METHODS FOR THE PRODUCTION OF INSULIN IN PLANTS

Inventors: MAURICE M. MOLONEY, Calgary (CA); JOSEPH BOOTHIE, Calgary (CA); RICHARD KEON, Calgary (CA); CORY NYKIFORUK, Calgary (CA); GLIS VAN ROOIJEN, Calgary (CA)

Correspondence Address:
BERESKIN AND PARR LLP/S.E.N.C.R.L., s.r.l.
40 KING STREET WEST, BOX 401
TORONTO, ON M5H 3Y2 (CA)

Assignee: SEMBIOSYS GENETICS INC., Calgary (CA)

Appl. No.: 12/464,207
Filed: May 12, 2009

Commercial production of human insulin can be effected via transgenic expression in plant seeds. Thus, levels of insulin accumulation exceeding 0.1% of total cellular protein can be achieved recombinantly, through expression of the insulin with a single-chain antibody as a fusion partner. Production in seeds offers flexibility in storage and shipment of insulin as a raw material, and insulin retains its activity upon extraction from stored seed. Further, the amount of biomass subjected to extraction is limited, due to the relatively low water content of plant seeds.
### FIGURE 5

<table>
<thead>
<tr>
<th>Construct</th>
<th>Line</th>
<th>Transgene (% total seed protein)</th>
<th>Mini-insulin (% total seed protein)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4404</td>
<td>2</td>
<td>1.20</td>
<td>0.20</td>
</tr>
<tr>
<td>4404</td>
<td>17</td>
<td>1.09</td>
<td>0.19</td>
</tr>
<tr>
<td>4404</td>
<td>20</td>
<td>1.39</td>
<td>0.24</td>
</tr>
<tr>
<td>4405</td>
<td>4</td>
<td>0.63</td>
<td>0.11</td>
</tr>
<tr>
<td>4405</td>
<td>13</td>
<td>0.62</td>
<td>0.11</td>
</tr>
<tr>
<td>4405</td>
<td>19</td>
<td>0.75</td>
<td>0.13</td>
</tr>
<tr>
<td>4414</td>
<td>9</td>
<td>6.78</td>
<td>1.15</td>
</tr>
<tr>
<td>4414</td>
<td>20</td>
<td>2.50</td>
<td>0.42</td>
</tr>
<tr>
<td>Wt</td>
<td>C24</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
METHODS FOR THE PRODUCTION OF INSULIN IN PLANTS

This application is a continuation of U.S. patent application Ser. No. 10/869,040 filed Jun. 17, 2004 (now allowed), and claims the benefit under 35 USC §119(e) from U.S. Provisional patent application Ser. No. 60/478,818 filed Jun. 17, 2003 and Ser. No. 60/549,539 filed Mar. 4, 2004, which are incorporated herein by reference in their entirety.

FIELD OF THE INVENTION

The present invention relates to plant genetic engineering methods and to the production of insulin. More specifically, the present invention relates to methods for the production of insulin in the seeds of plants.

BACKGROUND OF THE INVENTION

Insulin is an important peptide hormone required to maintain blood glucose homeostasis in mammals, including humans, and other vertebrates. In healthy individuals, an increase in blood glucose level stimulates the β-cells of the pancreas to secrete insulin. The insulin polypeptide then binds to specific receptors in muscle, liver, and adipose tissue leading to an increase in glucose uptake by these targeted tissues, an increase in metabolism, and a decrease in hepatic glucose production. The cumulative effects of these responses serve to keep blood glucose concentrations at a constant level.

In individuals suffering from diabetes mellitus, an abnormally low insulin concentration presents itself as chronic hyperglycemia. The clinical manifestations of chronic hyperglycemia are manifold, and include blindness, kidney failure and, if left untreated, will ultimately result in death. Estimates place diabetes mellitus as the third largest cause of death in industrialized countries, after cardiovascular diseases and cancer (Barfoed, H. C., 1987, Chem. Eng. Prog. 83:49-54). In order to allow efficient uptake and metabolism of blood glucose by the cells, diabetic individuals may be treated by the routine administration of insulin. Approximately 0.7% of the world’s population suffers from insulin-dependent diabetes (diabetes mellitus Type 1) (Winter, J. et al., 2000, J. of Biotechnol. 84:175-185). In addition, it is estimated that the number of individuals diagnosed with diabetes will double to approximately 300 million, in the next 25 years (Kjeldsen, I. et al., 2001, Biotechnol. Gen. Eng. Rev. 18:89-121). Consequently, the ability to cost-effectively manufacture human insulin in quantities to satisfy the anticipated growing world demand for insulin is highly desirable.

In vivo the human insulin polypeptide is produced by the pancreatic β-cells as a single 110 amino acid polypeptide chain precursor, proinsulin, which includes an N-terminally located 24 amino acid pre-sequence that is cleaved immediately upon completion of the chain’s biosynthesis (Steiner, D. F. 2000. J. Ped. Endocrinol. Metab. 13:229-239). Proinsulin consists of a B and A chain, linked by a connecting peptide (C-peptide). During packaging of the hormone for secretion the C-peptide is cleaved and removed by prohormone convertases, PC2 and PC1/PC3 (Steiner, D. F. 2000. J. Ped. Endocrinol. Metab. 13:229-239). What remains is mature human insulin, a 51 amino acid protein consisting of two polypeptide chains, A (21 amino acids in length) and B (30 amino acids in length), linked by two inter-chain disulphide bonds. Additionally, the A chain comprises one intrachain disulphide bond.

Human insulin has been prepared using a variety of different methodologies. Microorganisms such as Escherichia coli (Frank et al., 1981, in Peptides: Proceedings of the 7th American Peptide Chemistry Symposium (Rich & Gross, eds.), Pierce Chemical Co., Rockford, Ill. pp 729-739; Chan et al., 1981, Proc Natl. Acad. Sci. USA 78: 5434-5438), Saccharomyces cerevisiae (Thin et al., 1986, Proc Natl. Acad. Sci. USA 83: 6766-6770) are routinely employed to recombinantly produce insulin. Wang et al. (Biotechnol. Bioeng. 2001, 73:74-79) have shown that fungi, such as Pichia pastoris, are also suitable for insulin production. Alternative manufacturing options include production in non-human mammalian cell lines (Yanagita, M., et al., 1992, FEBS Lett 311:55-59), isolation from human pancreas, peptide synthesis, or the semisynthetic conversion to human insulin from porcine and bovine insulin. However, all of these methods suffer from lower yields and higher costs than desired.

The use of plants as bioreactors for the large scale production of recombinant proteins is well known, and numerous studies, including human therapeutic proteins, have been produced. For example, U.S. Pat. Nos. 4,596,282, 5,550,038 and 5,629,175 disclose the production of γ-interferon in plants; U.S. Pat. Nos. 5,650,307, 5,716,802 and 5,763,748 detail the production of human serum albumin in plants and U.S. Pat. Nos. 5,202,422, 5,639,947 and 5,955,177 relate to the production of antibodies in plants. One of the significant advantages offered by plant-based recombinant protein production systems is that by increasing the acreage of plants grown, protein production can be inexpensively scaled up to provide for large quantities of protein. By contrast, fermentation and cell culture systems have large space, equipment and energy requirements, rendering scale-up of production costly. However, despite the fact that the use of plants as bioreactors is amply documented, and despite the above mentioned anticipated prodigious increase in need for large volumes of insulin, the prior art provides only a limited number of methods which demonstrably result in the production of insulin in plants (see: Arakawa et al. Nature Biotech., 1998, 16: 934-938; PCT 0172959).

Arakawa et al. disclose the production of a fusion protein comprising insulin in the tubers of transgenic potato plants. However insulin represents only up to 0.05% of the total soluble protein content present in the transgenic tubers. At a level of 0.05% of total soluble protein, large amounts of biomass must be subjected to protein extraction rendering the production economics associated with the use of potato tubers unfavorable. Furthermore, Arakawa et al. are not concerned with the isolation of insulin from the potato tuber tissue, but instead suggest an approach prevent the onset of Type I diabetes by inducing immunotolerance which involves oral administration of insulin through the feeding of transgenic potato tubers.

PCT Patent Application WO 01/72959 discloses the production of a fusion protein comprising insulin in chloroplasts of transgenic tobacco. However, while purportedly addressing shortcomings with respect to the accumulation levels of human proteins in plant tissue, the invention to which WO 0172959 pertains is limited in that production in chloroplasts results in the accumulation of insulin in green tissue, primarily the tobacco leaves. Due to the relatively high water content of green tissue, a large amount of biomass must
be processed. Furthermore production of insulin would require immediate extraction from the biomass upon harvesting, as leaf material will rapidly deteriorate when stored.

[0010] Thus in view of the shortcomings associated with the methods for the recombinant production of insulin in plants provided by the prior art, it is presently unclear whether and how the synthetic capacity of plants may be harnessed to achieve the commercial production of insulin in plants. There is a need in the art to improve methods for the commercial production of insulin in plants.

SUMMARY OF THE INVENTION

[0011] The present invention relates to improved methods for the production of insulin in plants. In particular the present invention relates to methods for the production of insulin in seeds.

[0012] Accordingly, the present invention provides a method for the expression of insulin in plants comprising:

[0013] (a) providing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:

[0014] (i) a nucleic acid sequence capable of controlling expression in plant seed cells; and

[0015] (ii) a nucleic acid sequence encoding an insulin polypeptide;

[0016] (b) introducing the chimeric nucleic acid construct into a plant cell; and

[0017] (c) growing the plant cell into a mature plant capable of setting seed wherein the seed expresses insulin.

[0018] In a preferred embodiment of the present invention the nucleic sequence capable of controlling expression in plant seed cells is a seed-preferred promoter, such as a phaseolin promoter.

[0019] In a preferred embodiment of the present invention insulin is expressed in a manner that permits accumulation of the insulin polypeptide in a membrane enclosed intracellular compartment within the seed cells. Accordingly the present invention provides a method for the expression of insulin in plants which comprises:

[0020] (a) providing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:

[0021] (i) a nucleic acid sequence capable of controlling expression in plant seed cells; and

[0022] (ii) a nucleic acid sequence encoding an insulin polypeptide; and

[0023] (iii) a nucleic acid sequence encoding a polypeptide capable of retaining the insulin polypeptide in a membrane enclosed intracellular compartment

[0024] (b) introducing the chimeric nucleic acid construct into a plant cell; and

[0025] (c) growing the plant cell into a mature plant capable of setting seed wherein the seed expresses insulin.

[0026] In a further preferred embodiment of the present invention, the membrane enclosed intracellular compartment is the endoplasmic reticulum (ER) or an ER-derived storage vesicle. Accordingly, the present invention provides a method for the expression of insulin in plants which comprises:

[0027] (a) providing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:

[0028] (i) a nucleic acid sequence capable of controlling expression in plant seed cells; and

[0029] (ii) a nucleic acid sequence encoding an insulin polypeptide;

[0030] (iii) a nucleic acid sequence encoding a polypeptide capable of retaining the insulin polypeptide in the ER or in an ER derived storage vesicle

[0031] (b) introducing the chimeric nucleic acid construct into a plant cell; and

[0032] (c) growing the plant cell into a mature plant capable of setting seed wherein the seed expresses insulin.

[0033] In a further preferred embodiment the chimeric nucleic acid construct is introduced into the plant cell under nuclear genomic integration conditions. Under such conditions the chimeric nucleic acid sequence is stably integrated in the plant’s genome.

[0034] In a yet further preferred embodiment the nucleic acid sequence encoding insulin is optimized for plant codon usage and the nucleic acid sequence encoding the connecting peptide (C-peptide) is shortened. Preferred nucleic acid sequences used in accordance with the present invention encode human, bovine or porcine insulin. In accordance with the present invention a nucleic acid sequence encoding a proinsulin sequence is used wherein the proinsulin is modified in that the C-peptide is shortened in length.

[0035] In another aspect, the present invention provides a method of recovering plant seeds comprising insulin. Accordingly, pursuant to the present invention a method is provided for obtaining plant seeds comprising insulin comprising:

[0036] (a) providing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:

[0037] (i) a nucleic acid sequence capable of controlling expression in plant seed cells; and

[0038] (ii) a nucleic acid sequence encoding an insulin polypeptide;

[0039] (b) introducing the chimeric nucleic acid construct into a plant cell;

[0040] (c) growing the plant cell into a mature plant capable of setting seed; and

[0041] (d) obtaining seeds from said plant wherein the seed comprises insulin.

[0042] Preferably, at least 0.1% of the total seed protein present in the seed is insulin.

[0043] The seeds may be used to obtain a population of progeny plants each comprising a plurality of seeds expressing insulin.

[0044] The present invention also provides plants capable of setting seed expressing insulin. In a preferred embodiment of the invention, the plants capable of setting seed comprise a chimeric nucleic acid sequence comprising in the 5' to 3' direction of transcription:

[0045] (a) a first nucleic acid sequence capable of controlling expression in a plant seed cell operatively linked to:

[0046] (b) a second nucleic acid sequence encoding an insulin polypeptide, wherein the seed contains insulin.

[0047] Preferably, at least 0.1% of the total seed protein present in the seed is insulin.

[0048] In a preferred embodiment the chimeric nucleic acid sequence is integrated in the plant’s nuclear genome.
[0049] In a further preferred embodiment of the present invention the plant that is used is a sunflower, a flax plant or an Arabidopsis plant.

[0050] In yet another aspect, the present invention provides plant seeds expressing insulin. In a preferred embodiment of the present invention, the plant seeds comprise a chimeric nucleic acid sequence comprising in the 5' to 3' direction of transcription:

[0051] (a) a first nucleic acid sequence capable of controlling expression in a plant seed cell operatively linked to;

[0052] (b) a second nucleic acid sequence encoding an insulin polypeptide.

[0053] Preferably, at least 0.1% of the total seed protein present in the seed is insulin. The seeds are a source whence the desired insulin polypeptide, which is synthesized by the seed cells, may be extracted and the insulin may be used to treat diabetic patients.

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

BRIEF DESCRIPTION OF THE DRAWINGS

[0055] The invention will now be described in relation to the drawings in which:

[0056] FIG. 1 depicts the nucleotide sequence (SEQ ID NO. 1) and deduced amino acid sequence (SEQ ID NO. 2) of the insulin fusion protein (PR5-D9scFv-klip27-MI-KDEL) of pSH84404. The predicted amino acid sequence is shown in single letter code. The deduced amino acid sequence of the PRS signal peptide is in italics, the deduced amino acid sequence of the D9 scFv is in bold, the deduced amino acid sequence of the KLIP27 sequence is underlined, the deduced amino acid sequence of the mini-insulin sequence is in italics and bold and finally the KDEL sequence is bolded and underlined.

