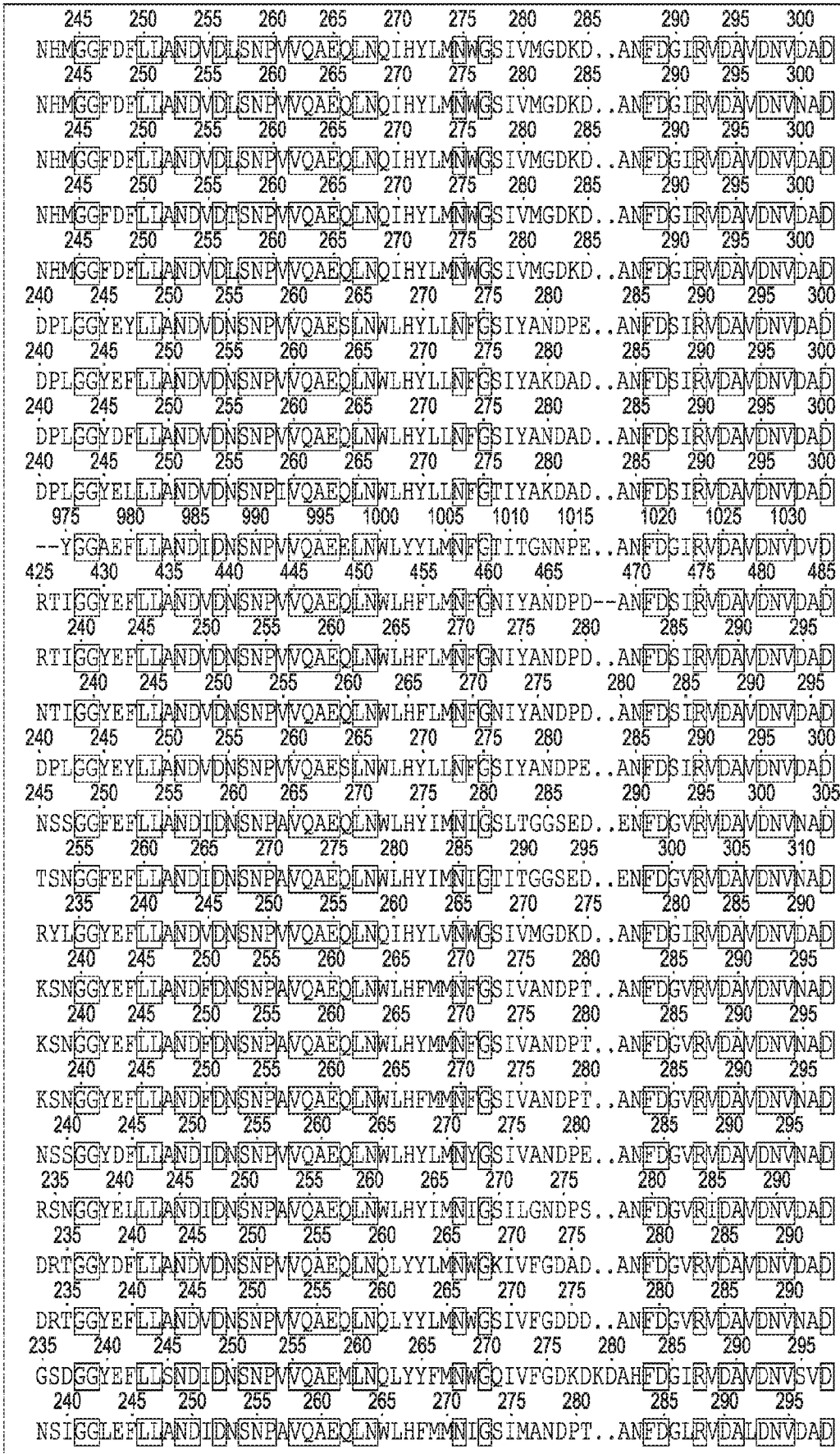

FIG. 2 Continued

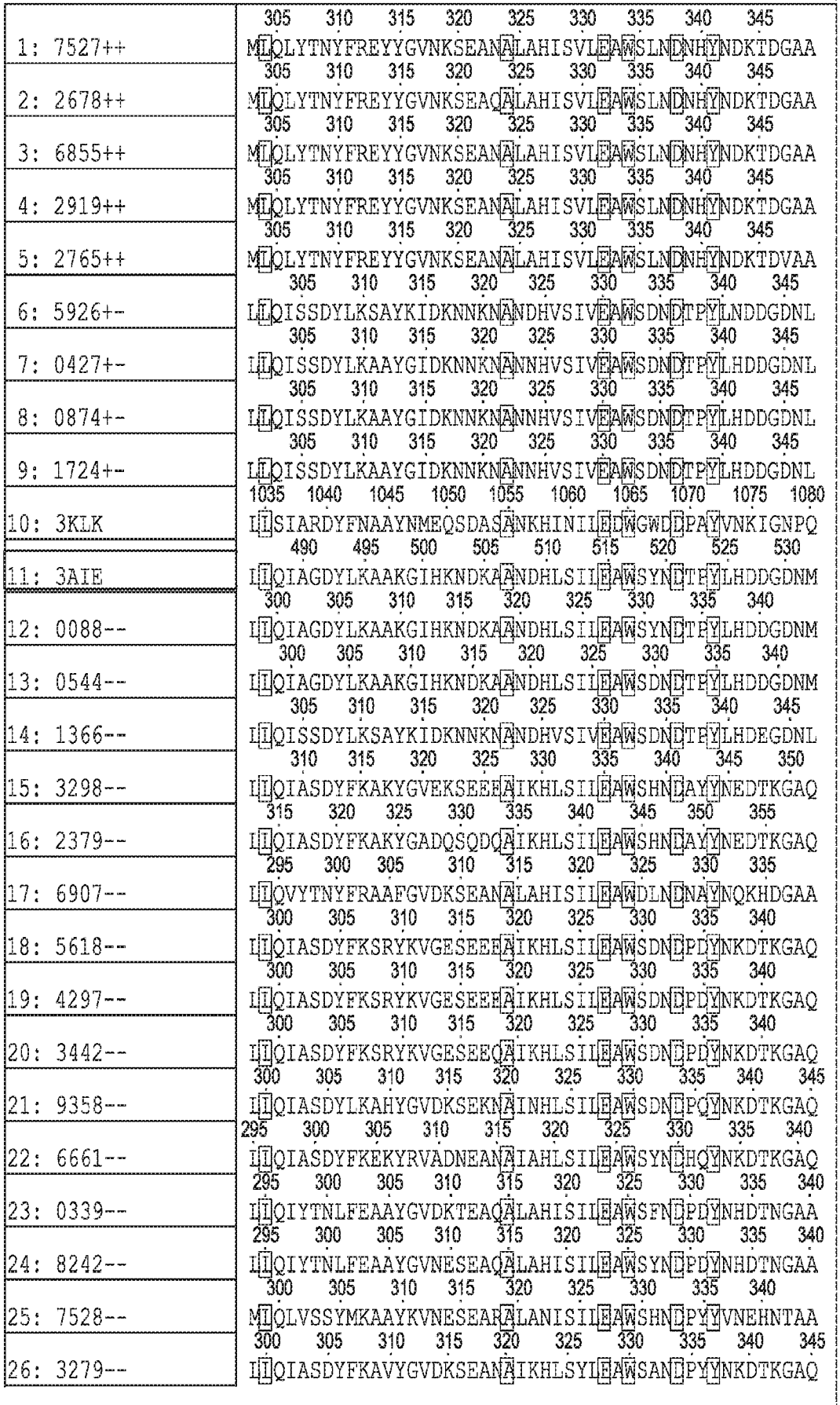

FIG. 2 Continued

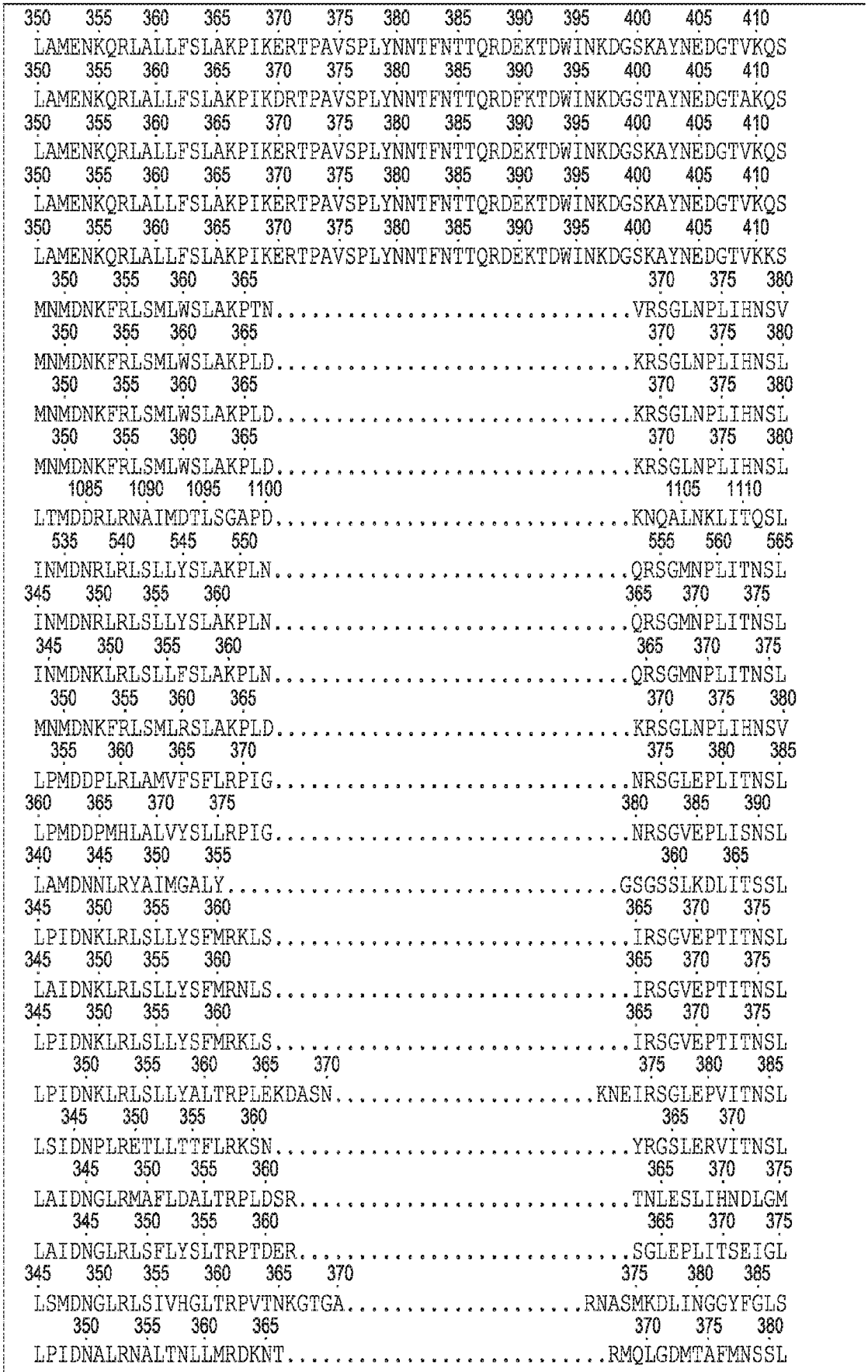

FIG. 2 Continued

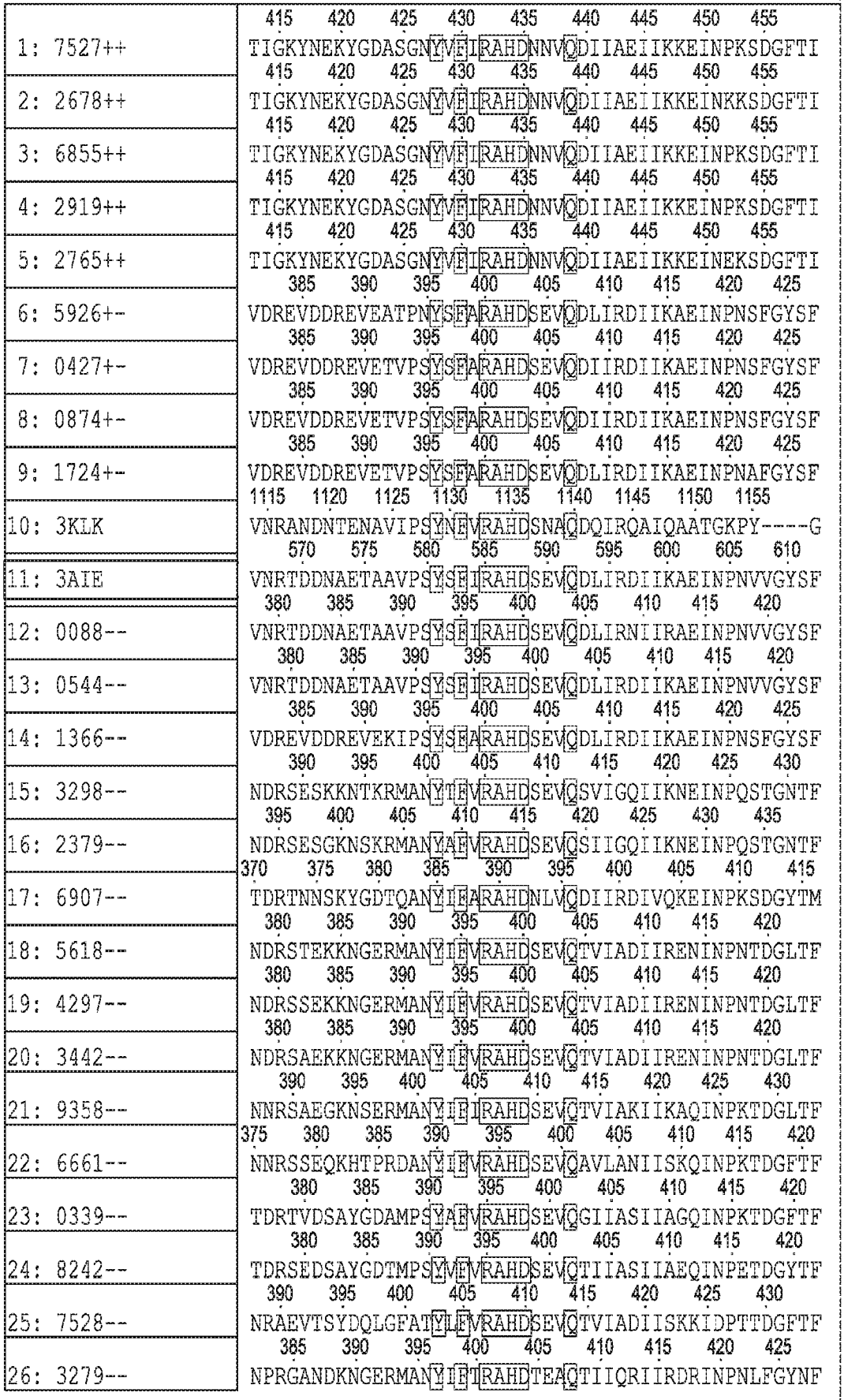

FIG. 2 Continued

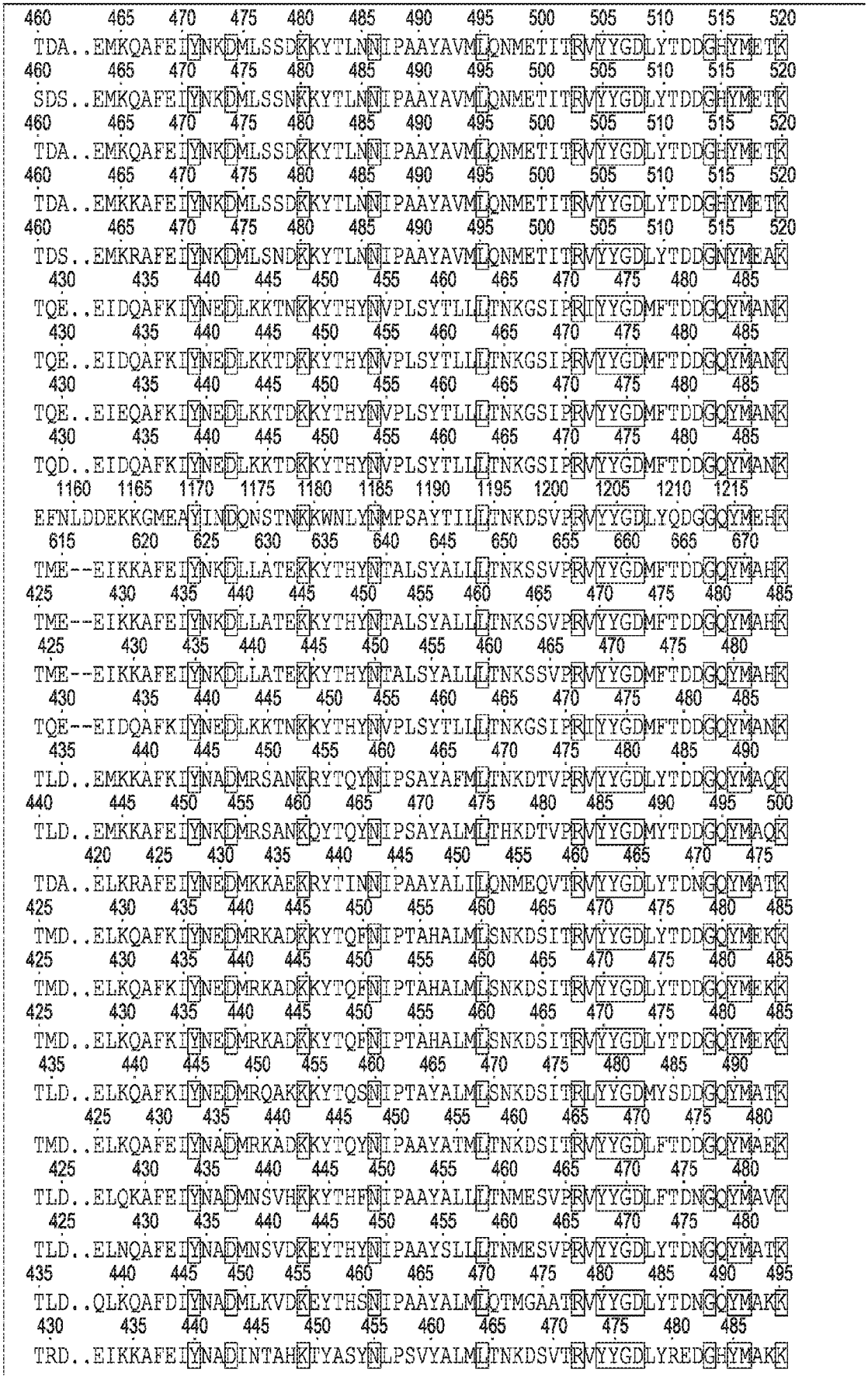

FIG. 2 Continued

17 Jun 2020
2016280700

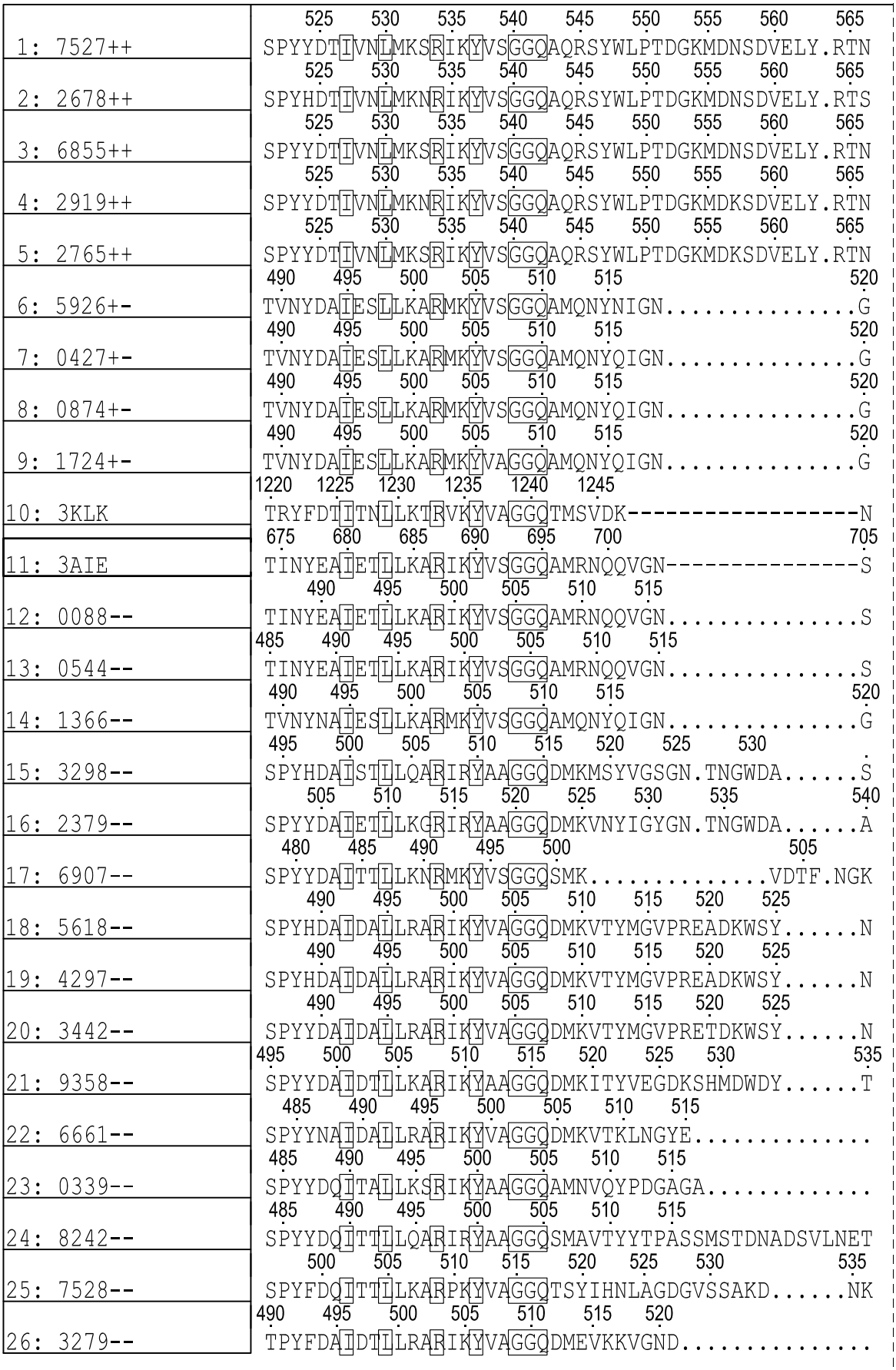

FIG. 2 Continued

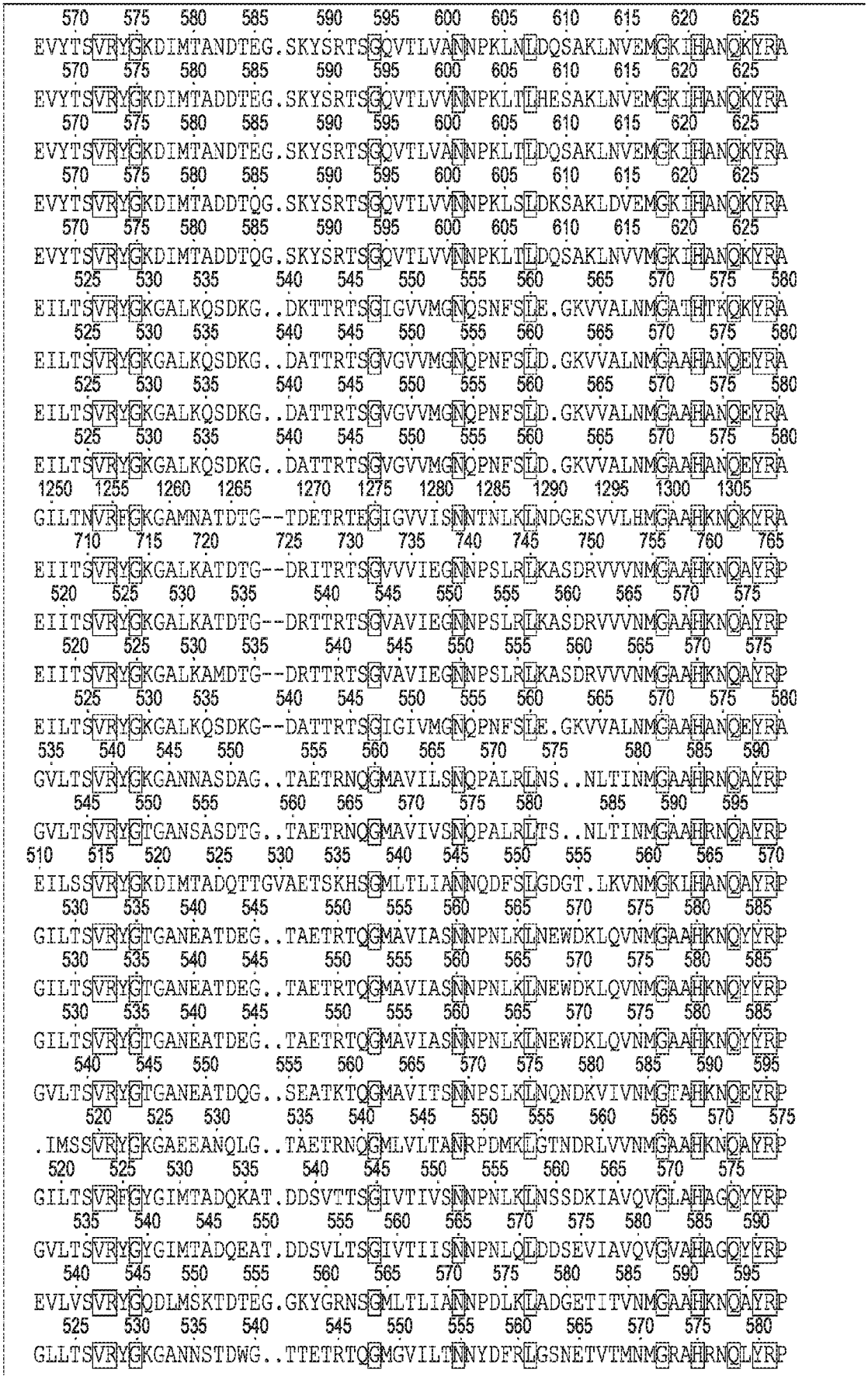

FIG. 2 Continued

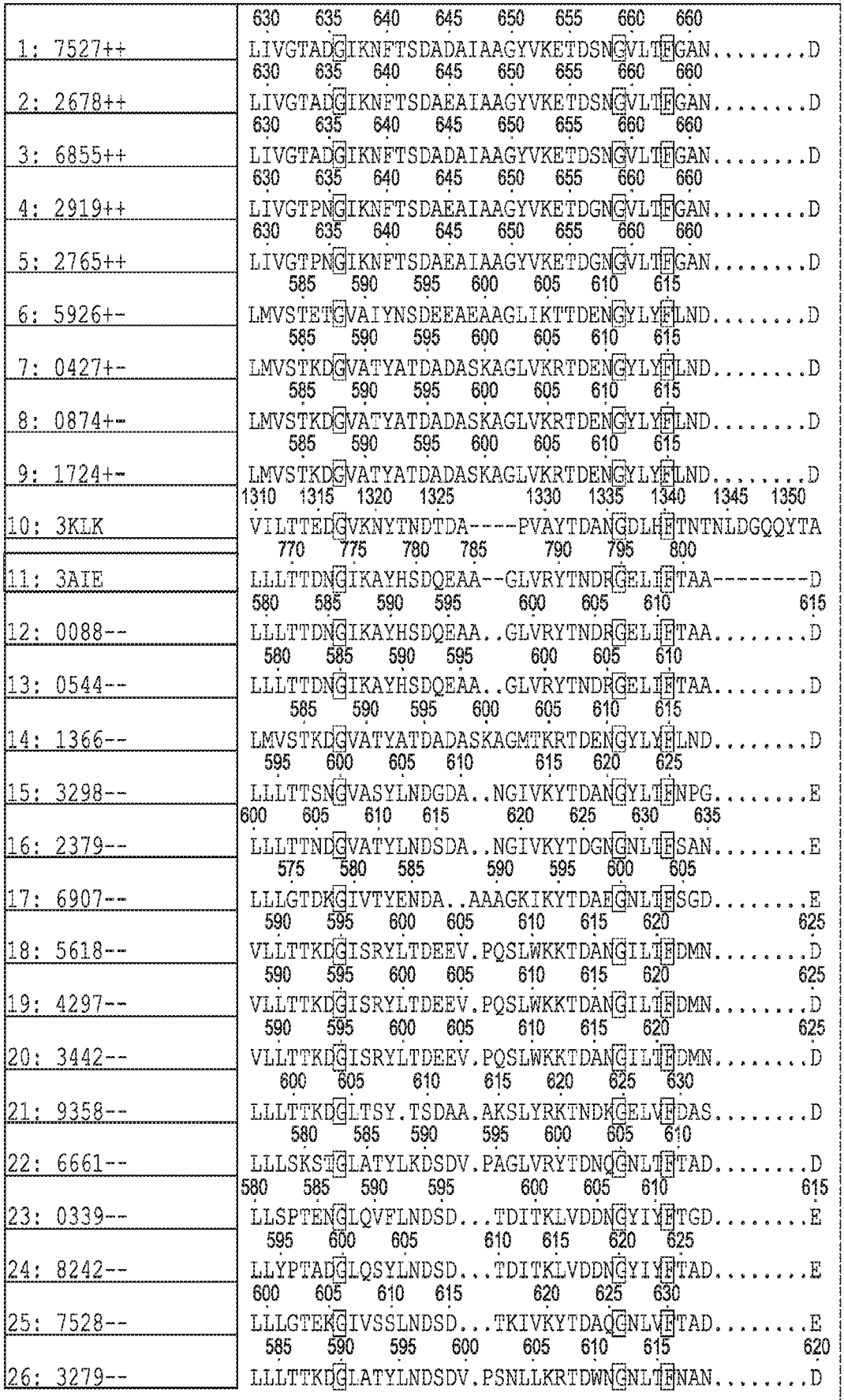

FIG. 2 Continued

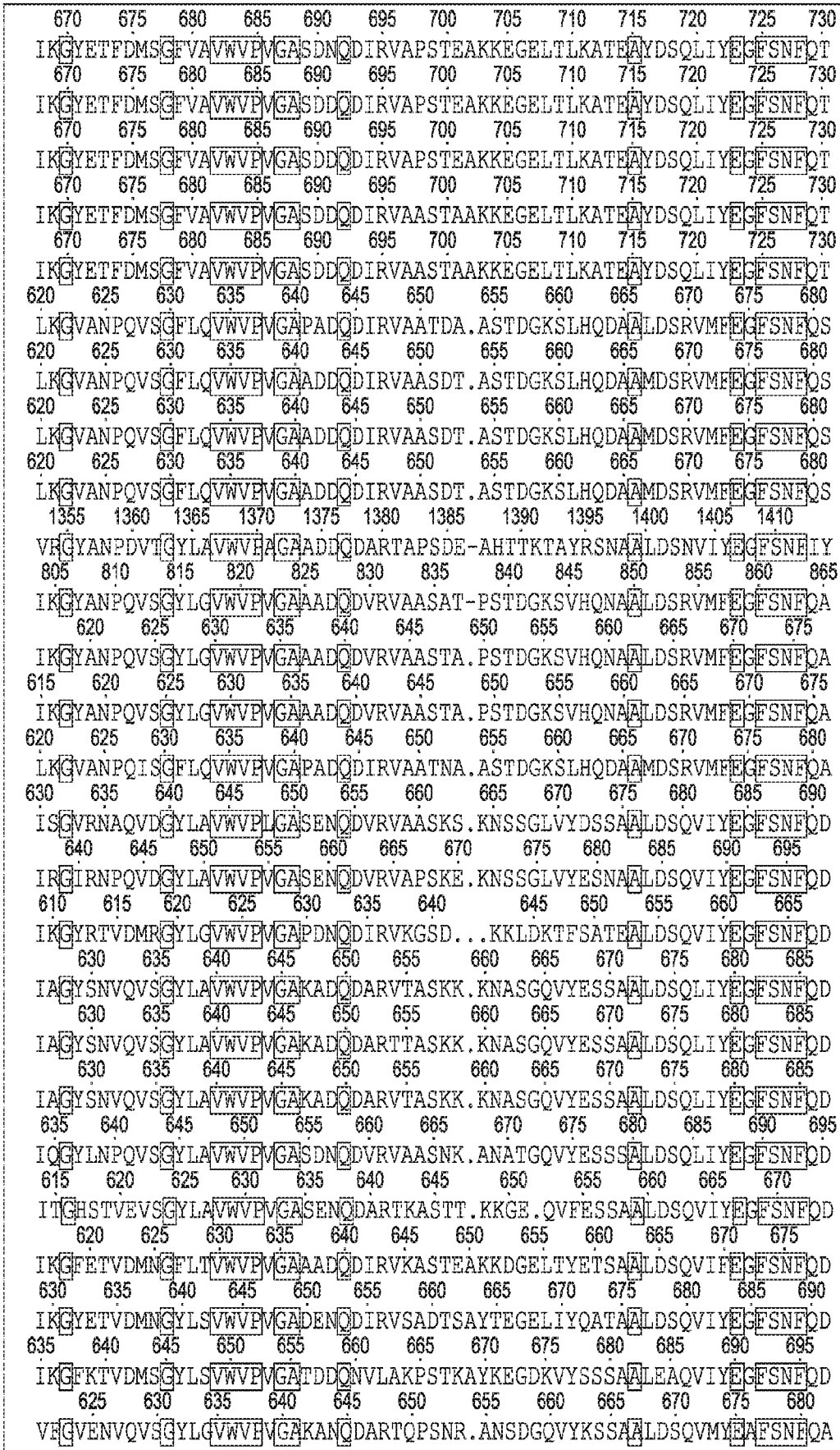

FIG. 2 Continued

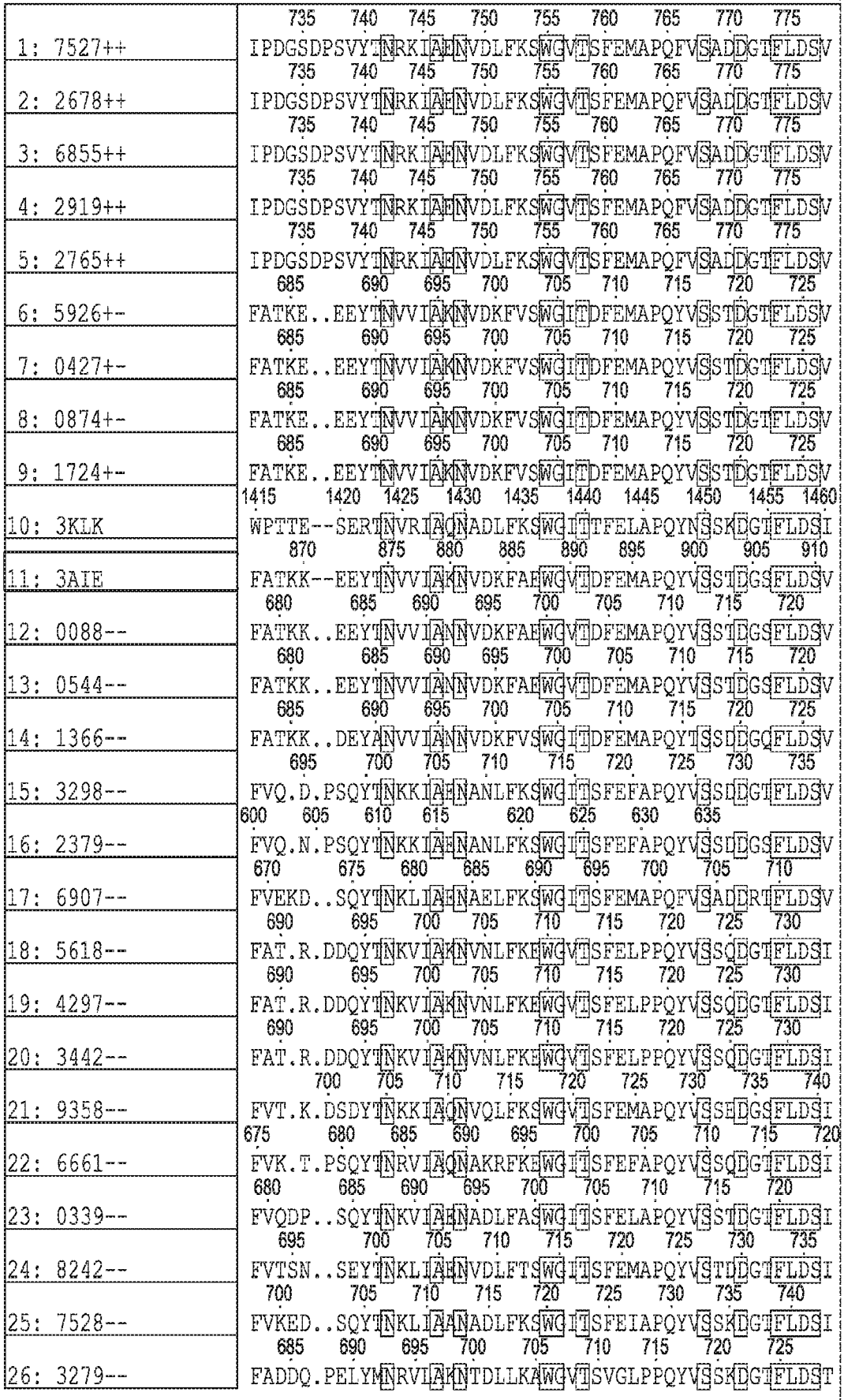

FIG. 2 Continued

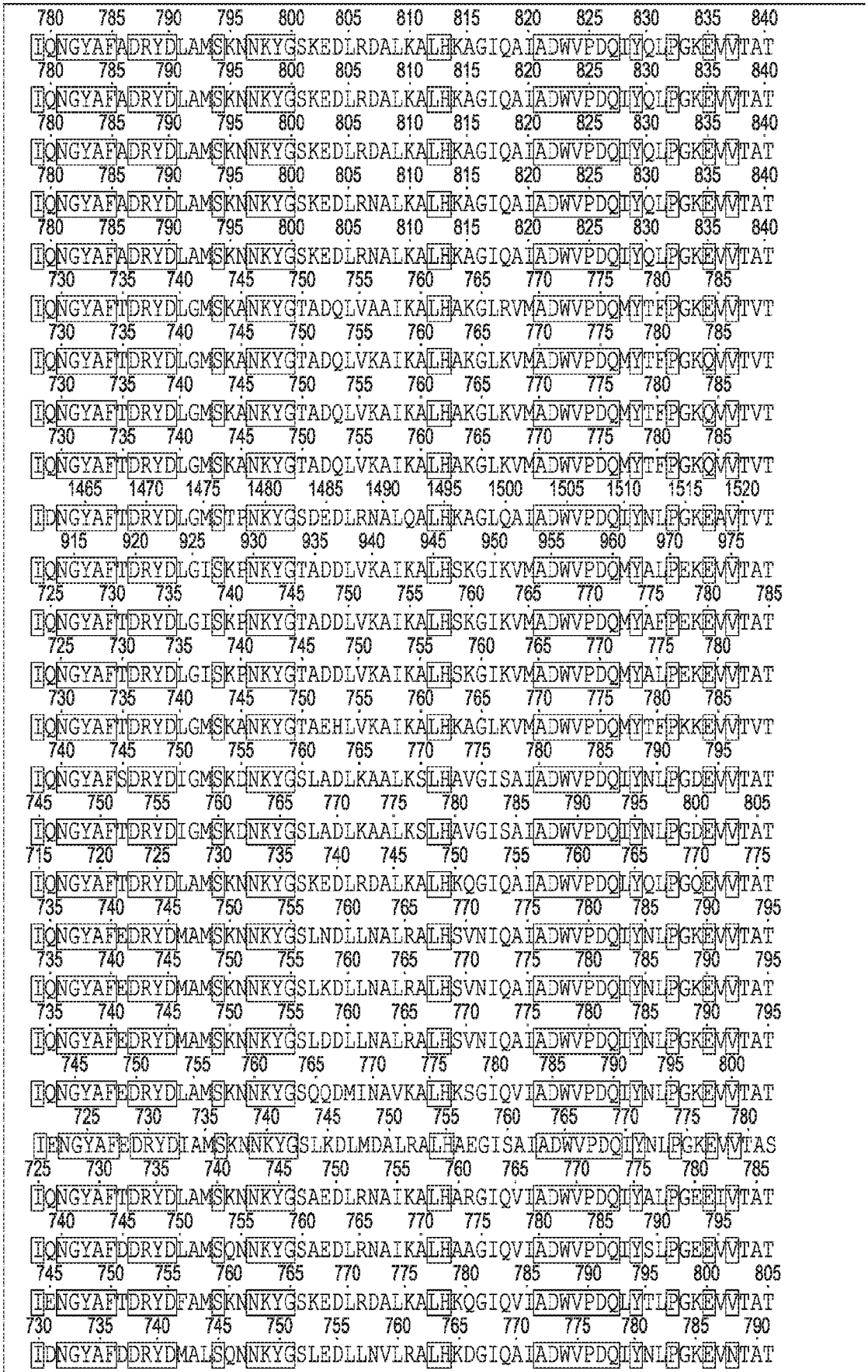

FIG. 2 Continued

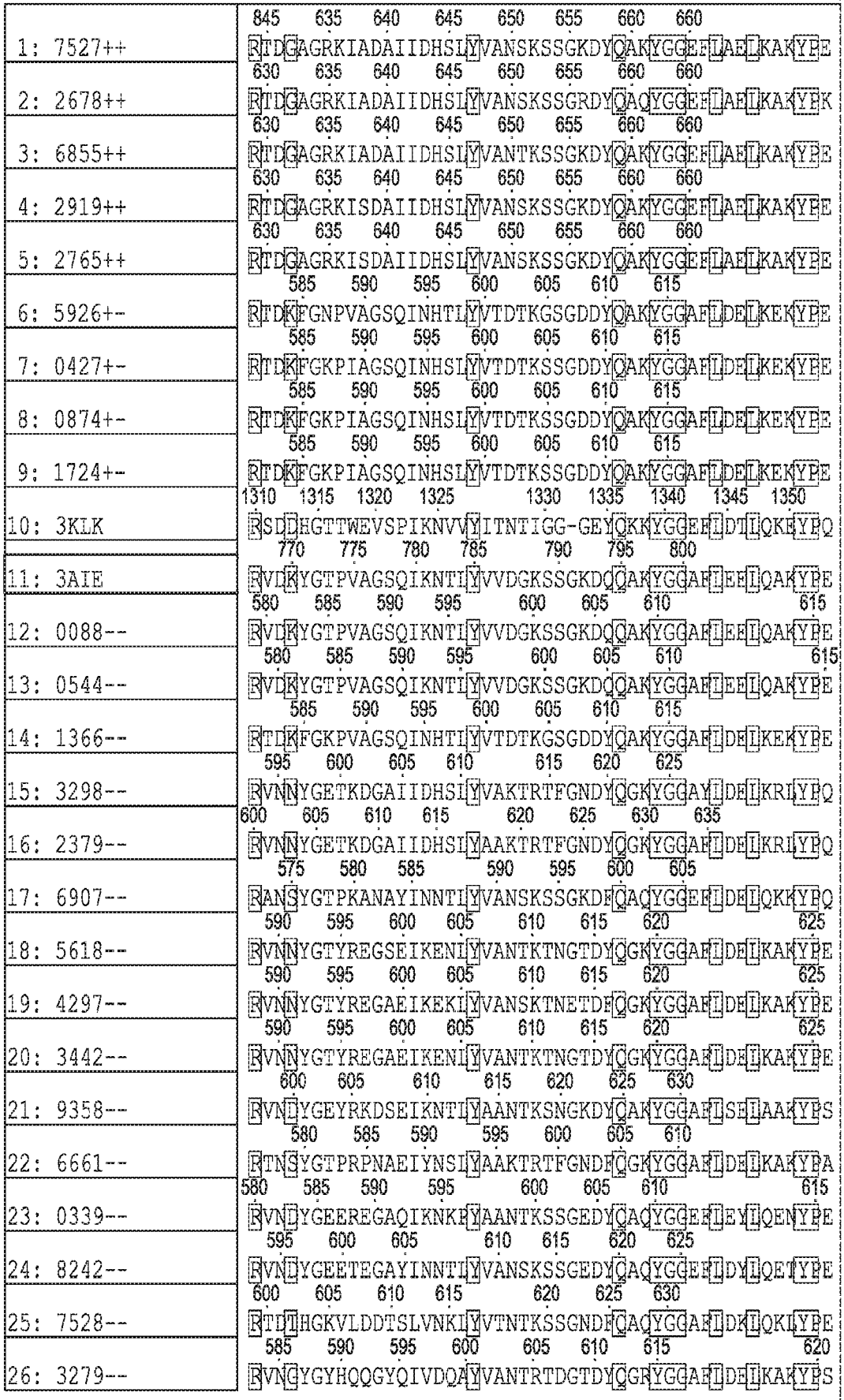

FIG. 2 Continued

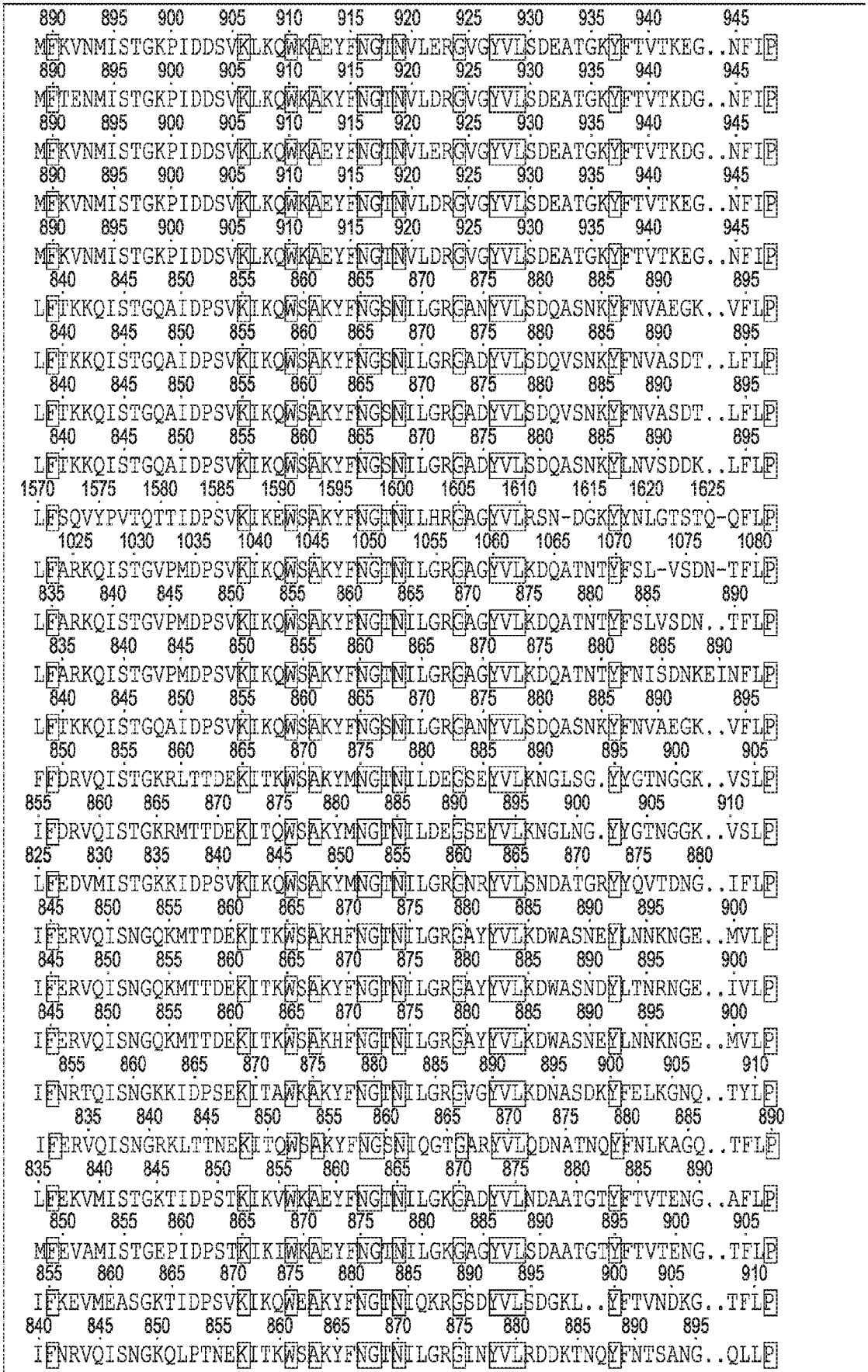

FIG. 2 Continued

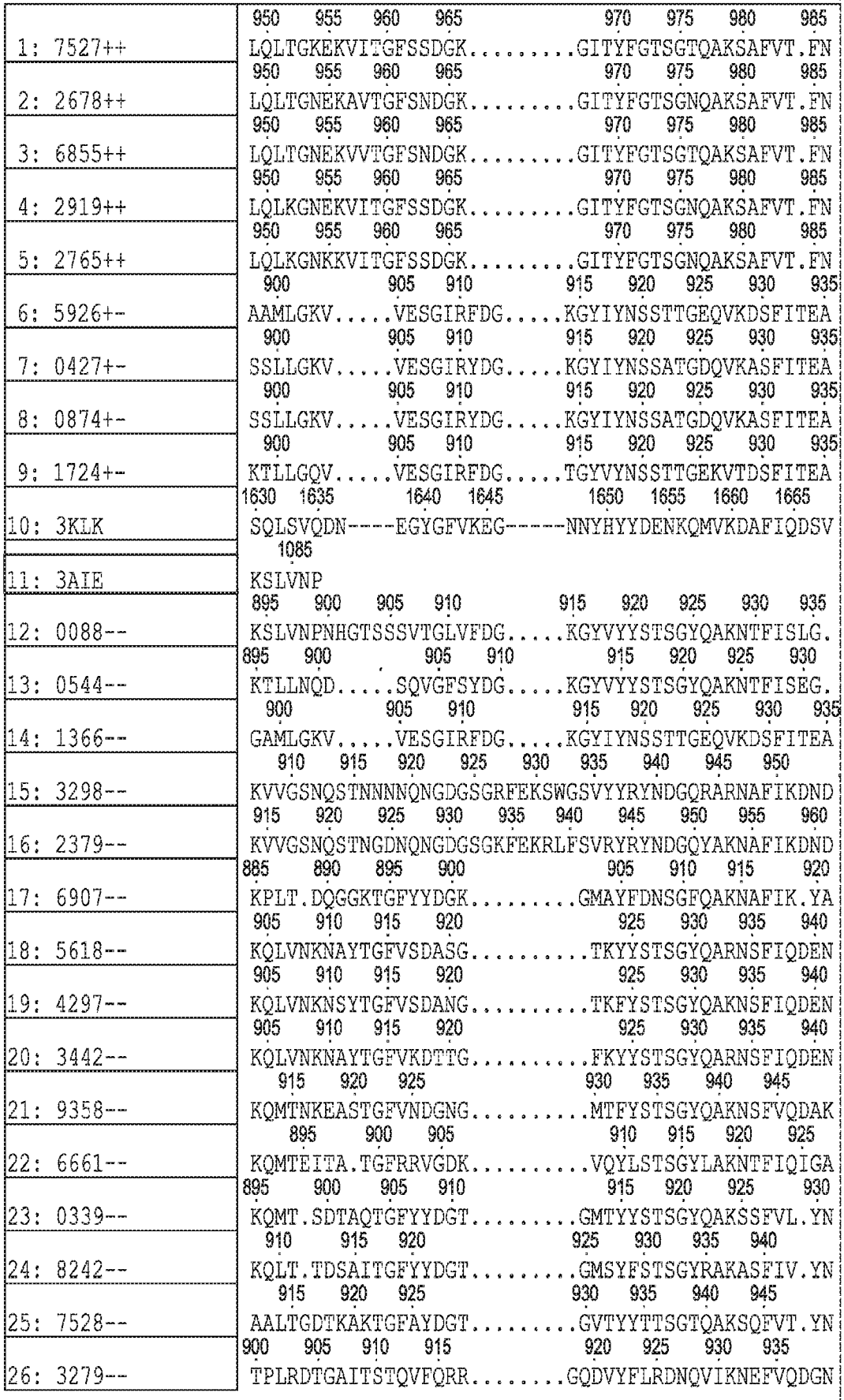

FIG. 2 Continued

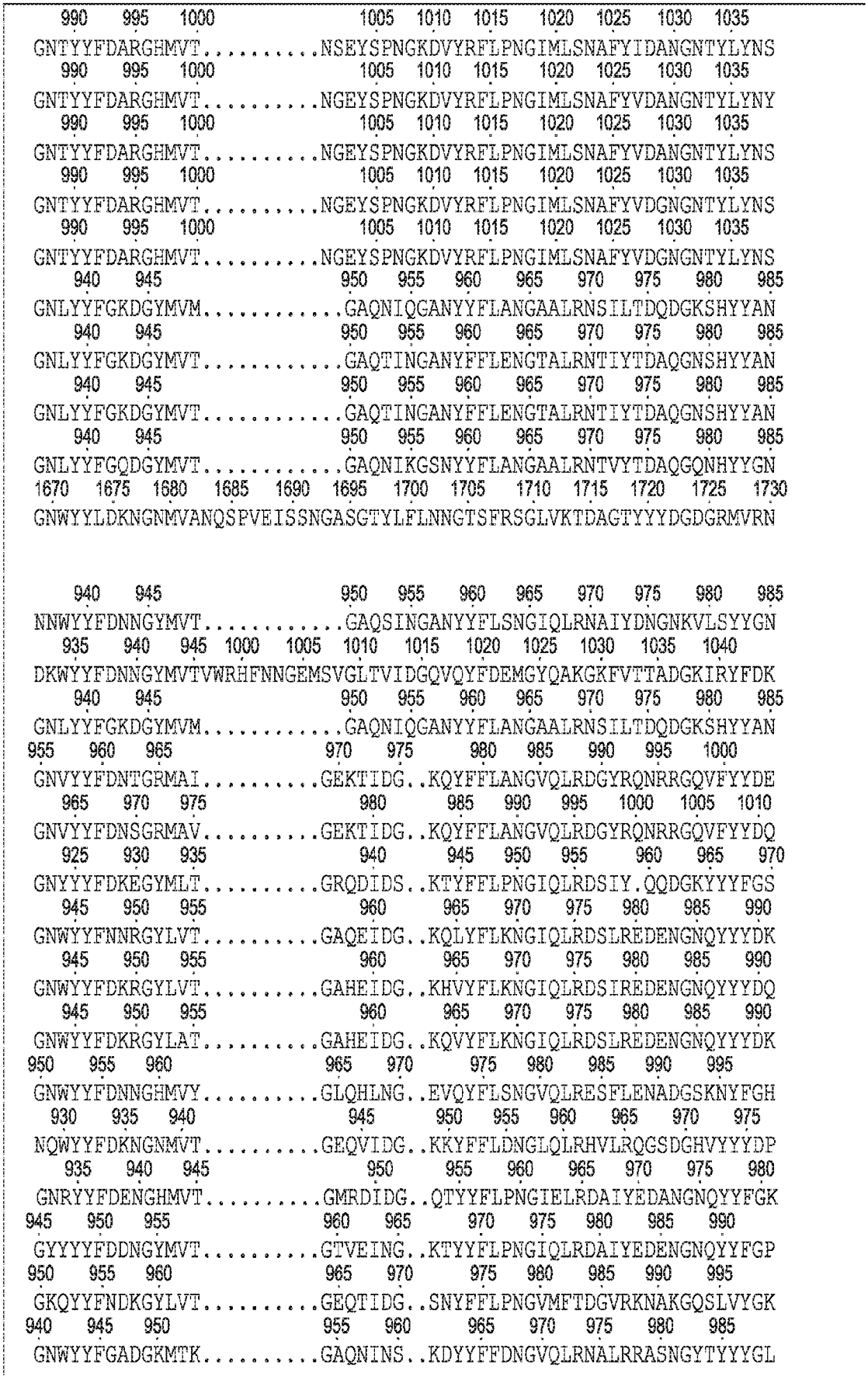

FIG. 2 Continued

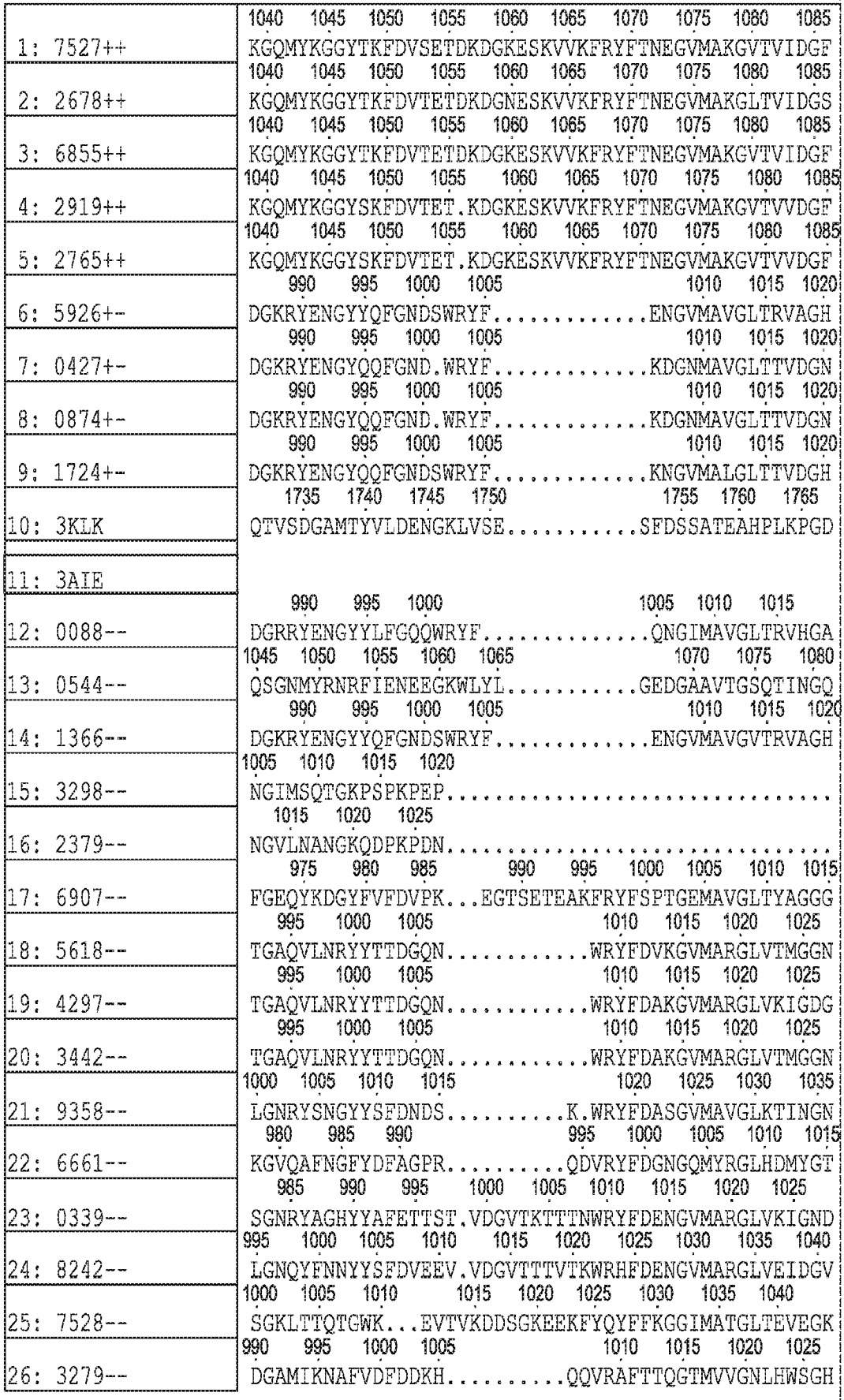

FIG. 2 Continued

```
   1090    1095    1100   1105  1110   1115   1120  1125  1130
TQYFGED.GFQAKD.KLVTFKGKTYYFDAHTGNGIKDTWRNINGKWYYF.............
  1090    1095    1100   1105  1110   1115   1120  1125  1130
TQYFGED.GFQTKD.KLATYKGKTYYFEAHTGNAIKNTWRNIDGKWYHF.............