[0057] FIG. 2 depicts the nucleotide sequence (SEQ ID NO. 3) and deduced amino acid sequence (SEQ ID NO. 4) of the insulin fusion protein (OEO-KLIP8-KLIP27-MI) of pSH84405. The predicted amino acid sequences are shown in single letter code. The deduced amino acid sequence of the Arabidopsis thaliana 18 kDa oleosin is in italics, the deduced amino acid sequence of the KLIP 8 sequence is in bold, the deduced amino acid sequence of the KLIP27 sequence is underlined and the deduced amino acid sequence of the mini-insulin is in italics and bold.

[0058] FIG. 3 depicts the complete nucleic acid sequence (SEQ ID NO:5) and amino acid sequence (SEQ ID NO:6) of the 4414 insulin fusion protein (PR5-MI-tetraspecific linker-D9scFv-KDEL). The predicted amino acid sequences are shown in single letter code. The deduced amino acid sequence of the PRS signal peptide is in italics, the deduced amino acid sequence of the mini-insulin (B30 tetraspecific) is in bold, the deduced amino acid sequence of the tetrabasic linker sequence is underlined, the deduced amino acid sequence of the D9 scFv is in italics and bold and finally the KDEL sequence is bolded and underlined.

[0059] FIGS. 4(A-D) depicts recombinant expression of insulin fusion proteins in transformed Arabidopsis thaliana lines (4404-2, -17, -20, and 4405-4) on the basis of Coomassie-stained SDS-PAGE and Western blot analysis. The arrows denote the position of the migrating 38.5 kDa and 34.2 kDa fusion polypeptides, PRS-D9(scFv)-KLIP27-MI/KDEL and Oleo-KLIP8-KLIP27-MI respectively, under reducing conditions. FIGS. 4A (Coomassie-stained gel) and 4B (corresponding Western Blot probed with anti-insulin E2E3) depicts total seed protein from wild type (wt) and transgenic seed lines expressing the 4404 and 4405 constructs. FIGS. 4C (Coomassie-stained gel) and 4D (corresponding Western Blot probed with anti-insulin E2E3) depicts oil body protein prepared from wild type and transgenic seed expressing the same 4404 and 4405 constructs. FIGS. 4E and 4F depicts the recombinant expression of insulin fusion proteins in transformed Arabidopsis thaliana lines (4419-9 and 4414-20) on the basis of Coomassie-stained SDS-PAGE (4E) and Western Blot analysis (4F). The molecular weight markers (M) are 10, 15, 25, 37, 50, 75, 100, 150 kDa. Controls include, hIN (recombinant human insulin standard) and hProIN (recombinant human proinsulin standard), separated under non-reducing conditions.

[0060] FIG. 5 depicts determined expression levels in the available T3 seed lines (4404-2, -17, -20, 4405-4, -13, -19) and 2 seed lines (4414-9 and -20). The levels of transgene and % molar MI expression were determined on the basis of densitometry.

[0061] FIG. 6 depicts Coomassie-stained SDS-PAGE (15%) analysis of oil body preps prior to elution (—OB), OB prep after elution with formic acid (-OH), and the concentrated eluted material (-E). The arrow denotes the position of the migrating fusion polypeptide. The wild type control is essentially free of any major proteins following elution whereas the concentrated 4404 material contains the fusion protein, some truncated products (possible hydrolyzed fusion protein) and possibly some albumins that co-eluted.

[0062] FIG. 7 depicts the chromatograms depicting the characteristic retention times of the human insulin standard (retention time 17.179 min) in comparison to the trypsin cleaved eluted 4404 fusion protein (retention time 17.011 min) on the C18 column. The hIN standard is recombinant human insulin standard (0.5 µg).

[0063] FIG. 8 depicts the mass spectral analysis of human insulin standard (A) in comparison to trypsin cleaved and HPLC purified 4404 (B) fractions collected from 17.0-17.5 minutes.

[0064] FIG. 9 depicts the Coomassie-stained SDS-PAGE (15%) analysis of total extractable seed protein and oil body (OB) prepared protein from lines expressing 4405 in comparison to wild type (nonrecombinant) seed. The arrow denotes the position of the migrating fusion polypeptide.

[0065] FIG. 10 depicts the chromatograms depicting the characteristic retention times of the human insulin standard (retention time 17.179 min) in comparison to the trypsin cleaved 4405 OB preparations (retention time 17.220 min) by RP-HPLC on the C18 column. The hIN standard is recombinant human insulin standard (0.5 µg).

[0066] FIG. 11 depicts the mass spectral analysis of human insulin standard (A) in comparison to trypsin cleaved and HPLC purified 4405 (B) fractions collected from 17.0-17.5 minutes.

[0067] FIG. 12 depicts the chromatogram of trypsin cleaved 4405 oil body preparations (dashed line) in compari-
son to human insulin standard (solid line). Fractions of the eluted cleaved insulin was collected between 7-35 mScm and concentrated by hyrophilization for insulin bioassay.

[0068] FIG. 13 depicts the changes in serum glucose levels in male B6 mice following injection of negative controls (open circles—saline placebo, closed circles—trypsin cleaved wild-type oil bodies), positive controls (open squares—human R11, open triangles—human hN) in comparison to plant derived insulin prepared from 4405 oil bodies (closed diamonds—SBS hN DesB10).

[0069] FIG. 14 depicts a Coomassie-stained gel of oil body proteins from two representative lines (4409-6 and 4409-8) comparing the migration of Oleosin-1PIN fusion protein (as denoted by the black arrow) to non-transformed (wt). Arabidopsis. The level of expression was determined by densitometry to measure on average about 0.10% of total seed protein. This level was calculated above and beyond the co-migration of an endogenous protein of the same molecular weight in the non-transformed seed (wt) which constituted approximately 0.04% of the total seed protein.

DETAILED DESCRIPTION OF THE INVENTION

[0070] As hereinbefore mentioned, the present invention relates to improved methods for the production of insulin in transgenic plants. The present inventors have surprisingly found that levels of insulin accumulation exceeding 0.1% of total cellular protein may be achieved in plants by recombinantly producing insulin in the seeds of plants. These expression levels, which are at least ten times higher as those here-tofore achieved, render commercial production of insulin in plants viable. Production in seeds offers flexibility in storage and shipment of insulin as a raw material, since insulin retains its activity upon extraction from stored seed. Furthermore, the amount of biomass that needs to be subjected to extraction is limited due to the relatively low water content present in plant seeds.

[0071] Accordingly, pursuant to the present invention a method for the expression insulin in plants is provided in which the method comprises:

[0072] (a) providing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:

[0073] (i) a nucleic acid sequence capable of controlling expression in plant seed cells; and

[0074] (ii) a nucleic acid sequence encoding an insulin polypeptide;

[0075] (b) introducing the chimeric nucleic acid construct into a plant cell; and

[0076] (c) growing the plant cell into a mature plant capable of setting seed wherein the seed expresses insulin.

[0077] In accordance with the present invention it has surprisingly been found that insulin accumulates to levels in plant seeds heretofore not achieved if insulin is expressed in seed in a manner that permits sequestration of the insulin polypeptide within the seed cells in a membrane enclosed intracellular compartment. Accordingly, pursuant to the present invention a preferred method for the expression of insulin in plants is provided in which the method comprises:

[0078] (a) providing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:

[0079] (i) a nucleic acid sequence capable of controlling expression in plant seed cells; and

[0080] (ii) a nucleic acid sequence encoding an insulin polypeptide;

[0081] (iii) a nucleic acid sequence encoding a polypeptide capable of retaining the insulin polypeptide in a membrane enclosed intracellular compartment

[0082] (b) introducing the chimeric nucleic acid construct into a plant cell; and

[0083] (c) growing the plant cell into a mature plant capable of setting seed wherein the seed expresses insulin.

Terms and Definitions

[0084] Unless defined otherwise, all technical and scientific terms used herein shall have the same meaning as is commonly understood by one skilled in the art to which the present invention belongs. Where permitted, all patents, applications, published applications, and other publications, including nucleic acid and polypeptide sequences from GenBank, SwissPro and other databases referred to in the disclosure are incorporated by reference in their entirety.

[0085] The term “nucleic acid sequence” as used herein refers to a sequence of nucleoside or nucleotide monomers consisting of naturally occurring bases, sugars and internuclear (backbone) linkages. The term also includes modified or substituted sequences comprising non-naturally occurring monomers or portions thereof. The nucleic acid sequences of the present invention may be deoxyribonucleic acid sequences (DNA) or ribonucleic acid sequences (RNA) and may include naturally occurring bases including adenine, guanine, cytosine, thymidine and uracil; and xanthine and hypoxanthine.

[0086] The terms “nucleic acid sequence encoding insulin” and “nucleic acid sequence encoding an insulin polypeptide”, which may be used interchangeably herein, refer to any and all nucleic acid sequences encoding an insulin polypeptide, including the insulin polypeptides listed in Table 1 (SEQ ID NO: 7 to 145) as well as any mammalian insulin polypeptide and any nucleic acid sequences that encode pro-insulin and preproinsulin. As used herein “proinsulin” refers to an insulin polypeptide which includes the connecting peptide or “C-peptide” linking the B and A insulin polypeptide chains. In native human insulin the C-peptide is the 31 amino acid residue polypeptide chain connecting residue B30 to residue A1. The term “preproinsulin” refers to a proinsulin molecule additionally comprising an N-terminal signal sequence which directs translation to occur on the ER ribosomes. Nucleic acid sequences encoding an insulin polypeptide further include any and all nucleic acid sequences which (i) encode polypeptides that are substantially identical to the insulin polypeptide sequences set forth herein; or (ii) hybridize to any nucleic acid sequences set forth herein at least moderately stringent hybridization conditions or which would hybridize thereto under at least moderately stringent conditions but for the use of synonymous codons.

[0087] By the term “substantially identical” it is meant that two polypeptide sequences preferably are at least 75% identical, and more preferably are at least 85% identical and most preferably at least 95% identical, for example 96%, 97%, 98% or 99% identical. In order to determine the percentage of identity between two polypeptide sequences the amino acid sequences of such two sequences are aligned, preferably
using the Clustal W algorithm (Thompson, J.D., Higgins D.G., Gibson T.J., 1994, Nucleic Acids Res. 22 (22): 4673-4680, together with BLOSUM 62 scoring matrix (Henikoff S. and Henikoff J. G., 1992, Proc. Natl. Acad. Sci. USA 89: 10915-10919) and a gap opening penalty of 10 and gap extension penalty of 0.1, so that the highest order match is obtained between two sequences wherein at least 50% of the total length of one of the sequences is involved in the alignment. Other methods that may be used to align sequences are the alignment method of Needleman and Wunsch (J. Mol. Biol., 1970, 48: 443), as revised by Smith and Waterman (Adv. Appl. Math., 1981, 2: 482) so that the highest order match is obtained between the two sequences and the number of identical amino acids is determined between the two sequences. Other methods to calculate the percentage identity between two amino acid sequences are generally art recognized and include, for example, those described by Carillo and Lipton (SIAM J. Applied Math., 1988, 48:1073) and those described in Computational Molecular Biology, Lesk, ed. Oxford University Press, New York, 1988, Biocomputing: Informatics and Genomics Projects. Generally, computer programs will be employed for such calculations. Computer programs that may be used in this regard include, but are not limited to, CCG (Devereux et al., Nucleic Acids Res., 1984, 12: 387) BLASTP, BLASTN and FASTA (Altschul et al., J. Molec. Biol., 1990: 215:403).

By “at least moderately stringent hybridization conditions” it is meant that conditions are selected which promote selective hybridization between two complementary nucleic acid molecules in solution. Hybridization may occur to all or a portion of a nucleic acid sequence molecule. The hybridizing portion is typically at least 15 (e.g. 20, 25, 30, 40 or 50) nucleotides in length. Those skilled in the art will recognize that the stability of a nucleic acid duplex, or hybrids, is determined by the Tm, which in sodium containing buffers is a function of the sodium ion concentration and temperature (Tm ~ 81.5°C -16.6 (log[Na+]) +0.41% (G+C) -6001), or similar equation). Accordingly, the parameters in the wash conditions that determine hybrid stability are sodium ion concentration and temperature. In order to identify molecules that are similar, but not identical, to a known nucleic acid molecule a 1% mismatch may be assumed to result in about a 1°C decrease in Tm, for example if nucleic acid molecules are sought that have a >95% identity, the final wash temperature will be reduced by about 5°C. Based on these considerations those skilled in the art will be able to readily select appropriate hybridization conditions. In preferred embodiments, stringent hybridization conditions are selected. By way of example the following conditions may be employed to achieve stringent hybridization: hybridization at 5x sodium chloride/sodium citrate (SSC)/5x Denhardt’s solution/1.0% SDS at Tm ~ 5°C. Based on the above equation, followed by a wash of 0.2xSSC/0.1% SDS at 60°C. Moderately stringent hybridization conditions include a washing step in 3xSSC at 42°C. It is understood however that equivalent stringencies may be achieved using alternative buffers, salts and temperatures. Additional guidance regarding hybridization conditions may be found in: Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 1989, 6.3. 1.-6.3.6 and in: Sambrook et al., Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989, Vol. 3.

As used herein the terms “insulin” and “insulin polypeptide”, which may be used interchangeably herein, refer to any and all insulin polypeptides including the insulin polypeptides listed in Table 1 (SEQ ID NO: 7 to 145) as well as a polypeptide molecule comprising a sequence of amino acid residues which is (i) substantially identical to the amino acid sequences constituting any insulin polypeptide set forth herein or (ii) encoded by a nucleic acid sequence capable of hybridizing under at least moderately stringent conditions to any nucleic acid sequence encoding insulin set forth herein or capable of hybridizing under at least moderately stringent conditions to any nucleic acid sequence encoding insulin set forth herein but for the use of synonymous codons. The terms insulin and insulin polypeptide include pro-insulin polypeptides and mini-insulin polypeptides. The insulin polypeptide is preferably of human, porcine or bovine origin.

The term “polypeptide capable of retaining the insulin polypeptide within a membrane enclosed intracellular compartment” as used herein refers to any polypeptide, which when linked to an insulin polypeptide, is capable of sequestering the insulin polypeptide in a subcellular structure surrounded by a membrane and located within the intracellular space of the plant cell as defined by the plant cell’s plasma membrane.

The term “a polypeptide capable of retaining the insulin polypeptide in the ER or in an ER derived storage vesicle” as used herein refers to any polypeptide, which when linked to an insulin polypeptide is capable of sequestering the insulin polypeptide either in the endoplasmic reticulum or in a storage compartment which is derived from an endoplasmic reticulum, such as for example an oil body, within a plant cell.