  1090    1095    1100   1105  1110   1115   1120  1125  1130
TQYFGED.GFQAKD.KLVTFKGKTYYFDAHTGNAIKDTWRNINGKWYHF.............
   1090    1095    1100   1105  1110   1115   1120  1125  1130
TQYFNED.GIQSKD.ELVTYKGKTYYFEAHTGNAIKNTWRNIKGKWYHF.............
   1090    1095    1100   1105  1110   1115   1120  1125  1130
TQYFNED.GIQSKD.ELVTYKGKTYYFEAHTGNAIKNTWRNIKGKWYHF.............
   1025    1030  1035   1040  1045   1050   1055  1060  1065
DQYFDKD.GIQAKNKIIVTRDGKVRYFDEHNGNAATNTFISDQAGHWYY.............
1020   1025    1030   1035  1040   1045   1050  1055  1060  1065
VQYFDKD.GVQAKNKIIVTRDGKVRYFDEHNGNAVTNTFIADKTGHWYY.............
1020   1025    1030   1035  1040   1045   1050  1055  1060  1065
VQYFDKD.GVQAKNKIIVTRDGKVRYFDEHNGNAATNTFIADKTGHWYY.............
   1025    1030  1035   1040  1045   1050   1055  1060  1065
VQYFDKD.GVQAKNKIIVTRDGKVRYFDEHNGNAVTNTFVADKTGHWYY.............
 1770   1775
LNGQKHH.HHHH


1020   1025    1030   1035  1040   1045   1050  1055  1060  1065
VQYFDAS.GFQAKGQFITTADGKLRYFDRDSGNQISNRFVRNSKGEWFL.............
   1085    1090  1095   1100  1105   1110   1115  1120  1125
HLYFRAN.GVQVKGEFVTDRHGRISYYDGNSGDQIRNRFVRNAQGQWFY.............
   1025    1030  1035   1040  1045   1050   1055  1060  1065