The term “oil body” or “oil bodies” as used herein refers to any oil or fat storage organelle in a plant seed cell (described in for example: Huang (1992) Ann. Rev. Plant Mol. Biol. 43: 177-200).

The term “chimeric” as used herein in the context of nucleic acid sequences refers to at least two linked nucleic acid sequences which are not naturally linked. Chimeric nucleic acid sequences also include nucleic acid sequences of different natural origins. For example a nucleic acid sequence constituting a plant promoter linked to a nucleic acid sequence encoding human insulin is considered chimeric. Chimeric nucleic acid sequences also may comprise nucleic acid sequences of the same natural origin, provided they are not naturally linked. For example a nucleic acid sequence constituting a promoter obtained from a particular cell-type may be linked to a nucleic acid sequence encoding a polypeptide obtained from that same cell-type, but not normally linked to the nucleic acid sequence constituting the promoter. Chimeric nucleic acid sequences also include nucleic acid sequences comprising any naturally occurring nucleic acid sequence linked to any non-naturally occurring nucleic acid sequence.

Preparation of Recombinant Expression Vectors Comprising Chimeric Nucleic Acid Sequences Encoding Insulin and a Promoter Capable of Controlling Expression in a Plant Seed Cell

The nucleic acid sequences encoding insulin that may be used in accordance with the methods and compositions provided herein may be any nucleic acid sequence encoding an insulin polypeptide, including any proinsulin and preproinsulin.

Exemplary nucleic acid sequences encoding insulin are well known to the art and are generally readily available from a diverse variety of mammalian sources including
human (Bell, G. I. et al., 1980, Nature 284:26-32), porcine (Clunie, R. E. et al., 1968, Science 161:165-167), bovine (D’Agostino, J. et al., 1987, Mol. Endocrinol. 1:327-331), ovine (Peterson, J. D. et al., 1972, Biol. Chem. 247:4866-4871) and the like, as well as from plant sources (Oliveira, A. E. A. et al., 1999, Protein Pept. Lett. 6:15-21). Insulin encoding sequences that may used include those encoding polypeptide chains set forth as SEQ ID NO:7 to SEQ ID NO:145. The respective corresponding nucleic acid sequences encoding the insulin polypeptide chains can be readily identified via the Swiss Protein identifier numbers provided in Table 1. Using these nucleic acid sequences, additional novel insulin encoding nucleic acid sequences may be readily identified using techniques known to those of skill in the art. For example libraries, such as expression libraries, cDNA and genomic libraries, may be screened, and databases containing sequence information from sequencing projects may be searched for similar sequences. Alternative methods to isolate additional nucleic acid sequences encoding insulin polypeptides may be used, and novel sequences may be discovered and used in accordance with the present invention. In preferred embodiments nucleic acid sequences encoding insulin are human, porcine and bovine insulin.

Numerous insulin analogs are known to the prior art (see for example U.S. Pat. Nos. 5,461,031; 5,474,978; 5,164,366 and 5,008,241) and may be used in accordance with the present invention. Analogs that may be used herein include human insulin molecules in which amino acid residue 28 of the B-chain (B28) has been changed from its natural proline residue to aspartate, lysine or isoleucine. In another embodiment the lysine residue at B29 is modified to a proline. Furthermore, the asparagine at A21 may be changed to alanine, glutamine, glutamate, glycine, histidine, isoleucine, leucine, methionine, serine, threonine, tryptophan, tyrosine or valine. Also, asparagine at B3 may be modified to lysine. Further examples of insulin analogs that may be used herein include: human insulin lacking the B30 residue, also frequently referred to as "desB30" or "B(1-29)"; those lacking the last 3 amino acid residues insulin "B(1-27)"; insulin molecules lacking the phenylalanine residue at B1; and analogs wherein the A-chain or the B-chain have an N-terminal or C-terminal extension, for example the B-chain may be N-terminally extended by the addition of two arginine residues.

In preferred embodiments, the nucleic acid sequence encoding insulin that is used is proinsulin. In further preferred embodiments nucleic acid sequence molecules encoding insulin are used in which the C-peptide has been modified relative to its native form. Amino acid residues in the C-peptide may be substituted and the C-peptide may be lengthened or shortened. In this regard, as used herein the term “mini-insulin” refers to an insulin polypeptide which has been modified so that the C-peptide has been shortened in length relative to its native form. In preferred embodiments a mini-insulin is used. Preferably the C-peptide of a mini-insulin molecule will be shorter than 20 amino acid residues, more preferably shorter than 15 amino acid residues and most preferably shorter than 9 amino acid residues, for example 7, 5, or 3 residues. Preferably, as is the case with natural insulin molecules, the mini-insulin C-peptide will comprise a cleavage site at its C- and N-termini. Such cleavage sites may be at any convenient sites known to the art, for example methionine, cleavable by cyanogen bromide, a single basic residue or a pair of basic residues, cleavable by trypsin or trypsin like proteases or by a carboxy-peptidase. For example, the C-peptide may comprise a C-terminal lysine, for example Ala-Ala-Lys (SEQ ID NO:146), or a dibasic processing site immediately prior to the GlyA1 residue, for example Asn-Ala-Lys (SEQ ID NO:147), or Arg-Arg-Lys-Gln-Lys-Arg (SEQ ID NO:148) or a tetrabasic processing site immediately prior to the Gly A1 residue, for example Arg-Arg-Lys-Lys (SEQ ID NO:149). Accordingly, mini-insulin molecules that may be used in accordance with the present invention include:

$$B(1-29/30) - X_1 - X_2 - X_3 - Y_1 - A(1-21)$$

wherein

- $X_1$ is any amino acid;
- $X_2$ is any amino acid;
- $X_3$ is Lys or Arg;
- $Y_1$ is peptide bond or 1-17 amino acid residues;

- $B(1-29/30)$ is the B chain of the human insulin B-chain containing amino acid residues 1-29 or 1-30; and

- $A(1-21)$ is the A chain of the human insulin A-chain containing amino acid residues 1-21.

In a preferred embodiment, $X_1$ is a basic amino acid residue (Lys or Arg), and $Y_1$ is either a peptide bond or 1-17 amino acid residues where the C-terminal residue is a basic amino acid residue (Lys or Arg).

Further, mini-insulin molecules that may be used herein comprise those that may be represented by the formula:

$$B(1-27) - X_2 - X_3 - X_4 - Y_1 - A(1-21)$$

wherein

- $X_2$ is a peptide of 1-8 amino acid residues comprising at least one aromatic amino acid residue;
- $X_3$ is one of Pro, Asp, Lys or Ile at position 28 of the B-chain;
- $X_4$ is one of Pro, Lys, Ala, Arg or Pro-Thr at position 29 of the B-chain;
- $Y_1$ is Lys or Arg;

- $B(1-27)$ is the B chain of the human insulin B-chain containing amino acid residues 1-27; and

- $A(1-21)$ is the A chain of the human insulin A-chain containing amino acid residues 1-21.

Additional examples of nucleic acid molecules encoding a mini-insulin polypeptides that may be used in accordance with the present invention include those described in: Markussen et al., Walter de Gruyter & Co. 1987, in: Peptides pp 189-194; Thim et al., 1989, in: Genetics and molecular biology of industrial microorganisms, American Society for Microbiology pp 322-328; and those set forth in U.S. Pat. Nos. 4,916,212; 5,324,641 and 6,521,738. Alterations to the nucleic acid sequence encoding insulin to prepare insulin analogs may be made using a variety of nucleic acid modification techniques known to those skilled in the art, including for example site directed mutagenesis, targeted mutagenesis, random mutagenesis, the addition of organic solvents, gene shuffling or a combination of these and other techniques known to those of skill in the art (Shraiberi et al., 1988, Arch. Biochem. Biophys. 358: 104-115; Gilkin et al., 1997, Protein Eng. 10: 687-690; Carugo et al., 1997, Proteins 28: 10-28; Hurley et al., 1996, Biochem. 35: 5670-5678; Holenberg et al., 1999, Protein Eng. 12: 851-856).

In accordance with the present invention it has surprisingly been found that insulin accumulates to levels in plant seeds heretofore not achieved if the insulin is expressed in seed, preferably in such a manner that the insulin polypeptide within the seed cells is sequestered in a membrane.
enclosed intracellular compartment. In preferred embodiments of the present invention the insulin polypeptide is sequestered in the ER or in an ER derived storage vesicle. In order to achieve such sequestration of insulin in the ER or an ER derived storage vesicle, in accordance with the present invention, the polypeptide encoding insulin is linked to a polypeptide which causes the insulin polypeptide to be retained in the ER or in an ER derived storage vesicle rather than being transported out of the ER to, for example, the apoplast. Polypeptides that may be used in accordance with the present invention to retain the insulin polypeptide in the ER include any polypeptide capable of sequestering the insulin in the ER. Such polypeptides may be synthesized or obtained from any biological source. In a preferred embodiment of the present invention, the polypeptide that is capable of retaining the insulin is a polypeptide comprising a C-terminal ER retention motif. Examples of such C-terminal ER retention motifs include KDEL, HDEL, DDEL, ADEL and SDEL sequences (SEQ ID NO:150 to 154 respectively). Other examples include HDEF (SEQ ID NO:155) (Lehmann et al., 2001, Plant Physiol 127(2): 436-49.), or two arginine residues dose to the N-terminus located at positions 2 and 3, 3 and 4, or 4 and 5 (Abstract from Plant Biology 2001 Program, ASPB, July 2001, Providence, R.I., USA.). Nucleic acid sequences encoding a C-terminal ER retention motif are preferably linked to the nucleic acid sequence encoding the insulin polypeptide in such a manner that the polypeptide capable of retaining the insulin in the ER is linked to the C-terminal end of the insulin polypeptide.

In order to achieve sequestration of the insulin polypeptide in an ER derived storage vesicle, the insulin polypeptide is linked to a polypeptide that is capable of retaining the insulin polypeptide in an ER derived storage vesicle. The polypeptide capable of retaining the insulin polypeptide in an ER derived storage vesicle may be used in accordance with the methods herein or may be any polypeptide capable of sequestering the insulin polypeptide in an ER derived storage vesicle. Polypeptides capable of retaining insulin in an ER derived storage vesicle may be synthesized or obtained from any biological source. In a preferred embodiment the ER derived storage vesicle is an oil body and the insulin polypeptide is linked to an oil body protein or a sufficient portion thereof capable of retaining the insulin polypeptide in the ER derived storage vesicle. Oil body proteins that may be used in this regard include any protein that naturally associates with an oil body. Oil body proteins that are particularly preferred are oleosins, for example an Arabidopsis oleosin (van Rooijen et al. (1991) Plant Mol. Biol. 18: 1177-1179) corn oleosin (Bowman-Vance et al., 1987, J. Biol. Chem. 262: 11275-11279; Qu et al., 1990, J. Biol. Chem. 265: 2238-2243), carrot oleosin (Hatzopoulos et al. (1990) Plant Cell 2: 457-457) or Brassica oleosa (Lee et al., 1991, Plant Physiol. 96: 1395-1397), oleosins, see for example Genbank accession number AF067857 and sterolesins (Lin et al., 2002 Plant Physiol. 128(4):1200-11). In a further preferred embodiment, the oil body polypeptide is a plant oleosin and shares sequence similarity with other plant oleosins such as the oleosin isolated from Arabidopsis thaliana (SEQ ID NO:156) or Brassica napus (SEQ ID NO:157). In another embodiment, the oil body polypeptide is a oleosin or calcium binding protein from plant, fungal or other sources and shares sequence homology with plant oleosins such as the oleosin isolated from Arabidopsis thaliana (SEQ ID NO:158 and SEQ ID NO:159) or in another embodiment the oil body protein is a steroleosin (SEQ ID NO:160), a sterol binding dehydrogenase (Lin L-J et al, (2002) Plant Physiol 128: 1200-1211). The polypeptide encoding insulin may be linked to the oil body protein to the N-terminus as well as to the C-terminus and to fragments of an oil body protein, such as for example the central domain of an oleosin. New oil body proteins may be discovered for example by preparing oil bodies (for methodologies to prepare oil bodis see for example U.S. Pat. No. 6,550,555) or in oil body preparations through for example SDS gel electrophoresis. Polyclonal antibodies may be raised against these proteins and used to screen cDNA libraries in order to identify nucleic acid sequences encoding oil body proteins. New oil body proteins further may be discovered using known nucleic acid sequences encoding oil body proteins, for example the herebefore mentioned oil body polypeptide sequences encoding oil body proteins, to probe for example cDNA or genomic libraries for the presence of oil body proteins. [0115] Polypeptides capable of retaining the insulin in the ER or an ER derived storage organelle are typically not cleaved and the insulin may accumulate in the form of a fusion protein, which is, for example, typically the case when a KDEL retention signal is used to retain the polypeptide in the ER or when an oil body protein is used to retain the polypeptide in an ER derived storage organelle. [0116] The chimeric nucleic acid sequence additionally may contain a nucleic acid sequence which targets the nucleic acid sequence to the endomembrane system (“signal peptide”). In embodiments of the present invention in which the insulin polypeptide is retained in the ER using a sequence capable of retaining the polypeptide in the ER, such as KDEL, HDEL or SDEL, polypeptide, it is particularly desirable to include a nucleic acid sequence encoding a signal peptide. Exemplary signal peptides that may be used herein include the tobacco pathogenesis related protein (PR-S) signal sequence (SEQ: ID NO:161) (Sijmons et al., 1990, Bio/technology, 8:217-221), lectin signal sequence (Boehn et al., 2000, Transgenic Res, 9(6):477-86), signal sequence from the barley α-amylase rich glycoprotein from Phaseolus vulgaris (Yan et al., 1997, Plant Physiol. 115(3):915-24 and Corbin et al., 1987, Mol Cell Biol 7(12):4337-44), potato patatin signal sequence (Iturria, G., et al., 1989, Plant Cell 1:381-390 and Bevan et al., 1986, Nuc. Acids Res. 41:4625-4638) and the barley alpha amylase signal sequence (Rasmussen and Johnsson, 1992, Plant Mol. Biol. 18(2):423-7). Such targeting signals may in vivo be cleaved off from the insulin sequence, which for example is typically the case when an apoplast targeting signal, such as the tobacco pathogenesis related protein-S (PR-S) signal sequence (Sijmons et al., 1990, Bio/technology, 8:217-221) is used. Other signal peptides can be predicted using the SignalP World Wide Web server (http://www. cbs.dtu.dk/services/SignalP/) which predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms. In general there is little conservation of the primary amino acid sequence, although general physicochemical properties are conserved to some extent. The generic structure of signal peptides has 5 regions, the short amino-terminal “n-region” contains positively charged residues, the central hydrophobic “h-region” ranges in size from 7 to 15 amino acids and the carboxy-terminal “c-region” contains polar amino acids and a cleavage site that is recognized by membrane bound signal peptidase enzymes (Nakai K., 2000, Advances in Protein Chem 54:277-344). A targeting signal that also may be used in accordance herewith
includes the native insulin signal sequence (24 amino acids in length in case of the human sequence). In preferred embodiments hereof an N-terminally located aprotinin targeting sequence, such as the hereinbefore mentioned tobacco PR-S sequence is used combined with a C-terminally located ER retention sequence such as the KDEL sequence.

[0117] In a further preferred embodiment, a nucleic acid sequence encoding a yeast α-factor leader sequence is linked to the N-terminal end of the nucleic acid sequence encoding insulin. Yeast leader sequences or sequences derived from yeast leader sequences that may be used in accordance here-with include those listed in SEQ ID NO:162 to SEQ ID NO:171 (Kjeldsen et al., 2001, Biotechnology and Genetic Engineering Reviews 18: 89-121). Such leader sequences may further comprise a spacer peptide located C-terminally of the nucleic acid encoding the leader sequence and N-terminally of the sequence encoding insulin. In accordance here-with such spacer sequences typically are between 2 and 20 amino acids in length. Thus, for example, spacer sequences SEQ ID NO:172 and SEQ ID NO:173 (Kjeldsen et al., 2001, Biotechnology and Genetic Engineering Reviews 18: 89-121) may be used. In embodiments of the present invention in which a yeast leader sequence is used, the nucleic acid sequence encoding the insulin polypeptide is preferably a mini-insulin polypeptide. In accordance herewith, in a particularly preferred embodiment a nucleic acid sequence encoding a single-chain antibody linked to a nucleic acid sequence encoding a yeast secretion leader peptide is used, as further described in Example 1 hereof.

[0118] The chimeric nucleic acid sequence may also comprise polypeptides resulting in N- and/or C-terminal stabilizing protein extensions. Such extensions may be used to stabilize and/or assist in folding of the insulin polypeptide chain and additionally may be used to facilitate purification of insulin. Polypeptide extensions that may be used in this regard include for example a nucleic acid sequence encoding a single chain antibody, a nucleic acid encoding a fibrinogen B-molecule (Affibody AB), a nucleic acid encoding the non-toxic B subunit of cholera toxin (CTB) (Arakawa, T. et al., 1998, Nat. Biotechnol. 16:938) or combinations of such polypeptides. In a particularly preferred embodiment, the insulin polypeptide is retained in a membrane enclosed compartment, such as the ER, using for example a KDEL sequence as hereinbefore described, combined with a stabilizing polypeptide that permits the association of the insulin polypeptide with an oil body upon breakage of the integrity of the plant cell such as will occur when the insulin polypeptide is recovered from the plant cell. An example of such a stabilizing polypeptide is a single chain antibody with specificity for an oil body. Nucleic acid sequences encoding single chain antibodies with specificity for an oil body may be prepared from hybridoma cell lines expressing monoclonal antibodies raised against an oil body protein. In one embodiment, the single chain antibody specifically binds an oleosin, as described by Alting-Moos et al. (2000) IBC’s International Conference on Antibody Engineering, Poster #1. This embodiment of the present invention is further detailed in Example 1 hereof.

[0119] In a further embodiment, a cleavage site may be located upstream of the N-terminus and downstream the C-terminus of the insulin allowing for the insulin polypeptide to be cleaved from the fusion partner, thereby obtaining isolated insulin. Examples of such cleavage sites can be found in WO 98/49326 (Method for the cleavage of fusion proteins) and related applications and LaValle et al. (1994) Enzymatic and chemical cleavage of fusion proteins in Current Protocols in Molecular Biology pp 16.4.5-16.4.17, John Wiley and Sons, Inc., New York N.Y. In a preferred embodiment, the cleavage site is a tetrasaccharide linker (for example Arg-Lys-Lys-Arg—SEQ ID NO:149) which is cleavable by trypsin. In a further preferred embodiment, the cleavage site is KLIP 8 (SEQ ID NO:174) which is cleavable by aspartic proteases including chymosin.

[0120] The invention further provides methods for the separation of heterologous proteins from host cell components by partitioning of the oil body fraction and subsequent release of the heterologous protein via specific cleavage of the heterologous protein—oil body protein fusion. Optionally a cleavage site may be located upstream of the N-terminus and downstream of the C-terminus of the heterologous polypeptide allowing the fusion polypeptide to be cleaved and separated by phase separation into its component peptides.

[0121] The nucleic acid sequence encoding insulin may be altered, to further improve expression levels for example, by optimizing the nucleic acid sequence in accordance with the preferred codon usage for a particular plant cell type which is selected for the expression of the insulin polypeptide, or by altering motifs known to destabilize mRNAs (see for example: PCT Patent Application 97/02352). Comparison of the codon usage of the nucleic acid sequence encoding the insulin polypeptide with the codon usage of the plant cell type will enable the identification of codons that may be changed. Construction of synthetic genes by altering the codon usage is described in for example PCT Patent Application 93/07278.

[0122] In a preferred embodiment, the nucleic acid sequence encoding insulin that is used is represented by SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:195.

[0123] In accordance herewith the nucleic acid sequence encoding insulin is linked to a promoter capable of controlling expression of the insulin polypeptide in a plant seed cell. Accordingly, the present invention also provides a nucleic acid sequence encoding insulin linked to a promoter capable of controlling expression in a plant seed cell. Promoters that may be used herein will be generally art recognized and include any plant derived promoter capable of controlling expression of polypeptides in plants. Generally, promoters obtained from dicotyledonous plant species will be used when a dicotyledonous plant is selected in accordance herewith, whereas a monocotyledonous plant species will be used when a monocotyledonous plant species is selected. Constitutive promoters that may be used include, for example, the 35S cauliflower mosaic virus (CaMV) promoter (Rothstein et al., 1987, Gene 53: 153-161), the rice actin promoter (McElroy et al., 1990, Plant Cell 2: 163-171; U.S. Pat. No. 6,429,357), a ubiquitin promoter, such as the corn ubiquitin promoter (U.S. Pat. Nos. 5,879,903; 5,273,894), and the parsley ubiquitin promoter (Kowalkeck, P. et al., 1993, Plant Mol. Biol. 21:673-684).

[0124] In preferred embodiments, the promoter that is used is a promoter which results in preferential expression of the insulin polypeptide in seed tissue. “Seed-preferred promoters” in this regard are promoters which control expression of a recombinant protein (i.e. insulin) so that preferably at least 80% of the total amount of recombinant protein present in the mature plant is present in the seed. More preferably at least 90% of the total amount of recombinant protein present in the mature plant is present in the seed. Most preferably at least 95% of the total amount of recombinant protein present in the
mature plant is present in the seed. Seed-preferred promoters that may be used in this regard include, for example, the bean phaseolin promoter (Sengupta-Gopalan et al., 1985, Proc. Natl. Acad. Sci. USA 82: 3320-3324); the Arabidopsis 18 kDa oleosin promoter (U.S. Pat. No. 5,792,922) or the 10 kDa oleosin promoter (WO 01/16340); the flux legumin like seed storage protein (Ibnin) promoter (WO 01/16340); the flux 2S storage protein promoter (WO 01/16340); an endosperm preferred promoter such as the Amy32b promoter (Rogers and Milliman, J. Biol. Chem., 1984, 259: 12234-12240, the Amy6-p promoter (Kursheed and Rogers, J. Biol. Chem., 1988, 263: 18953-18960 or the Aleurain promoter (Whittier et al., 1987, Nucleic Acids Res., 15: 2515-2535) or the bean aracelin promoter (Juenger G D. et al., 2002, Nat. Biotechnol. December 20: 1265-8). New promoters useful in various plants are constantly discovered. Numerous examples of plant promoters may be found in Ohamura et al. (Biochem. of Plnts., 1989, 15: 1-82).

[0125] Certain genetic elements capable of enhancing expression of the insulin polypeptide may be used herein. These elements include the untranslated leader sequences from certain viruses, such as the AMV leader sequence (Jobling and Gishke, 1987, Nature, 325: 622-625) and the intron associated with the maize ubiquitin promoter (U.S. Pat. No. 5,504,200). Generally the chimeric nucleic acid sequence will be prepared so that genetic elements capable of enhancing expression will be located 5’ to the nucleic acid sequence encoding the insulin polypeptide.

[0126] In accordance with the present invention the chimeric nucleic acid sequences comprising a promoter capable of controlling expression in plant cells linked to a nucleic acid sequence encoding an insulin polypeptide can be integrated into a recombinant expression vector which ensures good expression in the seed cell. Accordingly, the present invention includes a recombinant expression vector comprising in the 5’ to 3’ direction of transcription as operably linked components:

- [0127] (i) a nucleic acid sequence capable of controlling expression in plant seed cells; and
- [0128] (ii) a nucleic acid sequence encoding an insulin polypeptide;

[0129] wherein the expression vector is suitable for expression in a seed cell. The term “suitable for expression in a seed cell” means that the recombinant expression vector comprises the chimeric nucleic acid sequence of the present invention linked to genetic elements required to achieve expression in a seed cell. Genetic elements that may be included in the expression vector in this regard include a transcriptional termination region, one or more nucleic acid sequences encoding marker genes, one or more origins of replication and the like. Preferred embodiment of the expression vector further comprises genetic elements required for the integration of the vector or a portion thereof in the plant cell’s nuclear genome, for example the T-DNA left and right border sequences which facilitate the integration into the plant’s nuclear genome in embodiments of the invention in which plant cells are transformed using Agrobacterium.

[0130] As hereinbefore mentioned, the recombinant expression vector generally comprises a transcriptional terminator which besides serving as a signal for transcription termination further may serve as a protective element capable of extending the mRNA half-life (Guarnieros et al., 1982, Proc. Natl. Acad. Sci. USA, 79: 238-242). The transcriptional terminator is generally from about 200 nucleotides to about 1000 nucleotides and the expression vector is prepared so that the transcriptional terminator is located 3’ of the nucleic acid sequence encoding the insulin. Termination sequences that may be used herein include, for example, the nopaline termination region (Bevan et al., 1983, Nucl. Acids. Res., 11: 369-385), the phaseolin terminator (van der Geest et al., 1994, Plant J. 6: 413-423), the aracelin terminator (Juenger G D. et al., 2002, Nat. Biotechnol. December 20: 1265-8), the terminator for the octopine synthase genes of Agrobacterium tumefaciens or other similarly functioning elements. Transcriptional terminators may be obtained as described by An (An, 1987, Methods in Enzym. 153: 292).

[0131] Pursuant to the present invention the expression vector may further contain a marker gene. A marker gene that may be used in accordance with the present invention include all genes that allow the distinction of transformed cells from non-transformed cells, including all selectable and screenable marker genes. A marker gene may be a resistance marker such as an antibiotic resistance marker against, for example, kanamycin (U.S. Pat. No. 6,174,724), ampicillin, G418, bleomycin, hygromycin which allows selection of a trait by chemical means or a tolerance marker against a chemical agent, such as the normally phytotoxic sugar mannosone (Negrotto et al., 2000, Plant Cell Rep. 19: 798-803). Other convenient markers that may be used herein include markers capable of conveying resistance against herbicides such as glyphosate (U.S. Pat. Nos. 4,940,355; 5,188,642), phosphonothricin (U.S. Pat. No. 5,879,958) or sulphoxynil (U.S. Pat. No. 5,633,437). Resistance markers, when linked in close proximity to nucleic acid sequence encoding the insulin polypeptide, may be used to maintain selection pressure on a population of plant cells or plants that have not lost the nucleic acid sequence encoding the insulin polypeptide. Screenable markers that may be employed to identify transformants through visual inspection include GUS (β-glucuronidase) (GUS) (U.S. Pat. Nos. 5,268,463 and 5,599,670) and green fluorescent protein (GFP) (Niedz et al., 1995, Plant Cell Rep., 14: 403).

[0132] Recombinant vectors suitable for the introduction of nucleic acid sequences into plants include Agrobacterium and Rhizobium based vectors, such as the Ti and Ri plasmids, including, for example pBin19 (Bevan, Nucl. Acid. Res., 1984, 22: 8711-8721), pKG585 (Boucher et al., 1993), C R Acad. Sci. Paris, Life Sciences, 316:1188-1193), the pCGN series of binary vectors (McBride and Summerfelt, 1990, Plant Mol. Biol., 14:269-276) and other binary vectors (e.g. U.S. Pat. No. 4,940,838).

[0133] The recombinant expression vectors, nucleic acid sequences and chimeric nucleic acid sequences of the present invention may be prepared in accordance with methodologies well known to those skilled in the art of molecular biology. Such preparation will typically involve the bacterial species Escherichia coli as an intermediary cloning host. The preparation of the E. coli vectors as well as the plant transformation vectors may be accomplished using commonly known techniques such as restriction digestion, ligation, gel electrophoresis, DNA sequencing, the Polymerase Chain Reaction (PCR) and other methodologies. These methodologies permit the linking of nucleic acid sequences and polypeptides to which the present invention pertains. A wide variety of cloning vectors is available to perform the necessary steps required to prepare a recombinant expression vector. Among the vectors with a replication system functional in E. coli, are vectors
such as pBR322, the pUC series of vectors, the M13mp series of vectors, phuscript etc. Typically, these cloning vectors contain a marker allowing selection of transformed cells. Nucleic acid sequences may be introduced in these vectors, and the vectors may be introduced in *E. coli* grown in an appropriate medium. Recombinant expression vectors may readily be recovered from cells upon harvesting and lysing of the cells. Further, general guidance with respect to the preparation of recombinant vectors may be found in, for example: Sambrook et al., Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989, Vol. 3.

Preparation of Plants Comprising Seeds Capable of Expressing Insulin

[0134] In accordance with the present invention the chimeric nucleic acid sequence is introduced into a plant cell and the cells are grown into mature plants capable of setting seed, wherein the seed expresses the insulin polypeptide.

[0135] In accordance herewith any plant species or plant cell may be selected. Particular cells used herein include cells obtainable from *Arabidopsis thaliana*, Brazil nut (*Beeaeholetia excelsa*); castor bean (*Ricinus communis*); coconut (*Cocos nucifera*); coriander (*Coriandrum sativum*); cotton (*Gossypium spp.*); groundnut (*Arachis hypogaea*); jojoba (*Simmondsia chinesis*); linoed/flax (*Linum usitatissimum*); maize (*Zea mays*); mustard (*Brassica spp.* and *Sinapis alba*); oil palm (*Elaeis guineensis*); olive (*Olea europaea*); rapeseed (*Brassica spp.*); rice (*Oryza sativa*); safflower (*Carthamus tinctorius*); soybean (*Glycine max*); squash (*Cucurbita maxima*); barley (*Hordeum vulgare*); wheat (*Triticum aestivum*); and sunflower (*Helianthus annuus*).