DQYFDKD.GIQAKNKIIVTRDGKVRYFDEHNGNAVTNTFISDQAGHWYY.............
                                  1025   1030  1035  1040
..........................KPDNNTFSRNQFIQIGNNVWA.............
                                  1030   1035  1040  1045
.............................NNNASGRNQFVQIGNNVWA.............
   1020    1025  1030   1035  1040   1045   1050  1055  1060  1065  1070
LQYFDEN.GFQAKGTKYVTPDGKLYFFDKNSGNAYTNRWAEIDGIWYEFNDQGYAQA.....
  1030    1035  1040   1045  1050   1055   1060  1065  1070  1075
QQFFDQN.GYQVKGKIARAKDGKLRYFDKDSGNAAANRFAQGDNPSDWY.............
  1030    1035  1040   1045  1050   1055   1060  1065  1070  1075
QQFFDEN.GYQVKGKIVSAKDGKLRYFDKDSGNAVINRFAQGDNPSDWY.............
  1030    1035  1040   1045  1050   1055   1060  1065  1070  1075
QQFFDQN.GYQVKGKIARAKDGKLRYFDKDSGNAAANRFAQGDNPSDWY.............
   1040    1045  1050   1055  1060   1065   1070  1075   1080
TQYFDQD.GYQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKN.GDWY.............
   1020    1025  1030   1035  1040   1045   1050  1055  1060
TFYFDEKTGIQAKDKFIRFADGRTRYFIPDTGNLAVNRFAQNPENKAWY.............
 1030   1035    1040   1045  1050   1055   1060  1065   1070  1075
YQYYDDN.GNQIKGQLVTDKDGNTRYFKADSGAMVTGEFALV.NGGWYY.............
   1045    1050  1055   1060  1065   1070   1075  1080  1085