[0136] In accordance herewith in a preferred embodiment plant species or plant cells from oil seed plants are used. Oil seed plants that may be used herein include peanut (*Arachis hypogaea*); mustard (*Brassica spp.* and *Sinapis alba*); rapeseed (*Brassica spp.*); chickpea (*Cicer arietinum*); soybean (*Glycine max*); cotton (*Gossypium hirsutum*); sunflower (*Helianthus annuus*); *Lentil (Lens culinaris)*; linoed/flax (*Linum usitatissimum*); white clover (*Trifolium repens*); olive (*Olea europaea*); oil palm (*Elaeis guineensis*); safflower (*Carthamus tinctorius*); and barley (*Hordeum vulgare*).

[0137] In accordance herewith in a particularly preferred embodiment safflower, *Arabidopsis* or flax is used.

[0138] Methodologies to introduce plant recombinant expression vectors into a plant cell, also referred to herein as “transformation”, are well known to the art and typically vary depending on the plant cell that is selected. General techniques to introduce recombinant expression vectors in cells include, electroporation; chemically mediated techniques, for example CaCl₂ mediated nucleic acid uptake; particle bombardment (biolistics); the use of naturally infective nucleic acid sequences, for example virally derived nucleic acid sequences, or *Agrobacterium* or *Rhizobium* derived sequences, polyethylene glycol (PEG) mediated nucleic acid uptake, microinjection and the use of silicone carboide whiskeys.

[0139] In preferred embodiments, a transformation methodology is selected which will allow the integration of the chimeric nucleic acid sequence in the plant cell’s genome, and preferably the plant cell’s nuclear genome. In accordance herewith this is considered particularly desirable as the use of such a methodology will result in the transfer of the chimeric nucleic acid sequence to progeny plants upon sexual reproduction. Transformation methods that may be used in this regard include biolistics and *Agrobacterium* mediated methods.

[0140] Transformation methodologies for dicotyledenous plant species are well known. Generally, *Agrobacterium* mediated transformation is used because of its high efficiency, as well as the general susceptibility by many, if not all, dicotyledenous plant species. *Agrobacterium* transformation generally involves the transfer of a binary vector, such as one of the hereinbefore mentioned binary vectors, comprising the chimeric nucleic acid sequence of the present invention from *E. coli* to a suitable *Agrobacterium* strain (e.g., LBA4404 and LBA4404) by, for example, tri-parental mating with an *E. coli* strain carrying the recombinant binary vector and an *E. coli* strain carrying a helper plasmid capable of mobilizing the binary vector to the target *Agrobacterium* strain, or by DNA transformation of the *Agrobacterium* strain (Hofgen et al., Nucl. Acids Res., 1988, 16:3977). Other techniques that may be used to transform dicotyledenous plant cells include biolistics (Sanford, 1988, Trends in Biotechn. 6:299-302); electroporation (Fromm et al., 1985, Proc. Natl. Acad. Sci. USA., 82:5824-5828); PEG mediated DNA uptake (Potrykus et al., 1985, Mol. Gen. Genetics, 199:169-177); microinjection (Reich et al., BioTech., 1986, 4:1001-1004); and silicon carboide whiskeys (Kaeppler et al., 1990, Plant Cell Rep., 9:415-418) or in plants transformation using, for example, a flower dipping methodology (Clough and Bent, 1998, Plant J., 16:735-743).


[0142] The exact plant transformation methodology may vary depending on the plant species and the plant cell type (e.g. seedling derived cell types such as hypocotyls and cotyledons or embryonic tissue) that is selected as the cell target for transformation. As hereinbefore mentioned in a particularly preferred embodiment safflower, *Arabidopsis* or flax is used. A methodology to obtain safflower transformants is available in Baker and Dyer (Plant Cell Rep., 1996, 16:106-110). Additional plant species specific transformation protocols may be found in: Biotechnology in Agriculture and Forestry 46: Transgenic Crops I (Y. F. S. Bajaj ed.), Springer-Verlag, New York (1999), and Biotechnology in Agriculture and Forestry 47: Transgenic Crops II (Y. F. S. Bajaj ed.), Springer-Verlag, New York (2001).

[0143] Following transformation, the plant cells are grown and upon the emergence of differentiating tissue, such as shoots and roots, mature plants are regenerated. Typically a plurality of plants is regenerated. Methodologies to regenerate plants are generally plant species and cell type dependent and will be known to those skilled in the art. Further guidance with respect to plant tissue culture may be found in, for example: Plant Cell and Tissue Culture, 1994, Vasil and Torpe Eds., Kluwer Academic Publishers; and in: Plant Cell Culture Protocols (Methods in Molecular Biology 111), 1999, Hall Eds., Humana Press.

[0144] In one aspect, the present invention provides a method of recovering plant seeds comprising insulin. Accord-
ingly, the present invention provides a method for obtaining plant seeds comprising insulin comprising:

- [0145] (a) providing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:
  - [0146] (i) a nucleic acid sequence capable of controlling expression in plant seed cells; and
  - [0147] (ii) a nucleic acid sequence encoding an insulin polypeptide;
- [0148] (b) introducing the chimeric nucleic acid construct into a plant cell;
- [0149] (c) growing the plant cell into a mature plant capable of setting seed; and
- [0150] (d) obtaining seeds from said plant wherein the seed comprises insulin.

[0151] In preferred embodiments, a plurality of transformed plants is obtained, grown, and screened for the presence of the desired chimeric nucleic acid sequence, the presence of which in putative transformants may be tested by, for example, growth on a selective medium, where herbicide resistance markers are used, by direct application of the herbicide to the plant, or by Southern blotting. If the presence of the chimeric nucleic acid sequence is detected, transformed plant may be selected to generate progeny and ultimately mature plants comprising a plurality of seeds comprising the desired chimeric nucleic acid sequence. Such seeds may be used to isolate insulin or they may be planted to generate two or more subsequent generations. It will generally be desirable to plant a plurality of transgenic seeds to obtain a population of transgenic plants, each comprising seeds comprising a chimeric nucleic acid sequence encoding insulin. Furthermore, it will generally be desirable to ensure homozygosity in the plants to ensure continued inheritance of the recombinant polypeptide. Methods for selecting homozygous plants are well known to those skilled in the art. Methods for obtaining homozygous plants that may be used include the preparation and transformation of haploid cells or tissues followed by the regeneration of haploid plantlets and subsequent conversion to diploid plants for example by the treatment with colchicine or other microtubule disrupting agents. Plants may be grown in accordance with otherwise conventional agricultural practices.

[0152] In another aspect, the present invention also provides plants capable of setting seed expressing insulin. In a preferred embodiment of the invention, the plants capable of setting seed comprise a chimeric nucleic acid sequence comprising in the 5' to 3' direction of transcription:

- [0153] (a) a first nucleic acid sequence capable of controlling expression in a plant seed cell operatively linked to
- [0154] (b) a second nucleic acid sequence encoding an insulin polypeptide, wherein the seed contains insulin.

[0155] In a preferred embodiment the chimeric nucleic acid sequence is stably integrated in the plant’s nuclear genome.

[0156] In yet another aspect, the present invention provides plant seeds expressing insulin. In a preferred embodiment of the present invention, the plant seeds comprise a chimeric nucleic acid sequence comprising in the 5' to 3' direction of transcription:

- [0157] (a) a first nucleic acid sequence capable of controlling expression in a plant seed cell operatively linked to
- [0158] (b) a second nucleic acid sequence encoding an insulin polypeptide.

[0159] In accordance with the present invention, seed is obtained which in which preferably at least 0.1% of the total soluble protein present in the seed is insulin. In further preferred embodiments of the present invention, seed is obtained in which at least 0.2%, 0.3%, 0.5%, or even 1.0% of the total soluble protein present in the seed is insulin. The insulin polypeptide may be present in a variety of different types of seed cells including, for example, the hypocotyl and the embryonic axis, including in the embryonic roots and embryonic leaves, and where monocotyledonous plant species, including cereals and corn, are used in the endosperm tissue. Preparation of Insulin from Plant Seeds

[0160] Once the plant seeds have been obtained the insulin protein may be purified from the seed using any protein purification methodologies known in the art. Accordingly, pursuant to the present invention a method of purifying insulin from plant seeds is provided in which the method comprises:

- [0161] (a) providing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:
- [0162] (i) a nucleic acid sequence capable of controlling expression in plant seed cells; and
- [0163] (ii) a nucleic acid sequence encoding an insulin polypeptide;
- [0164] (b) introducing the chimeric nucleic acid construct into a plant cell;
- [0165] (c) growing the plant cell into a mature plant capable of setting seed wherein the seed expresses insulin;
- [0166] (d) obtaining seed expressing insulin; and
- [0167] (e) purifying said insulin from the seed.

The plant seeds may be ground using any comminuting process resulting in a substantial disruption of the seed cell membrane and cell walls. Both dry and wet milling conditions (U.S. Pat. No. 3,971,856; Lawhon et al., 1977, J. Am. Oil Chem. Soc., 63:533-534) may be used. Suitable milling equipment in this regard include colloid mills, disc mills, IKA mills, industrial scale homogenizers and the like. The selection of the milling equipment will depend on the seed type and throughput requirements. Solid seed contaminant such as seed hulls, fibrous materials, undissolved carbohydrates, proteins and other water insoluble contaminants may be removed from the seed fraction using for example size-exclusion based methodologies, such as filtering or gravitational based processes such as centrifugation. In preferred embodiments, the use of organic solvents commonly used in oil extraction, such as hexane, is avoided because such solvents may damage the insulin polypeptide. Substantially pure insulin may be recovered from the seed using a variety of additional purification methodologies such as centrifugation based techniques, size exclusion based methodologies, including for example membrane ultrafiltration and crossflow ultrafiltration; and chromatographic techniques, including for example ion-exchange chromatography, size exclusion chromatography, affinity chromatography, high performance liquid chromatography (HPLC); fast protein liquid chromatography (FPLC). Hydrophobic interaction chromatography and the like. Generally, a combination of such techniques will be used to obtain substantially pure insulin.

[0168] In a particularly preferred embodiment of the present invention the insulin polypeptide is isolated from the seed contaminants by contacting the insulin polypeptide with oil bodies. This method is considered to be particularly
advantageous as it permits the removal seed contaminants including seed proteins in a particularly efficacious and inexpensive manner. As hereinbefore mentioned such contacting of the insulin polypeptide with the oil bodies may be achieved by linking the insulin polypeptide to an oil body protein or by linking the insulin polypeptide to a polypeptide with affinity for an oil body, such as a single chain antibody with affinity for an oil body. In the former embodiment, insulin polypeptide will be sequestered within the cell on the oil bodies and hence co-purify with the oil bodies. In the latter embodiment, upon being expressed in a membrane enclosed intracellular compartment such as the ER, the insulin polypeptide will associate with the oil body upon breakage of the seed cells during the comminuting process. A process for isolating oil bodies is described in U.S. Pat. No. 5,650,554.

[0169] Pharmaceutical insulin formulations may be prepared from the purified insulin and such formulations may be used to treat diabetes. Generally the purified insulin will be admixed with a pharmaceutically acceptable carrier or diluent in amounts sufficient to exert a therapeutically useful effect in the absence of undesirable side effects on the patient treated. To formulate an insulin composition, the weight fraction of insulin is dissolved, suspended, dispersed or otherwise mixed in a selected carrier or diluent at an effective concentration such that the treated condition is ameliorated. The pharmaceutical insulin formulations are preferably formulated for single dosage administration. Therapeutically effective doses for the parenteral delivery of human insulin are well known to the art. Where insulin analogs are used or other modes of delivery are used therapeutically effective doses may be empirically determined by those of skill in the art using known testing protocols or by extrapolation of in-vivo or in-vitro test data. It is understood however that concentrations and dosages may vary in accordance with the severity of the condition alleviated. It is further understood that for any particular subject, specific dosage regimens may be adjusted over time according to individual judgement of the person administering or supervising administration of the formulations.

[0170] Pharmaceutical solutions or suspensions may include for example a sterile diluent such as, for example, water, lactose, sucrose, dibasic sodium phosphate, or carboxymethyl cellulose. Carriers that may be used include water, saline solution, aqueous dextrose, glycerol, glycols, ethanol and the like, to thereby form a solution or suspension. If desired the pharmaceutical compositions may also contain non-toxic auxiliary substances such as a wetting agents; emulsifying agents; solubilizing agents; antimicrobial agents, such as benzyl alcohol and methyl parabens; antioxidants, such as ascorbic acid and sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); pH buffering agents such as acetate, citrate or phosphate buffers; and combinations thereof.

[0171] The final formulation of the insulin preparation will generally depend on the mode of insulin delivery. The insulin prepared in accordance with the present invention may be delivered in a desired manner; however parenteral, oral, pulmonary, buccal and nasal forms of delivery are considered the most likely used modes of delivery. Parenteral preparations can be enclosed in ampoules, disposable syringes or single or multiple dose vials made of glass, plastic or other such suitable materials.

EXAMPLES

[0172] The following examples are offered by way of illustration and not by limitation.

Example 1
Preparation of an Insulin Protein Expressed as a Mini-Insulin (MI) Fusion Protein with a Trypsin Cleavable Pro-Peptide


[0174] One of the fusion proteins studied began with the tobacco pathogen related sequence (PRS) (Sijmons et al., 1990, Bio/technology, 8:217-221) which served as the signal peptide to target expression to the ER in a co-translational manner. Immediately downstream was a sequence encoding for a single-chain Fv antibody (scFv) with species-specific affinity against the 18 kDa oleosin from Arabidopsis thaliana denoted D9scFv, followed by a trypsin cleavable pro-peptide (KLP27) derived from the TA57 pro-peptide of yeast (Kjeldsen et al., 2001, Biotechnology and Genetic Engineering Reviews 18:89-121). This was followed by mini-insulin (MI) described by Kjeldsen et al. (2001) with the addition of a KDEL ER-retention signal at the C-terminal end of the polypeptide.