YQYYDEN.GYQVKGELITDADGNLRYFKEDSGEMVVSDFVKIGDNNWYY.............
1045   1050    1055   1060  1065   1070   1075  1080  1085  1090  1095  1100  1105
EKYFYDN.GYQAKGVFVPTKDGHLMFFCGDSGERKYSGFFEQDGNWYYANDKGYVATGFTKV
  1030    1035  1040   1045  1050   1055   1060  1065  1070  1075
HFYFDRETGIQAKDRIVRTDDGKLHYYVAQTGDMGRNVFATDSRTGKRY.............
```

FIG. 2 Continued

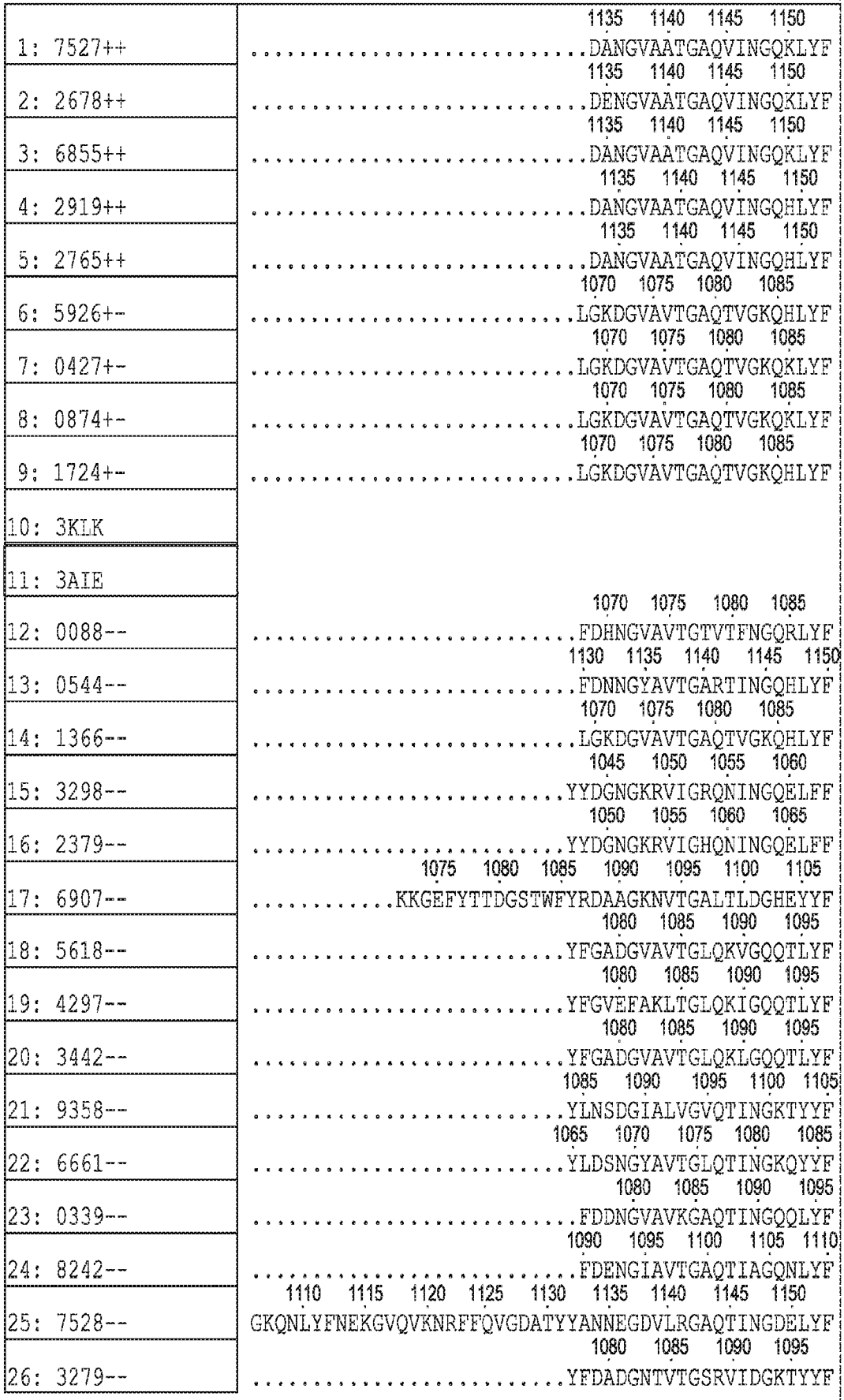

FIG. 2 Continued

```
1155        1160  1165                                          1170
NEDGS.....QVKGGVV.....KNADG
1155        1160  1165                                          1170
NEDGS.....QVKGGVV.....KNADG
1155        1160  1165                                          1170
NEDGS.....QVKGGVV.....KNADG
 1155        1160                                    1165
NEDGS.....QVKGGVV.....KNADG
 1155        1160                                    1165
NEDGS.....QVKGSIV.....KNADG
1090        1095  1100                                          1105
EANGQ.....QVKGDFV.....TAKD
1090        1095  1100
EANGE.....QVKGDFV.....TSHE
1090        1095  1100
EANGQ.....QVKGDFV.....TSDE
1090        1095  1100                                          1105
EANGQ.....QVKGDFV.....TAKD



1090        1095  1100
KPNGV.....QAKGEFI.....RDAD
   1155        1160                                           1165
RANGV.....QVKGEFV.....TDRH
1090        1095  1100                                          1105
EANGQ.....QVKGDFV.....TAKD
1065        1070  1075
DNNGV.....QVKGRTA.....QVDG
1070        1075  1080
DNNGV.....QVKGRTV.....NENG
 1110        1115  1120  1125  1130  1135  1140  1145  1150  1155  1160
RANGA.....QVKGEFVTENGKISYYTVDNGYKVKDKFFEVNGKWYHADKDGNLATGRQTID
 1100        1105                                          1110
DQDGK.....QVKGKVV.....TLAD
 1100        1105                                          1110
DQDGK.....QVKGKIV.....TLSD
 1100        1105                                          1110
DQEGK.....QVKGKIV.....TLAD
   1110        1115                                           1120
GQDGK.....QIKGKII.....TDNG
  1090        1095                                           1100
DNEGR.....QVKGHFV.....TINN
  1100        1105                                          1110
DENGV.....QAKGVFV.....TNEDG
   1115        1120                                           1125
DDNGV.....QAKGAFV.....TNADG
1155        1160  1165  1170  1175  1180  1185                1190
DESGK.....QVKGEFVNNPDGTTSYYDAITGVKLVD.....TSLVVDGQ
1100  1105  1110  1115
NQDGSVGTAYSNRADSI.....IFEN
```