[0175] The backbone of this plasmid, pSBS4055, was based on the plant binary vector, pLZP200, described by Hjedukiewicz et al. (Plant Molecular Biology, 1994, 25:989-994). In place of the described multiple cloning site, a gene conferring host plant phosphonothricine resistance (Wohlleben et al., 1988, Gene 70:25-37) driven by the ubiquitin promoter/terminator from Petunia splendens (Kawalek et al., 1993, Plant. Mol. Bio., 21:673-684), was inserted between the left and right border sequences. In addition to this cassette, the β-phoscolin promoter/terminator from Phaseolus vulgaris (Slighthom et al., 1983, Proc. Natl. Acad. Sc. USA 80:1897-1901) driving PRS was sub-cloned. Standard PCR (Horton et al., 1989, Gene 77:61-68) was used to fuse the synthetic PRS-encoding sequence with attached Sphl/HindIII restriction endonuclease sites to the 3'-end of the phuscolin promoter to yield pSBS4011. A Sphl-D9scFv-Xhol, Swal, HindIII insert sequence was generated by PCR amplification of a D9scFv CDNA clone (Sean Hemmingsen lab, unpublished) with primers 1325 (GACATCGACATGTGATGACACGCT)SEQ ID NO:175 and 1326 (AAGCTTGGCCATTTAAATATCGAGACTGTGAGACTTTGCGCTTTG)SEQ ID NO:176. Subsequent ligation of this fragment at the Sphl/HindIII sites of pSBS4011 resulted in plasmid pSBS4055.

[0176] The Klip27-MI sequence was synthesized from four partially overlapping oligonucleotides which incorporated Arabidopsis thaliana codon usage to increase the success of efficient translation in a plant-based expression system. Oligonucleotides 1324 (GAAGAAAGAGAGCCGTTAAGCTTTGGTATTAACGAAACATCGTGGGGGAGGCGGAGGCGGGGAGGCGGGGGG)SEQ ID NO:177 and 1323 (CCTGTCGTAACAAAAATCTTCTTCTCCACACACAAATGTAAGAGAGGCTGCTCTGACAACA)SEQ ID NO:178 were annealed at their complimentary 20 nucleotide overlap and extended to form the 5'-end of the Klip27-MI fusion while the same was done with oligonucleotides 1322 (CTAAAGCTTGCTAAAGGAAATG)SEQ ID NO:179 and 1321 (GAAGCTTGGCCATTTAAATATCGAGACTGTGAGACTTTGCGCTTTG)SEQ ID NO:176.

(GAAGAAAGAGAGCCGTTAAGCTTTGGTATTAACGAAACATCGTGGGGGAGGCGGAGGCGGGGAGGCGGGGGG)
TCACAATTCCCTTACGACGCTT—SEQ ID NO:180 to form the 3' end. The two halves were ligated following restriction digestion with Bsu36I, to yield the full Klip27-MI coding sequence. PCR of this gene using primers 1364 (CTCGAGTCAAACTGTATGACACTGAAATC)—SEQ ID NO:181 and 1334 (AAGGCTCAATTCTTCTGTGGCAAXAGTTCACCAATG)—SEQ ID NO:182 attached a 5' Xhol restriction endonuclease cleavage site and the 3' KDEL DNA sequence plus HindIII cleavage site for subsequent ligation into Xhol/HindIII-cut pSB8405. The result was plasmid pSB84404: a DNA sequence encoding the PRS-D9scFv-Klip27-M1-KDEL fusion protein being placed in a binary vector under expression control of the phaseolin promoter/terminator. The phaseolin promoter controls the specific-temporal and tissue-specific expression of the transgene during seed development. The complete nucleic acid sequence (SEQ ID NO:1) and amino acid sequence (SEQ ID NO:2) of the 4404 insulin fusion protein (PRS-D9scFvKlip27-M1-KDEL) is shown in Fig. I. Construction of pSB84405:OLEO-Klip8-Klip27-M1 Fusion Protein

[0177] The second fusion protein studied began with the 18 kDa oleosin from Arabidopsis thaliana followed in-frame by a chymosin cleavable pro-peptide (Klip8)—SEQ ID NO:175. Immediately downstream was a sequence encoding for the trypsin cleavable pro-peptide (Klip27) derived from the TA57 pro-peptide of yeast as described above (Kjeldsen et al., 2001, Biotechnology and Genetic Engineering Reviews 18:89-121). This was fused to mini-insulin (MI) described above (Kjeldsen et al., 2001). The expression of this fusion protein was targeted to the nascent oil bodies formed during the development of the embryo.

[0178] The backbone of this plasmid, pSB84055, was based on the plant binary vector, pZP200, described by Hajdukiewicz et al. (Plant Molecular Biology, 1994, 25:989-994). In place of the described multiple cloning site, a pat gene conferring host plant phosphonothricine resistance (Wohlleben et al., 1988, Gene 70:25-37) driven by the ubiquitin promoter/terminator from Petroselinum crispum (Kawalleck et al., 1993, Plant. Mol. Bio., 21:673-684), was inserted between the left and right border sequences. In addition to this cassette, the β-phaseolin promoter/terminator from Phaseolus vulgaris (Slightom et al., 1983, Proc. Natl. Acad. Sc. USA 80:1897-1901) driving the Arabidopsis 18 kDa oleosin genomic sequence-Klip8 fusion was sub-cloned. Standard PCR (Horton et al., 1989, Gene 77:61-68) was used to fuse the oleosin gene-Klip8 sequence with an attached Xhol/HindIII restriction endonuclease sites to the 3' end of the phaseolin promoter to yield pSB84010.

[0179] The Klip27-MI sequence was synthesized from four partially overlapping oligonucleotides which incorporated Arabidopsis thaliana codon usage to increase the success of efficient translation in a plant-based expression system. Oligonucleotides 1324 (GAAGAAGAGGAGCCTTAAGTTTGTTAACACATTCTTGTGTGATCTCATC TTGTGATGAGCTCTACTATTAG) —SEQ ID NO:177 and 1325 (CCTTGGAGTTGTAGAAATAACCTTCTCTTCCACACAAGTGGTACCCAGG) —SEQ ID NO:178 were annealed at their complimentary 20 nucleotide overlap and extended to form the 5' end of the Klip27-MI fusion while the same was done with oligonucleotides 1322 (CTAAGGCTGTGCAAGGGAATTTG) —SEQ ID NO:179 and 1321 (AAGCTCTAGTGCAAAATGGTCTCCTGAATG) —GAGCAATAGAA GTGCAACCTGTGTCAACAATCCCTTAGCAGGTT)—SEQ ID NO:180 to form the 3' end. The two halves were ligated following restriction digestion with Bsu36I, to yield the full Klip27-MI coding sequence. PCR of this gene using primers 1364 (CTCGAGTCAAACTGTATGACACTGAAATC)—SEQ ID NO:181 and 1329 (AAGGCTCAATTCTTCTGTGGCAAXAGTTCACCAATG)—SEQ ID NO:183 attached a 5' Xhol restriction endonuclease cleavage site and a 3' HindIII cleavage site, respectively, for subsequent ligation into Xhol/HindIII-cut pSB84010. The result was plasmid pSB84405: a DNA sequence encoding the Oleosin-Klip8-Klip27-M1 fusion protein being placed in a binary vector under expression control of the phaseolin promoter/terminator. The phaseolin promoter controls the specific-temporal and tissue-specific expression of the transgene during seed development. The complete nucleic acid sequence SEQ ID NO:3 and amino acid sequence SEQ ID NO:4 of the 4405 insulin fusion protein (OLEO-Klip8-Klip27-MI) is shown in Fig. 2. Construction of pSB84414: PRS-MI-Tetrabasic Linker-D9scFv-KDEL Fusion Protein

[0180] Another fusion protein studied began with the tobacco pathogen related sequence (PRS) (Sijmons et al., 1990, Bio/technology, 8:217-221) which served as the signal peptide to target expression to the ER in a co-translationally manner. Immediately downstream was the sequence encoding mini-insulin (MI) described by Kjeldsen et al. (2001) with the exception that the mini-C propeptide region (AAK—SEQ ID NO:146) was replaced with an intervening B$_2$, Threonine-tetrabasic site (B$_6$TTRRKR) (SEQ ID NO:149) sequence between the B$_3$-$29$- and A$_{14-21}$-chains of human insulin. This was immediately followed by a sequence encoding for a second tetrabasic linker followed by a single-chain Fv antibody (scFv) with species-specific affinity against the 18 kDa oleosin from Arabidopsis thaliana denoted D9scFv. At the C-terminal end of the polypeptide was the addition of a KDEL ER-retention signal.

[0181] The plasmid backbone, pSB84055, was based on the plant binary vector, pZP200, described by Hajdukiewicz et al. (Plant Molecular Biology, 1994, 25:989-994). In place of the described multiple cloning site, a pat gene conferring host plant phosphonothricine resistance (Wohlleben et al., 1988, Gene 70:25-37) driven by the ubiquitin promoter/terminator from Petroselinum crispum (Kawalleck et al., 1993, Plant. Mol. Bio., 21:673-684), was inserted between the left and right border sequences. In addition to this cassette, the β-phaseolin promoter/terminator from Phaseolus vulgaris (Slightom et al., 1983, Proc. Natl. Acad. Sc. USA 80:1897-1901) driving the PRS was sub-cloned. Standard PCR (Horton et al., 1989, Gene 77:61-68) was used to fuse the described multiple cloning site, a pat gene conferring host plant phosphonothricine resistance (Wohlleben et al., 1988, Gene 70:25-37) driven by the ubiquitin promoter/terminator from Petroselinum crispum (Kawalleck et al., 1993, Plant. Mol. Bio., 21:673-684), was inserted between the left and right border sequences. In addition to this cassette, the β-phaseolin promoter/terminator from Phaseolus vulgaris (Slightom et al., 1983, Proc. Natl. Acad. Sc. USA 80:1897-1901) driving the PRS was sub-cloned. Standard PCR (Horton et al., 1989, Gene 77:61-68) was used to fuse the synthetic PRS-encoding sequence with attached Splh/HindIII restriction endonuclease sites to the 3' end of the phaseolin promoter to yield pSB84011.

[0182] The Klip27-MI sequence was synthesized from four partially overlapping oligonucleotides which incorporated Arabidopsis thaliana codon usage to increase the success of efficient translation in a plant-based expression system. Oligonucleotides 1324 (GAAGAAGAGGAGCCTTAAGTTTGTTAACACATTCTTGTGTGATCTCATC TTGTGATGAGCTCTACTATTAG) —SEQ ID NO:177 and 1325 (CCTTGGAGTTGTAGAAATAACCTTCTCTTCCACACAAGTGGTACCCAGG) —SEQ ID NO:178 were annealed at their complimentary 20 nucleotide overlap and extended to form the 5' end of the Klip27-MI fusion while the same was done with oligonucleotides 1322 (CTAAGGCTGTGCAAGGGAATTTG) —SEQ ID NO:179 and 1321 (AAGCTCTAGTGCAAAATGGTCTCCTGAATG) —GAGCAATAGAA GTGCAACCTGTGTCAACAATCCCTTAGCAGGTT)—SEQ ID NO:180 to form the 3' end. The two halves were ligated following restriction digestion with Bsu36I, to yield the full Klip27-MI coding sequence. PCR of this gene using primers 1364 (CTCGAGTCAAACTGTATGACACTGAAATC)—SEQ ID NO:181 and 1329 (AAGGCTCAATTCTTCTGTGGCAAXAGTTCACCAATG)—SEQ ID NO:183 attached a 5' Xhol restriction endonuclease cleavage site and a 3' HindIII cleavage site, respectively, for subsequent ligation into Xhol/HindIII-cut pSB84010. The result was plasmid pSB84405: a DNA sequence encoding the Oleosin-Klip8-Klip27-M1 fusion protein being placed in a binary vector under expression control of the phaseolin promoter/terminator. The phaseolin promoter controls the specific-temporal and tissue-specific expression of the transgene during seed development. The complete nucleic acid sequence SEQ ID NO:3 and amino acid sequence SEQ ID NO:4 of the 4405 insulin fusion protein (OLEO-Klip8-Klip27-MI) is shown in Fig. 2.
overlap and extended to form the 5' end of the Klip27-MI fusion while the same was done with oligonucleotides 1322 (CTAAGGCTGCTAAGGGAATG) SEQ ID NO:179 and 1321 (AAGCTCAGTGGCACAATTCTCCAT- TGTAAAGTAGGACAAATAAGGATGACACTAGT) SEQ ID NO:180 to form the 3' end. The two halves were ligated following restriction digestion with Bsu361, to yield the full Klip27-MI coding sequence. PCR of this gene fusion using primers 1363 (GCATGTCACCAACGATTGACACTAGT) SEQ ID NO:184 and 1334 (AAGCTCAGTGGCACAATTCTCCATTTG- GCAATACTTCTCCAATTG) SEQ ID NO:182 attached a 5' SplI restriction endonuclease cleavage site and the 3' Ddel DNA sequence plus HindIII cleavage site for subsequent ligation into SplI/HindIII-cut pSB4011. The result was plasmid pSB4402: a DNA sequence encoding the PRS-Klip27-MI-KDEL fusion protein being placed in a binary vector under expression control of the phaeo carrot/ terminator. The phaeo carrot promoter controls the specific-temporal and tissue-specific expression of the transgene during seed development. The plant expression vector, pSB4402, served as the template to introduce tetrasaccharide sites between the B and A chains of insulin and between M1 and D9 ScFv.

[01834] An intron-containing tetrasaccharide (B_{15},T-RRRK) site was placed between the authentic B_{15},T and A_{21},T chains of human insulin by PCR using primers 1515 (GCATGTCACCATTTGCATAGCATTGGATATCAACTGTTGGG) SEQ ID NO:185 and 1518 (ACAGTTCGGAACACATTTCTCCTTCTCTGC- TCTAGTGTAGGAGTGTTAAAACCTC) SEQ ID NO:186 using pSB4402 as the template. The resulting 124 bp fragment was used in combination with primer 1517 (GCAATACTTCTCCATTTGAGC) SEQ ID NO:187 using pSB4300 as a template. Note that pSB4300 contains a truncated D9 scFv-KDEL fragment with a HindIII restriction site. This PCR reaction produced a 555 bp product which introduced a tetrasaccharide (RRRK)-D9 ScFv-KDEL-HindIII onto the 124 bp SplI-M1 fragment. The 955 bp fragment was then ligated and subcloned in pGem-T (Promega) to result in pSB3403. The entire SplI-M1 (with B_{15},T-RRRK modified C propeptide)-RRRK-D9 ScFv-KDEL-HindIII fragment was inserted into precut SplI/Hin-dIII pSB4402 to generate pSB4441. The complete nucleic acid sequence SEQ ID NO:5 and amino acid sequence SEQ ID NO:6 of the 4414 insulin fusion protein (PRS-MI-tetrasac- inner-linker-D9Scf-KDEL) is shown in Fig. 3.