FIG. 2 Continued

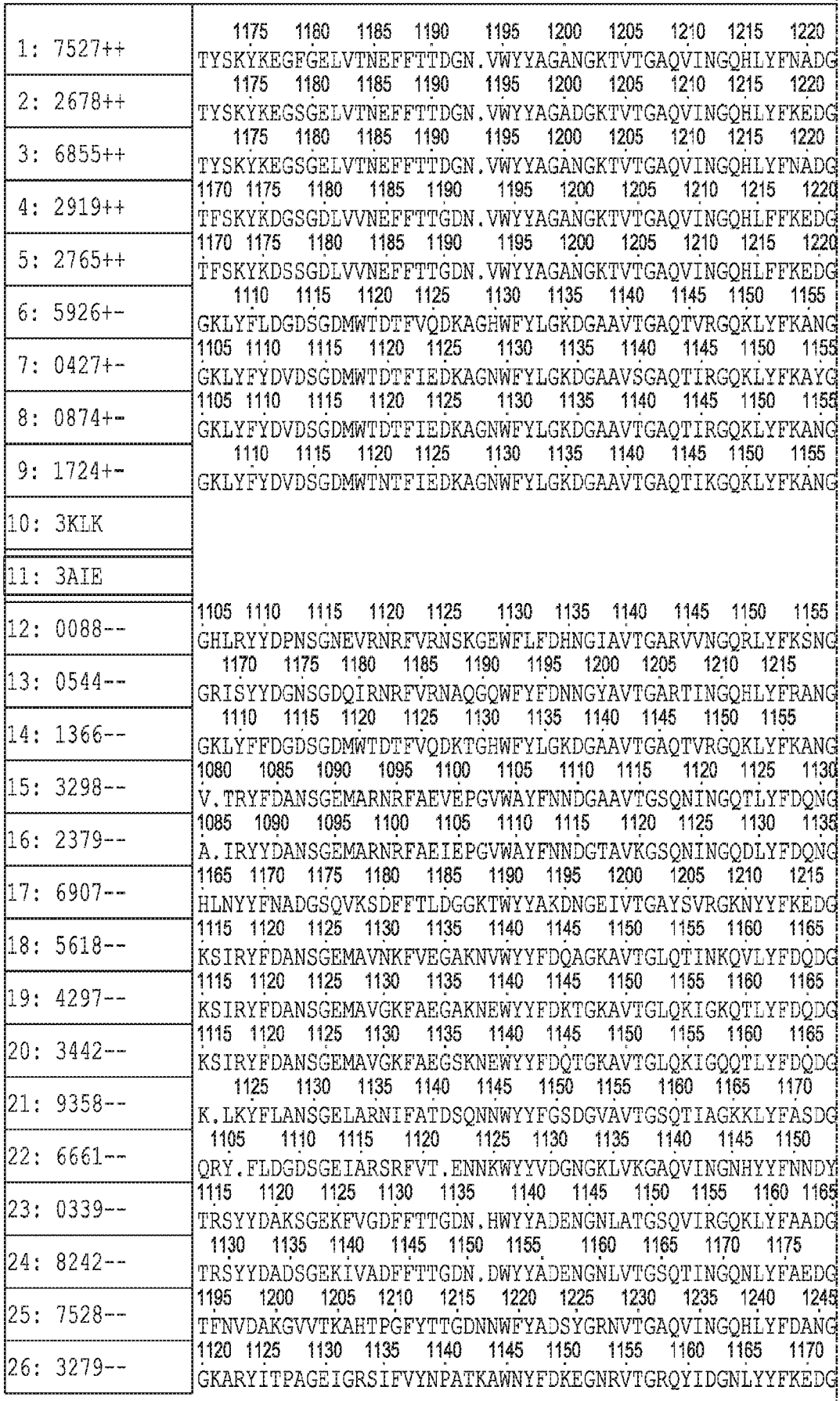

FIG. 2 Continued

FIG. 2 Continued

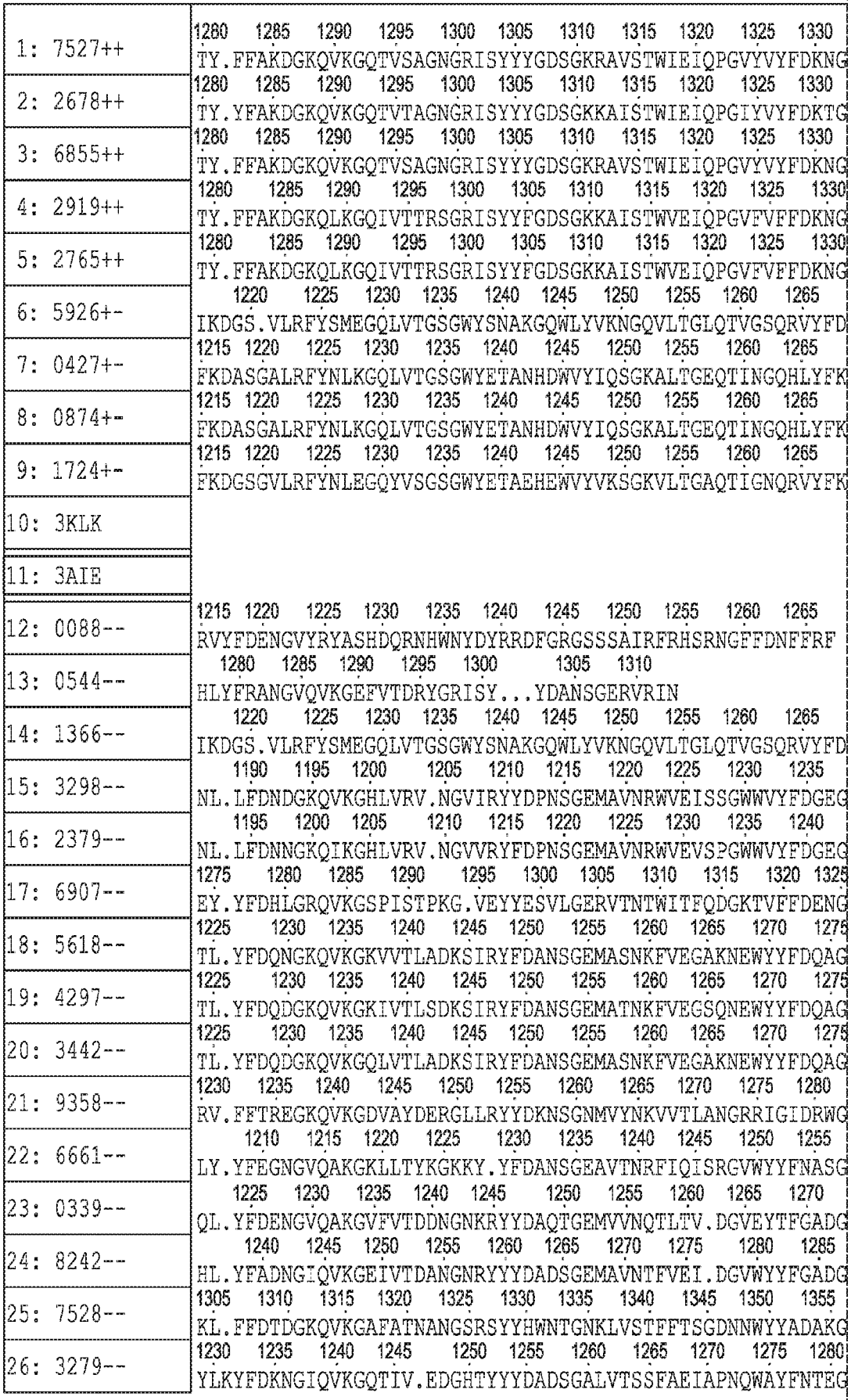

FIG. 2 Continued

```
 1335  1340
LAYPPRVLN
 1335  1340
IAYPPRVLN
 1335  1340
IAYPPRVLN
  1335  1340
LAYPPENMN
  1335  1340
LAYPPENMN
1270  1275  1280  1285  1290  1295  1300  1305  1310  1315  1320
ANGIQAKGKAVRTSDGKLRYFDANSGSMITNQWKEVNGQYYYFDNNGVAIYRGWN
1270  1275  1280  1285  1290  1295  1300  1305  1310  1315  1320
EDGHQVKGQLVTRTDGKVRYYDANSGDQAFNKSVTVNGKTYYFGNDGTAQTAGNPK
1270  1275  1280  1285  1290  1295  1300  1305  1310  1315  1320
EDGHQVKGQLVTGTDGKVRYYDANSGDQAFNKSVTVNGKTYYFGNDGTAQTAGNPK
1270  1275  1280  1285  1290  1295  1300  1305  1310  1315  1320
DNGHQVKGQLVTGNDGKLRYYDANSGDQAFNKSVTVNGKTYYFGSDGTAQTQANPK
```

```
1270  1275  1280  1285  1290  1295  1300  1305  1310  1315  1320
ANGIQAKGKAVRTSDGKLRYFDANSGSMITNQWKEVNGQYYYFDNNGVAIYRGWN
1240
RGQI
1245
RGQI
   1330
YADFDK
  1280  1285  1290  1295  1300  1305  1310  1315  1320  1325  1330
KAVTGLQQIGQQTLYFDQNGKQVKGKIVYVNGANRYFDANSGEMARNKWIQLEDGS
  1280  1285  1290  1295  1300  1305  1310  1315  1320  1325  1330
KAVTGLQQVGQQTLYFTQDGKQVKGKVVDVNGVSRYFDANSGDMARSKWIQLEDGS
  1280  1285  1290  1295  1300  1305  1310  1315  1320  1325  1330
KAVTGLQQIGQQTLYFDQNGKQVKGKIVYVNGANRYFDANSGEMARNKWIQLEDGS
 1285
IARYY
 1260  1265  1270  1275  1280  1285  1290  1295  1300  1305
QAVTGEQVINGQHLYFDASGRQVKGRYVWIKGQRRYYDANTGAWVRNR
1275  1280  1285  1290  1295  1300  1305  1310
VAVVNAQDSDEQSE..STDETQVTSDDATVAKTETSSAE
 1290  1295  1300  1305  1310  1315  1320  1325  1330  1335  1340
IAVTGAQVIDGQNLYFNADGSQVKGDVVRINGLRYYYDANSGEQVRNQWVTLPDGT
  1360  1365  1370  1375  1380  1385  1390  1395  1400  1405  1410
EVVVGEQTINGQHLYFDQTGKQVKGATATNPDGSISYYDVHTGEKAINRWVKIPSG
  1285  1290  1295  1300  1305  1310  1315  1320  1325  1330  1335
QALKGKWTINGKEYYFDQNGIQYKGKAVKVGSRYKYYDENDGQPVTNRFAQIEPNV
```

FIG. 2 Continued

1: 7527++	
2: 2678++	
3: 6855++	
4: 2919++	
5: 2765++	
6: 5926+-	
7: 0427+-	1325 1330 1335 1340 1345 1350 1355 1360 1365 1370 1375 GQIFKDGS.VLRFYSMEGQLVIGSGWYSNAQGQWLYVKNGKVLTGLQTVGSQRVYF
8: 0874+-	1325 1330 1335 1340 1345 1350 1355 1360 1365 1370 1375 GQTFKDGS.DIRFYSMEGQLVTGSGWYENAQGQWLYVKNGKVLTGLQTVGSQRVYF
9: 1724+-	1325 1330 1335 1340 1345 1350 1355 1360 1365 1370 1375 1380 GQTFKDGSGVLRFYNLEGQYVSGSGWYKNAQGQWLYVKDGKVLTGLQTVGNQKVYF
10: 3KLK	
11: 0088--	
12: 0544--	
13: 1366--	
14: 3298--	
15: 2379--	
16: 6907--	
17: 5618--	1335 1340 1345 WMYFDRNGRGRRFGWN
18: 4297--	1335 1340 1345 WMYFDRDGRGQNFGRN
19: 3442--	1335 1340 1345 WMYFDRNGRGRRFGWN
20: 9358--	
21: 6661--	
22: 0339--	
23: 8242--	1345 1350 1355 VVFFNARGYTWG
24: 7528--	1415 1420 1425 QWVYFNAQGKGYVSN
25: 3279--	1340 1345 1350 1355 1360 1365 1370 1375 1380 1385 1390 WAYFGADGYAVTGEQVINGQHLYFDQSGRQVKGAYVTVNGQRRYYDANTGEYIPGR