Transformation and Growth of Recombinant E. coli and Agrobacterium with pSB4404, pSB4405 and pSB4414.

[0184] After confirming the integrity of the cDNA encoding for the fusion protein by sequence analysis, the plasmids pSB4404, pSB4405 and pSB4414 were transformed into E. coli strain DH10a to allow for high levels of expression. Isolated plasmid DNA (100 ng) was mixed on ice with 100 μl of DH10a competent cells for 20 min. The cells were then heat shocked at 42°C for 45 seconds and returned to ice for 2 min. Then 1 ml of SOC media was added and the cells were incubated at 37°C on an endo-shaker at 225 rpm for 1 hr before placing transformed cells on LB-spectinomycin plates (10 μg/mL spectinomycin, 5 g/L yeast extract, 5 g/L NaCl, 15 g/L agar) and incubating overnight at 37°C. A single colony was used to inoculate 5 ml LB-spectinomycin broth. These cultures were grown overnight at 37°C. The recombinant plasmid was isolated from 1 ml of the overnight culture according to QiAprepSpin Miniprep Kit (Qiagen). The isolated plasmid was then used to transform competent Agrobacterium strain E1101 (Hood et al., 1986; J. Bacteriol. 144: 732-743) by electroporation (25 AF, 2.5 kV, 200 μF). Recombinant Agrobacterium were plated on AB-spectinomycin/kanamycin (20xAB salts, 2 M glucose, 0.25 mg/mL FeSO_4_7H_2O, 1 M MgSO_4, 1 M CaCl_2) and a single colony was used to inoculate 5 ml of AB-spectinomycin/kanamycin broth. These cultures were grown overnight at 28°C. The recombinant Agrobacterium were then used to transform Arabidopsis thaliana plants by the flower dipping method (Clough et al., 1998, Plant J., 16:755-743). Arabidopsis thaliana cv. (C24) is used for all the experiments. Seeds are planted on the surface of a soil mixture (two-thirds REDI-EARTH® professional growing mix and one-third perlite with a pH of 6.7) or an Arabidopsis soil mixture supplied by Lehle Seeds (perlite, peat, vermiculite, peat, terracotta, with a pH of 5.5) in 4 inch pots. The seedlings are allowed to grow to a rosette stage of 6-8 leaves to a diameter of approximately 2.5 cm. The pots are placed inside a dome at 4°C for four days for cold treatment and subsequently moved to 24°C growth room with constant light at about 150 μE and 60-70% relative humidity. The plants are irrigated at 2-3 day interval and fertilized weekly with 1% of Peters 20-19-18. Each pot contains five to six plants. When plants reach about 2 cm in height, the primary bolts are cut to encourage the growth of secondary and tertiary bolts. 4 to 5 days after cutting the primary bolts, the plants are ready to be infected with Agrobacterium. The plants with Arabidopsis plants are treated to allow the Arabidopsis plants being infected with 500 μl of a 1:10 suspension in an overnight Agrobacterium culture containing the plant transformation vector of interest for 20 seconds. It is important that the Agrobacterium culture contains 5% sucrose and 0.05% of the surfactant SILWET L77® surfactant (Lehle Seeds). The pots are subsequently covered with a transparent plastic dome for 24 hours to maintain higher humidity. The plants are allowed to grow to maturity and a mixture of seeds, untransformed and transformed, are harvested. For selection of transgenic lines, the putative transformed seeds are sterilized in a quick wash of 70% ethanol, then a 20% commercial bleach for 15 min and then rinsed at least four times with ddH_2O. About 1000 sterilized seeds are mixed with 0.6% melted top agar and evenly spread on a half strength MS plate (Murashige and Skoog, 1962, Physiologia Plantarum 15: 473-497) containing 0.3% sucrose and 80 μM of the herbicide phosphinotricin (PPT) DL. The plates are then placed in a growth room with light regime 8 hr dark and 16 hr light at 24°C. After 7 to 10 days, putative transgenic seedlings are green and growing whereas untransformed seedlings are bleached. After the establish- ment of roots the putative transgenic seedlings are individu- ally transferred to pots (the individually plants are irrigated in 3 day interval and fertilized with 1% Peters 20-19-18 in 7 day interval) and allowed to grow to maturity. The pots are covered with a transparent plastic dome for three days to protect the sensitive seedlings. After 7 days the seedlings are covered with a seed collector system from Lehle Seeds to prevent seed loss due to scattering. Seeds from these transgenic plants are harvested individually and ready for analysis.

Example 2

Expression Levels of Insulin in Arabidopsis thaliana

[0185] In the second example, expression levels of the fusion protein D9Scf-KLIP27-MI-KDEL (4404), OLEO- KLIP9-KLIP27-MI (4405) and PRS-MI-RRRK-D9Scf- KDEL (4414) were determined in transgenic Arabidopsis
thaliana mature seed. The transgene product was shown to be present in the cellular extracts of mature seed. Approximately 40 transgenic Arabidopsis thaliana seeds were ground with a mortar and pestle in 50 μl of 50 mM Tris-HCl pH 8.0. Then, a reducing SDS-PAGE sample buffer (6×SDS sample buffer, 0.35 M Tris-HCl pH 6.8, 30% glycerol, 10% SDS, 0.012% bromophenol blue, 5% mercaptoethanol) was added to the slurry and mixed by briefly vortexing. The sample was then briefly centrifuged and placed at 90°C for 10 minutes. After cooling on ice for 2 minutes the sample was centrifuged briefly. Samples were loaded (10 μl—equivalent to approximately 7 seeds) under reducing conditions.

[0186] For oil body prepared samples the transgenic and wild type seed (20 mg) were ground in 250 μl oil body extraction buffer (0.4 M sucrose, 0.5 M NaCl, 50 mM Tris-HCl pH 8.0). Samples were microfuged at 10000 g for 10 min. The soluble aqueous fraction was removed with a 26 G 1/2 ml syringe and the fat pad was re-suspended in 100 μl phosphate buffer supplemented with salt (20 mM Na₂HPO₄ pH 8.0, 0.5 M NaCl). The re-suspended fat pad was transferred to a clean microfuge tube and centrifuged again at 10000 g for 10 min. The procedure was repeated 3 more times with a final re-suspension of the fat pad in 100 μl phosphate buffer without salt (20 mM Na₂HPO₄ pH 8.0). An additional two more washes in phosphate buffer without salt were performed with intermittent centrifugation steps as outlined above. The final fat pellet was re-suspended in 10 μl phosphate buffer (20 mM Na₂HPO₄ pH 8.0). A 5 μl aliquot was taken and the oil body protein solubilized by boiling in 1/10 (v/v) 50 mM Tris-HCl pH 8.0 with 2% SDS. The sample was cooled on ice for 2 min and centrifuged at 10000 g for 5 min. The protein content of the undenatured were determined by BCA protein assay (Pierce, Rockford, Ill.). For coomassie-stained gels and Western blot analysis, 20 μg of total protein was separated on 15% SDS-PAGE gels under reducing conditions using the SDS-PAGE sample buffer.

[0187] The sample(s) were then loaded on discontinuous 15% SDS-PAGE gels and separated at 150 volts for approximately 1.5 hours. Gels were then either Coomassie-stained or blotted onto PVDF membrane (IMMIBOLON-PRO transfer membrane, Millipore Corporation, Bedford, Mass.) for Western blot analysis. Blotted samples were probed with monoclonal antibody directed against insulin (Clone E22E3; Roth et al., 1992) purchased from Abcam (Cambridge, UK). Insulin bands were detected using a secondary Sheep X mouse IgG F(ab)2 AP-conjugate (Chemicon International, Temecula, Calif.) and developed using NBT-BCIP in GARAP buffer (Tris-HCl pH 9.5, 100 mM NaCl, 5 mM Mg Cl₂). The immunoreactive band corresponded to a polypeptide band, migrating at the predicted molecular weight of the protein, as shown in Figs. 4A-4F. Figs. 4A-4F show recombinant expression of insulin fusion proteins in transformed Arabidopsis thaliana lines (4404-2, -17, -20, 4405-4, 4414-19 and 4414-20) on the basis of Coomassie-stained SDS-PAGE and Western Blot analysis. The arrows denote the position of the migrating 38.5 kDa, 34.2 kDa and 34.2 kDa fusion polypeptides, PRS-D9(scfv)-KLIP27-Mw/KDEL (4404), OLEO-KLIPr-KLIP27-MI (4405) and PRS-MI-RKRR-D9Scv-KDEL (4414) respectively, under reducing conditions. It should be noted 4414 fusion protein has an expected molecular weight of 34.2 kDa but has a higher apparent molecular weight on a SDS-PAGE gel. Figs. 4A (Coomassie-stained gel) and 4D (corresponding Western blot probed with anti-insulin E2E3) show total seed protein from wild type (wt) and transgenic seed lines expressing the 4404 and 4405 constructs. Figs. 4C (Coomassie-stained gel) and 4D (corresponding Western blot probed with anti-insulin E2E3) show oil body protein prepared from wild type and transgenic seed expressing the same 4404 and 4405 constructs. Figs. 4D (Coomassie-stained gel) and 4E (corresponding Western Blot probed with anti-insulin E2E3) show oil body protein prepared from wild type and transgenic seed expressing the same 4414 constructs. The molecular weight markers (M) are 10, 15, 20, 25, 37, 50, 75, 100, 150, 250 kDa. Controls include, hLN (recombinant human insulin standard) and hProLN (recombinant human proinsulin standard), separated under non-reducing conditions. Differences in expression levels are the result of clonal variation amongst transplants. The approximate protein levels of transgenic and MI expression are shown in Fig. 5. Levels of expression were determined using the 18 kDa oleosin band as an internal standard (equivalent to 1.5% total seed protein) by densitometry of the transgene band. The average level of expression for the PRS-D9(scfv)-KLIP27-Mw/KDEL (4404), OLEO-KLIPr-KLIP27-MI (4405) and PRS-MI-RKRR-D9Scv-KDEL (4414) constructs were 0.21% total seed protein, 0.12% total seed protein and 0.79% total seed protein respectively.

Example 3
Cleavage of pSBS4404 and HPLC Purification

[0188] Elution from Oil Body
[0189] In the third example, 1 g of transgenic seed was homogenized in 12 ml extraction buffer (0.4 M sucrose, 0.5 M NaCl, 50 mM Tris-HCl pH 8.0) and centrifuged at 10000 g for 10 min, the fat pads were removed and placed in 1 ml of 20 mM Na₂HPO₄, 0.5 M NaCl and re-centrifuged as above. This was repeated twice, before washing and centrifuging the fat pad twice in 750 μl 20 mM Na₂HPO₄. The 4404 fusion protein was eluted from the oil body into the undenatured by washing the final fat pad 5 times in 750 μl 20 mM formic acid pH 4.1, with 10000 g centrifugation steps in between each wash. The collected elution fractions (undenatured) were pooled and neutralized with 2 N NaOH to pH 8.0. The entire solution was then placed at ~80°C to freeze, and lyophilized overnight to concentrate the fusion protein. The lyophilized sample was re-suspended in 500 μl 50 mM Tris-HCl pH 8.0. The resuspended 4404 fusion protein was then desalted on a NAP-5 column (Amersham Pharmacia Biotech Ab, Upplands, Sweden) and re-exchanged with buffer (50 mM Tris-HCl pH 8.0). The desalted fraction was then frozen again and lyophilized overnight to concentrate. The final concentrated sample was re-suspended in a final volume of 105 μl double distilled H₂O. The elution results are presented in Fig. 6. Fig. 6 is a Coomassie-stained SDS-PAGE (15%) analysis of oil body prep prior to elution (--OB), OB prep after elution with formic acid (--OBF), and the concentrated eluted material (--E). The arrow denotes the position of the migrating fusion polypeptide. The wild type control is essentially free of any major proteins following elution whereas the concentrated 4404 material contains the fusion protein, some truncated products (possible hydrolyzed fusion protein) and possibly some albumins that co-eluted.

Cleavage and HPLC Analysis of 4404 Expressing Arabidopsis Seed
[0190] The concentrated sample was re-suspended in 105 μl of double distilled water and protein content was assessed by
BCA protein assay according to the manufacture (Pierce, Rockford, Ill., USA). Samples were then cleaved with trypsin (1:300 trypsin:total protein ratio, in 50 mM Tris-HCl pH 8.0, on ice for 90 min). The reactions were stopped with a 10 fold molar excess of TLCK (N-p-tosyl-L-lysine chloromethyl ketone). The entire reactions were then filtered through 0.2 μm filters (AERODISC® 13 mm Syringe filter with 0.2 μm SUPOR® membrane, Pall Corporation, Ann Arbor, Mich., USA) and analyzed by reversed phase (RP)—HPLC using a C18 column (ZORBAX® HPLC column 300SB-C18, Agilent Technologies, Waldbronn, Germany). Samples were loaded onto the column and eluted off at 1.0 ml/min using a 19-min linear gradient of 5-50% (v/v) acetonitrile in 0.1% (v/v) TFA. The chromatograph resulting from this analysis is seen in FIG. 7. The trace reveals a trypsin cleaved product from 4404 fusion protein, with nearly identical properties on the column as the human insulin standard (retention times of 17.011 min and 17.179 min, respectively). The HPLC fraction was collected from 17.0-17.5 min and analyzed by PSD MALDI/TOF mass spectrometry using a Voyager-DE STR mass spectrometer (Applied Biosystems). MS analysis was performed by the BioAnalytical Spectroscopy service provided through NRC-Plant Biotechnology Institute, Saskatoon, Saskatchewan, Canada. Resolution of the cleaved 4404 product purified by HPLC as described above is shown in FIG. 8B in comparison to human insulin standard shown in FIG. 8A. The observed mass of cleaved 4404 fusion protein with trypsin was 6191.51 Da. The discrepancy between human insulin standard (FIG. 8A) and the cleaved 4404 product (FIG. 8B) corresponds to a Des-B30 Insulin with the KDEL signal retained on the A-chain of the cleaved product (Des-B30 Insulin-KDEL).

**Example 4**

Cleavage of pBSB4405 and HPLC Purification

**Oil Body Preparation**

[0191] Fusion protein (OLEO-KLIP8-KLIP27-MI) can be partially purified by performing oil body preparations as described below. Approximately, 1 g of transgenic seed was homogenized in 12 ml extraction buffer (0.4 M sucrose, 0.5 M NaCl, 50 mM Tris-HCl pH 8.0) and centrifuged at 10000 g for 10 min, the fat pads were removed and placed in 1 ml of 50 mM Tris-HCl pH 8.0, 0.5 M NaCl and re-centrifuged as above. This was repeated twice, before washing and centrifuging the fat pad twice in 750 μl 50 mM Tris-HCl pH 8.0. The oil body preparation results in the removal of the majority of background proteins. The typical protein profile of an oil body preparation from transgenic Arabidopsis seed expressing the 4405 construct is demonstrated in FIG. 9.