FIG. 2 Continued

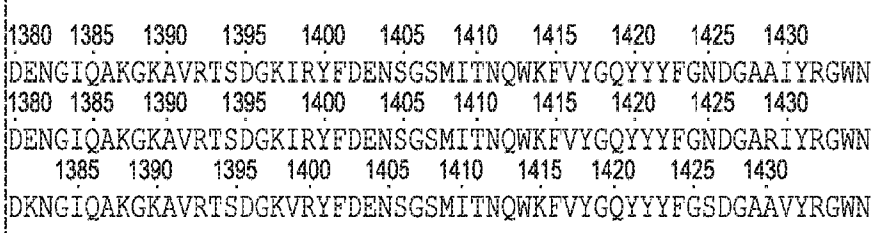

FIG. 2 Continued

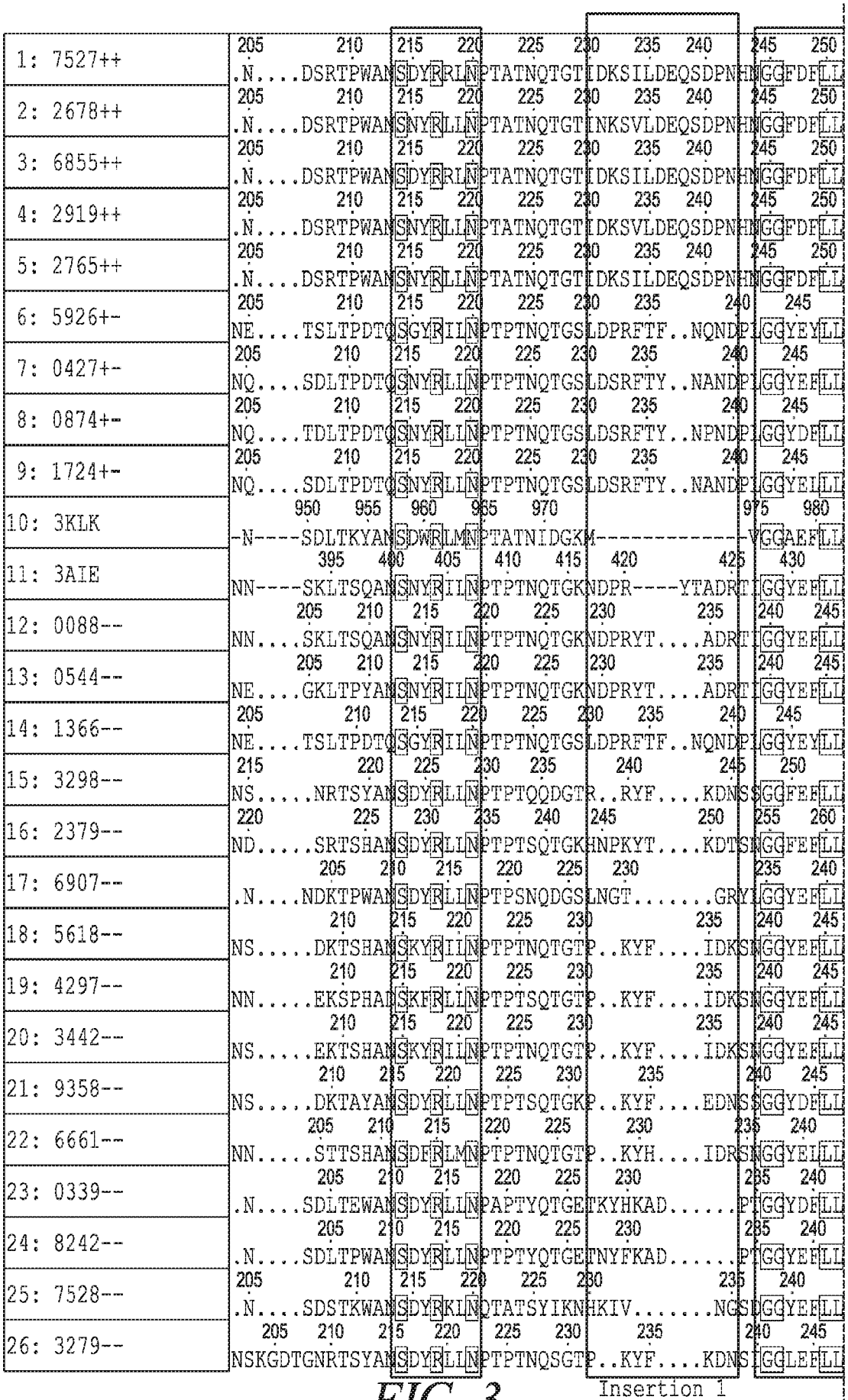
1: 7527++
2: 2678++
3: 6855++
4: 2919++
5: 2765++
6: 5926+-
7: 0427+-
8: 0874+-
9: 1724+-
10: 3KLK
11: 3AIE
12: 0088--
13: 0544--
14: 1366--
15: 3298--
16: 2379--
17: 6907--
18: 5618--
19: 4297--
20: 3442--
21: 9358--
22: 6661--
23: 0339--
24: 8242--
25: 7528--
26: 3279--
Insertion 1

FIG. 3

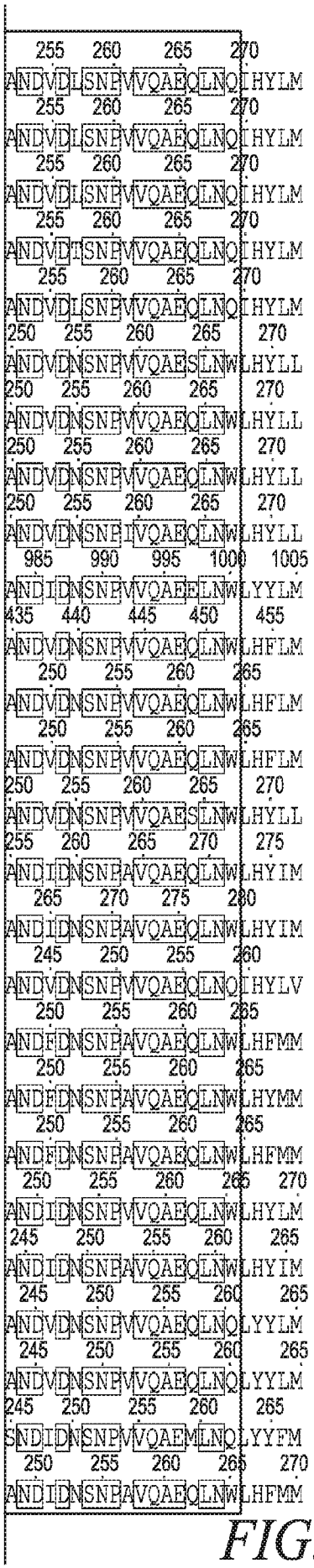

FIG. 3 Continued

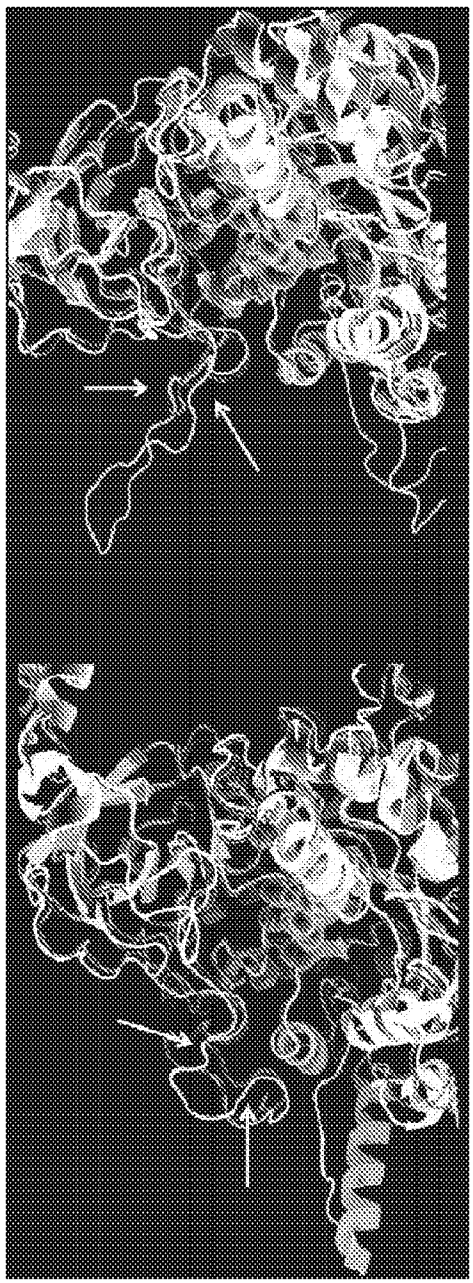

FIG. 4B

FIG. 4A

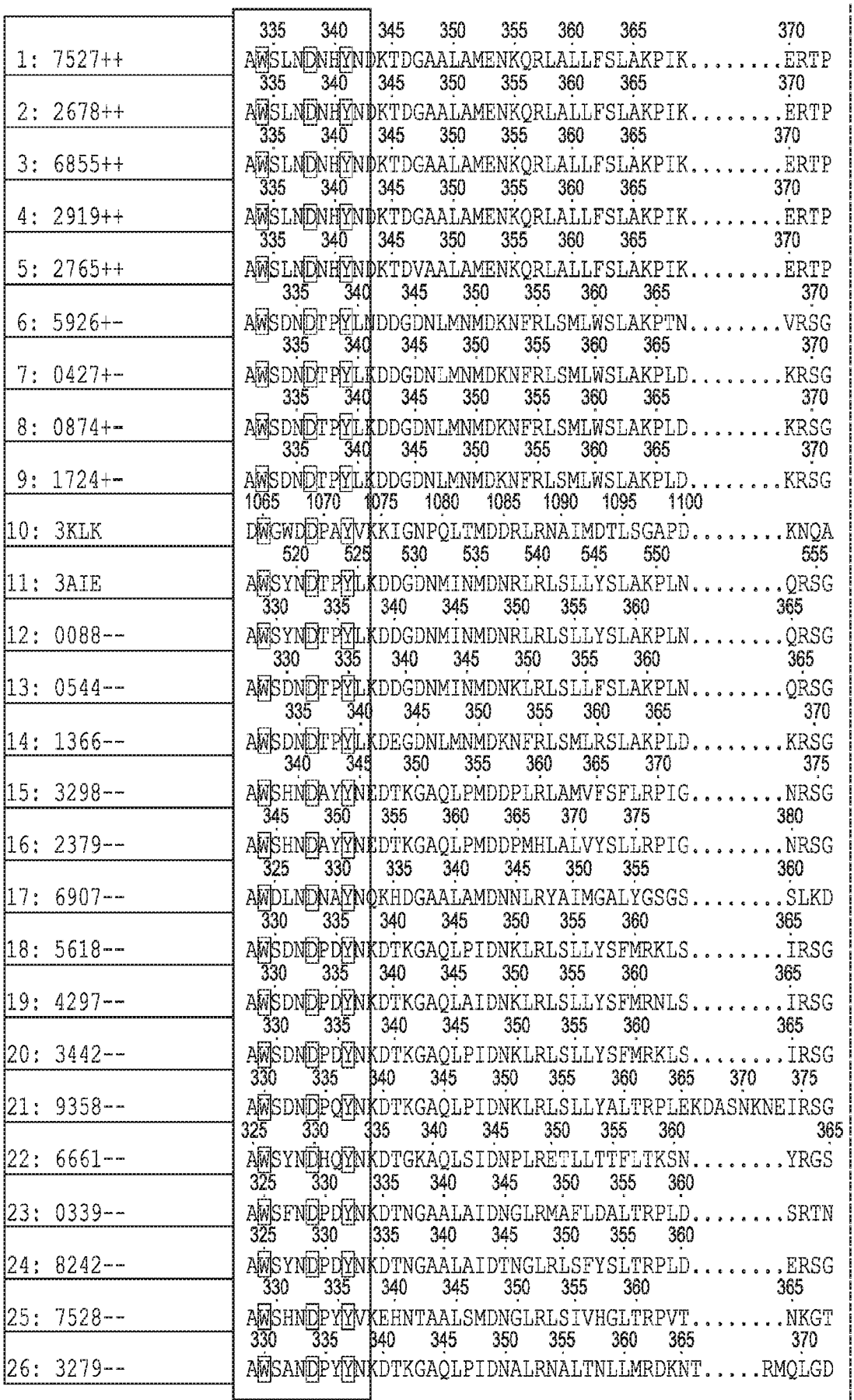
335 340 345 350 355 360 365 370
1: 7527++ AWSLNDNHYNDKTDGAALAMENKQRLALLFSLAKPIK.....ERTP
335 340 345 350 355 360 365 370
2: 2678++ AWSLNDNHYNDKTDGAALAMENKQRLALLFSLAKPIK.....ERTP
335 340 345 350 355 360 365 370
3: 6855++ AWSLNDNHYNDKTDGAALAMENKQRLALLFSLAKPIK.....ERTP
335 340 345 350 355 360 365 370
4: 2919++ AWSLNDNHYNDKTDGAALAMENKQRLALLFSLAKPIK.....ERTP
335 340 345 350 355 360 365 370
5: 2765++ AWSLNDNHYNDKTDVAALAMENKQRLALLFSLAKPIK.....ERTP
335 340 345 350 355 360 365 370
6: 5926+- AWSDNDTPYLNDDGDNLMNMDKNFRLSMLWSLAKPTN.....VRSG
335 340 345 350 355 360 365 370
7: 0427+- AWSDNDTPYLKDDGDNLMNMDKNFRLSMLWSLAKPLD.....KRSG
335 340 345 350 355 360 365 370
8: 0874+- AWSDNDTPYLKDDGDNLMNMDKNFRLSMLWSLAKPLD.....KRSG
335 340 345 350 355 360 365 370
9: 1724+- AWSDNDTPYLKDDGDNLMNMDKNFRLSMLWSLAKPLD.....KRSG
1065 1070 1075 1080 1085 1090 1095 1100
10: 3KLK DWGWDDPAYVKKIGNPQLTMDDRLRNAIMDTLSGAPD.....KNQA
520 525 530 535 540 545 550 555
11: 3AIE AWSYNDTPYLKDDGDNMINMDNRLRLSLLYSLAKPLN.....QRSG
330 335 340 345 350 355 360 365
12: 0088-- AWSYNDTPYLKDDGDNMINMDNRLRLSLLYSLAKPLN.....QRSG
330 335 340 345 350 355 360 365
13: 0544-- AWSDNDTPYLKDDGDNMINMDNKLRLSLLFSLAKPLN.....QRSG
335 340 345 350 355 360 365 370
14: 1366-- AWSDNDTPYLKDEGDNLMNMDKNFRLSMLRSLAKPLD.....KRSG
340 345 350 355 360 365 370 375
15: 3298-- AWSHNDAYYNEDTKGAQLPMDDPLRLAMVFSFLRPIG.....NRSG
345 350 355 360 365 370 375 380
16: 2379-- AWSHNDAYYNEDTKGAQLPMDDPMHLALVYSLLRPIG.....NRSG
325 330 335 340 345 350 355 360
17: 6907-- AWDLNDNAYNQKHDGAALAMDNNLRYAIMGALYGSGS.....SLKD
330 335 340 345 350 355 360 365
18: 5618-- AWSDNDPDYNKDTKGAQLPIDNKLRLSLLYSFMRKLS.....IRSG
330 335 340 345 350 355 360 365
19: 4297-- AWSDNDPDYNKDTKGAQLAIDNKLRLSLLYSFMRNLS.....IRSG
330 335 340 345 350 355 360 365
20: 3442-- AWSDNDPDYNKDTKGAQLPIDNKLRLSLLYSFMRKLS.....IRSG
330 335 340 345 350 355 360 365 370 375
21: 9358-- AWSDNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLEKDASNKNEIRSG
325 330 335 340 345 350 355 360 365
22: 6661-- AWSYNDHQYNKDTGKAQLSIDNPLRETLLTTFLTKSN.....YRGS
325 330 335 340 345 350 355 360
23: 0339-- AWSFNDPDYNKDTNGAALAIDNGLRMAFLDALTRPLD.....SRTN
325 330 335 340 345 350 355 360
24: 8242-- AWSYNDPDYNKDTNGAALAIDTNGLRLSFYSLTRPLD.....ERSG
330 335 340 345 350 355 360 365
25: 7528-- AWSHNDPYYVKEHNTAALSMDNGLRLSIVHGLTRPVT.....NKGT
330 335 340 345 350 355 360 365 370
26: 3279-- AWSANDPYYNKDTKGAQLPIDNALRNALTNLLMRDKNT.....RMQLGD

FIG. 5

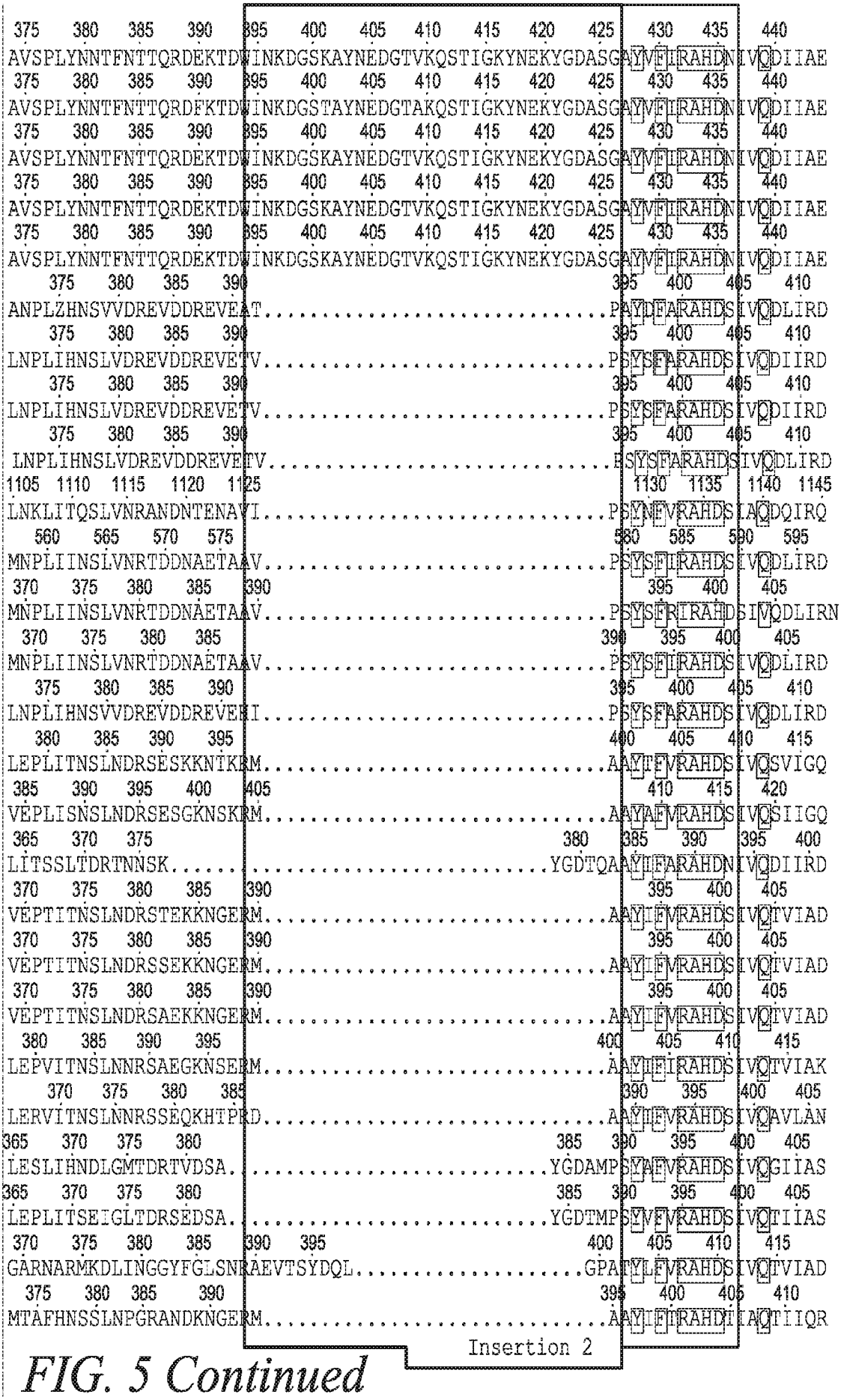

FIG. 5 Continued

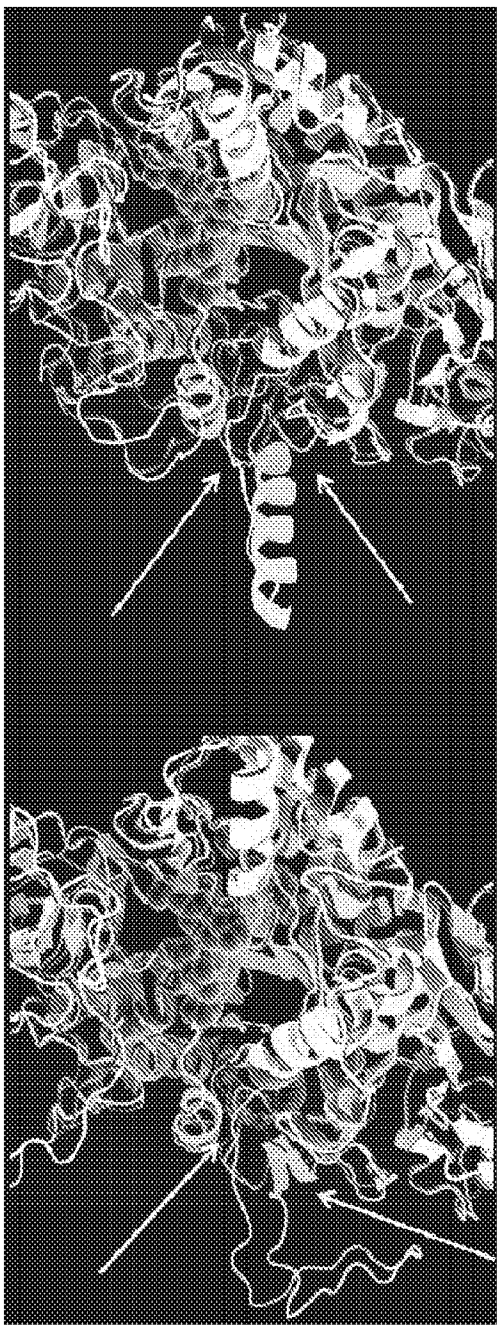

FIG. 6B

FIG. 6A

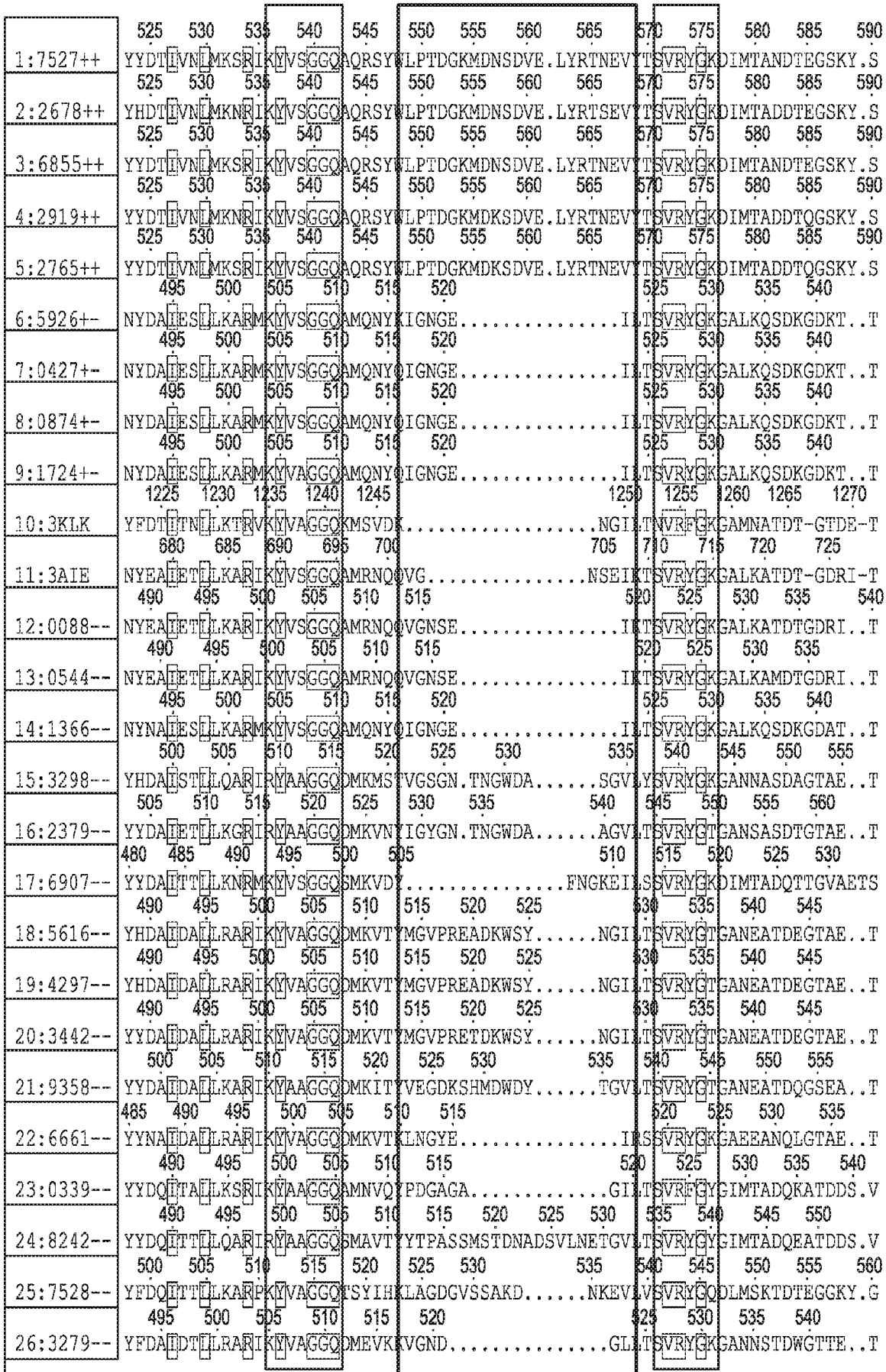

FIG. 7

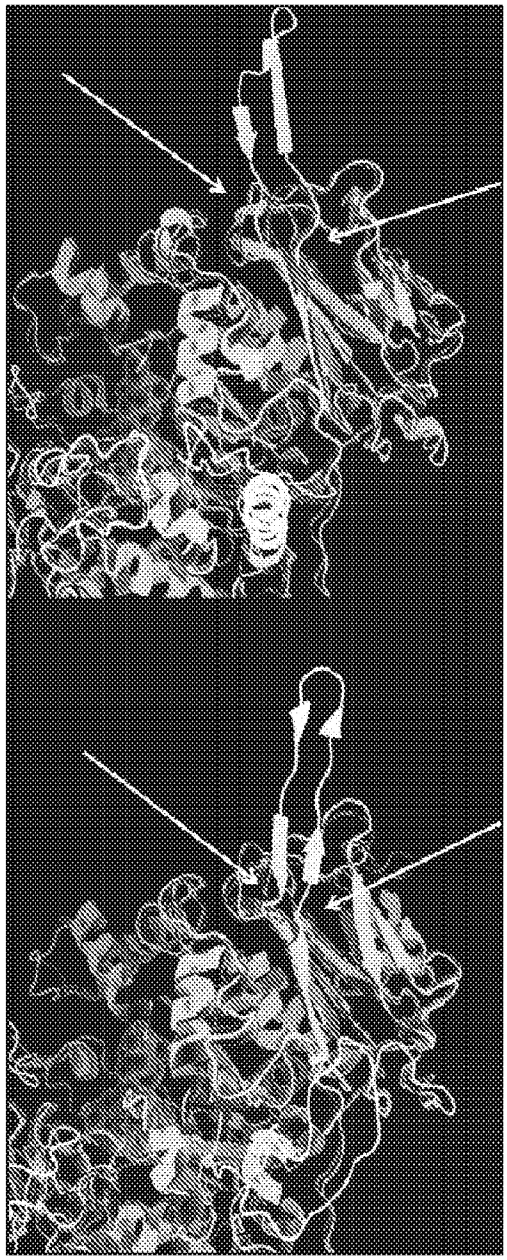

FIG. 8B

FIG. 8A