Cleavage and HPLC Analysis of 4405 Expressing Arabidopsis Seed

[0192] The total protein content from resuspended oil bodies was assessed by solubilizing a fraction of the preparation (5 μl) diluted 10 times in 2% SDS, 50 mM Tris-HCl pH 8.0, boiled for 5 min and centrifuged for 3 min at 10000 g. Thereafter, the protein content was determined by BCA protein assay according to the manufacture (Pierce, Rockford, Ill., USA). Samples were then cleaved with trypsin (1:300 trypsin:total protein ratio, in 50 mM Tris-HCl pH 8.0, on ice for 90 min) to release the Klip27-MI fragment from the fusion protein. The reactions were stopped with a 10 fold molar excess of TLCK (N-p-tosyl-L-lysine chloromethyl ketone). Samples were centrifuged at 10000 g for 10 min and the undernatants of the entire reactions were then filtered through 0.2 μm filters (AERODISC® 13 mm Syringe filter with 0.2 μm SUPOR® membrane, Pall Corporation, Ann Arbor, Mich., USA). FIG. 9 depicts the Coomassie-stained SDS-PAGE (15%) analysis of total extractable seed protein and oil body (OB) prepared protein from lines expressing 4405 in comparison to wild type (nonrecombinant) seed. The arrow denotes the position of the migrating fusion polypeptide. The undernatants were further analyzed by reversed phase (RP)—HPLC using a C18 column (ZORBAX® HPLC column 300SB-C18, Agilent Technologies, Waldbronn, Germany). Samples were loaded onto the column and eluted off at 1.0 ml/min using a 19-min linear gradient of 5-50% (v/v) acetonitrile in 0.1% (v/v) TFA. The chromatograph resulting from this analysis is seen in FIG. 10. The trace reveals a trypsin cleaved product from 4405 fusion protein, with nearly identical properties on the column as the human insulin standard (retention times of 17.220 min and 17.179 min, respectively). The HPLC fraction was collected from 17.0-17.5 min and analyzed by PSD MALDI/TOF mass spectrometry using a Voyager-DE STR mass spectrometer (Applied Biosystems). MS analysis was performed by the BioAnalytical Spectroscopy service provided through NRC-Plant Biotechnology Institute, Saskatoon, Saskatchewan, Canada. As shown in FIG. 11, the observed mass of cleaved 4405 fusion protein with trypsin was 5706.30 Da. The discrepancy between human insulin standard (FIG. 8A) and the cleaved 4405 product (FIG. 11) correspond to a Des-B30 Insulin product (Des-B30 Insulin). The Des-B30 Insulin is the product expected from correct trypsin maturation of the 4405 fusion.

**Example 5**

Purification of Trypsin Cleaved MI Using AKTA Explorer (IFLC)

[0193] Purification of cleaved MI from 4405 was also partially purified from upscaled cleavage reactions by anion exchange (Mono Q FF 1 ml, Amersham Pharmacia) on an AKTA explorer (Amersham Pharmacia). Cleavage reactions were performed on 4405 oil body as described above prepared from up to 30 g transgenic seed. The undernatant from cleavage reactions were either filtered through 0.2 μm filters or concentrated by lyophilization on a Savant SPEEDVAC® concentrator. Sample reactions filtered could be applied to the column directly, but concentrated samples required the removal of salts to effectively bind the column. Concentrated samples could be desalted by passing cleaved material through a PD-10 column (Amersham Pharmade), by dialysis, or dilution to a salt concentration equivalent or less than 5 mS/cm. Desalted samples were equilibrated with 20 mM Tris-HCl pH 6.5. Samples were separated using a step gradient with NaCl of 0-40% NaCl with a 1 ml/min flow rate. Detection was performed at 214 nm (detection at 280 nm is relatively poor because of the low content of aromatic amino acids in insulin). Solvent A was 20 mM Tris-HCl pH 6.5 while solvent B was 20 mM Tris-HCl pH 6.5, 1 M NaCl. Fractions (1 ml) eluting at the same conductivity as Roche insulin standard, between 7-35 mS/cm, were collected (refer to FIG. 12). FIG. 12 shows the chromatogram of trypsin cleaved 4405 oil body preparations (dashed line) in comparison to human insulin standard (solid line). The presence of Insulin was verified in the collected fractions by HPLC, ELISA, or West-
ern analysis (data not shown). Samples collected were then concentrated by lyophilization and used in the insulin bioassay described in example 6.

Example 6
Insulin Tolerance Test: Bioassay in C57Bl/6 (B6) Male Mice

[0194] This bioassay was performed to determine the in vivo effect of recombinant plant-derived (Des-B8-IN) from trypsin cleaved 4405 in comparison to human insulin. Glucose plasma levels in B6 mice were determined prior to and following the intraperitoneal injection of insulin standards, negative controls, and SBS insulin. Fifteen male C57Bl/6 (B6) mice approximately 2 months of age were purchased from Jackson Laboratories (Bar Harbor, Me.). Plasma glucose levels were determined with an automatic glucometer (UltraOneTouch Ultra® blood glucose monitor, LifeScan, Johnson and Johnson). Positive controls included HUMULIN® insulin (Eli Lilly) and yeast recombinant human insulin standard from Roche. A saline solution served as the placebo. A negative control was included which represented trypsin cleaved oil bodies purified from wild type (non-recombinant) Arabidopsis seed that was processed identically to recombinant 4405 trypsin cleaved oil body preparations.

[0195] B6 mice were housed and fed ad libitum on a 12-hour dark-light cycle. For insulin tolerance tests mice were injected intraperitoneally (IP) with insulin (1 U/kg body weight) and glucose levels determined at 0, 15, 30, and 60 minutes using the automatic glucometer. All insulin tolerance tests were performed at the same interval each day (0:00 am). Insulin tolerance tests were performed with at least 2 days intervening between administering the next test. The results for the insulin tolerance tests are depicted in FIG. 13. The SBS DesB8 insulin derived from 4405 seed (closed diamonds) behaved almost identically (was statistically not different, p<0.05) to Huminulin R® (open squares) and Roche Insulin (open triangles) standards following injection over the course of the study. All insulin(s) tested significantly reduced plasma glucose levels (p<0.05) in comparison to saline placebo (open circles) and trypsin cleaved wild type Arabidopsis oil bodies (closed circles) (negative control).

Example 7
Construction of pSBS4401: PRS-Klpp27-M1-Fusion Protein

[0196] One of the fusion proteins studied began with the tobacco pathogen related sequence (PRS) (Stijmons et al., 1998, Bio/technology, 8:217-221) which served as the signal peptide to target expression to the ER in a co-translational manner. Immediately downstream was a trypsin cleavable pro-peptide (KLIP27) derived from the TA57 pro-peptide of yeast (Kjeldsen et al., 2001, Biotechnology and Genetic Engineering Reviews 18:89-121). This was followed by mini-insulin (M1), described by Kjeldsen et al. (2001).

[0197] The backbone of this plasmid pSBS4055, was based on the plant binary vector, pPZP200, described by Hajdukiewicz et al. (Plant Molecular Biology, 1994, 25:989-994). In place of the described multiple cloning site, a pat gene conferring host plant phosphinothricine resistance (Wohlleben et al., 1988, Gene 70:25-37) driven by the ubiquitin promoter/terminator from Petuniaeum crista (Kowallock et al., 1993, Plant Mol. Biol., 21:673-684), was inserted between the left and right border sequences. In addition to this cassette, the B-phaselins promoter/terminator from Phaseolus vulgaris (Slighton et al., 1983, Proc. Natl. Acad. Sci. USA 80:1897-1901) driving PRS was sub-cloned. Standard PCR (Horton et al., 1989, Gene 77:61-68) was used to fuse the synthetic PRS-encoding sequence with attached Sphl/HindIII restriction endonuclease sites to the 3’ end of the phaselins promoter to yield pSBS4011.

[0198] The Klpp27-M1 sequence was synthesized from four partially overlapping oligonucleotides which incorporated Arabidopsis thaliana codon usage to ensure the success of efficient translation in a plant-based expression system. Oligonucleotides 1524 (GAAAGAAGGAGGCT


nant Agrobacterium were then used to transform Arabidopsis thaliana plants by the flower dipping method (Clough et al., 1998, Plant J., 16:735-743) as described in Example 1.

Expression Levels of Insulin in Arabidopsis thaliana

[2000] Expression levels of the fusion protein KLIP27-MI (4401) was determined in transgenic Arabidopsis thaliana mature seed using the procedure outlined in Example 2 above. The transgene product was not found to be present in the cellular extracts of mature seed.

Example 8
Construction of pBSIS4409:OLEO-Human Proinsulin (OLEO-hP1N) Fusion Protein

[2001] This fusion protein began with the N-terminal 18 kDa oleosin from Arabidopsis thaliana followed in-frame by the gene encoding for human proinsulin (hP1N). The expression of this fusion protein was targeted to the nascent oil bodies formed during the development of the embryo.

[2002] The backbone of this plasmid, pBSIS4409, was based on the plant binary vector, pZLP200, described by Hajdukiewicz et al. (Plant Molecular Biology, 1994, 25:989-994). In place of the described multiple cloning site, a pat gene conferring host plant phosphonothricine resistance (Wohleben et al., 1988, Gene 70:25-31) driven by the ubiquitin promoter/terminator from Petroselinum crispum (Kawalleck et al., 1993, Plant. Mol. Bio., 21:673-684), was inserted between the left and right border sequences. In addition to this cassette, the β-phaseolin promoter/terminator from Phaseolus vulgaris (Slighston et al., 1983, Proc. Natl. Acad. Sci. USA 80:1997-1991) driving the Arabidopsis 18 kDa oleosin genomic sequence was sub-cloned. Standard PCR (Horton et al., 1989, Gene 77:61-68) was used to fuse the oleosin gene sequence (minus stop codon) with attached NcoI and HindIII restriction endonuclease sites to the 3′ end of the phaseolin promoter to yield pBSIS4408.

[2003] An NcoI-human pro-proinsulin gene-HindIII was synthesized as a single 335 bp piece by AptaGene using preferred plant codon usage. Subsequent ligation into NcoI/HindIII-cut pBSIS4408 resulted in the plasmid pBSIS4400: a DNA sequence encoding the Oleosin-human pre-proinsulin fusion protein being placed in a binary vector under expression control of the phaseolin promoter/terminator. The plasmids pBSIS4400 plasmid served as the template to generate human proinsulin (hP1N) by standard PCR using phi DNA polymerase with primers directed against the 5′ end (1457 oligo TTCTTG-GAACACACTTG—SEQ ID NO:190) and 3′ end (1458 oligo AAAGTTCAGTTAAGTATG—SEQ ID NO:191) including the HindIII site of the existing proinsulin region of the vector. A second fragment was amplified using phi DNA polymerase with a primer directed against the available Sphi site (oligo 1455 GCAATGAGTGTGCGAG—SEQ ID NO:192) to the 3′ end of the Arabidopsis oleosin gene (oligo 1456 GTATGTTGTTGGCCCA—SEQ ID NO:193) within the pBSIS4400 vector. Following PCR, products were separated on an agarose gel and bands corresponding to a 267 bp (hP1N-HindIII) and 360 bp (Sphi-OLEO(3′ end)) fragment were gel purified using a gel extraction kit (Qiagen). The two fragments were fused by a second round of PCR amplification using Taq DNA polymerase with primers 1455 (SEQ ID NO:192) and 1458 (SEQ ID NO:193) in combination with 0.001 μM of an overlapping bridging PCR primer (oligo 1459 GAGTGGCTGAGCACAACCACACTTTG—SEQ ID NO:194) for two cycles with an annealing temperature of 58°C followed by 31 cycles at 52°C in order to amplify a 627 bp Sphi-OLEO(3′ end)-hP1N-HindIII fragment. The 627 bp Sphi-OLEO(3′ end)-hP1N-HindIII fragment was then ligated into the T/A overhang of pGEMT Easy Vector System™ (Promega) and used to transform DH5a bacteria to result in pBSIS3409 (pGEMT-Sphi-OLEO(3′ end)-hP1N-HindIII).

[2004] The Sphi/HindIII fragment from pBSIS3409 was exchanged with the Sphi/HindIII fragment from pBSIS4400. Standard restriction digests on both pBSIS3409 and pBSIS4404 were performed using Sphi/HindIII (New England Biolabs). Fragments were separated on 1.5% agarose gels and purified using a gel extraction kit (Qiagen). The 617 bp Sphi/HindIII fragment liberated from pBSIS3409 was then ligated into the Sphi/HindIII acceptor site in pre-cut pBSIS4400 (internal Sphi/HindIII fragment removed) vector backbone using T4 DNA ligase (NEB) overnight at 15°C.

[2005] The resulting plasmid pBSIS4409: a DNA sequence (SEQ ID NO:195) encoding the OLEOSIN-hP1N fusion protein (SEQ ID NO:196) being placed in a binary vector under the expression control of the phaseolin promoter/terminator. The phaseolin promoter controls the specific-temporal and tissue-specific expression of the transgene during seed development.

Transformation and Growth of Recombinant E. coli and Agrobacterium with pBSIS4409.

[2006] After confirming the integrity of the cDNA encoding for the fusion protein by sequence analysis, the plasmids pBSIS4409 was transformed into E. coli strain DH5a to allow for high level of expression. Isolated plasmid DNA (100 ng) was mixed on ice with 100 μl of DH5a competent cells for 20 min. The cells were then heat shocked at 42°C for 45 seconds and returned to ice for 2 min. Then 1 ml of SOC media was added and the cells were incubated at 37°C on an incubator at 225 rpm for 1 hr prior to plating transformed cells on LB-spectinomycin plates (10 g/L, tryptone, 5 g/L yeast extract, 5 g/L NaCl, 15 g/l agar) and incubating overnight at 37°C. A single colony was used to inoculate 5 ml LB-spectinomycin broth. These cultures were grown overnight at 37°C. The recombinant plasmid was isolated from 1 ml of the overnight culture according to Qiagen mini prep. The isolated plasmid was then used to transform competent Agrobacterium strain EHA105 (Hood et al., 1986; J. Bacteriol. 144: 732-743) by electroporation (25 μF, 2.5 kV, 0.2Ω). Recombinant Agrobacterium were plated on AB-spectinomycin/ kanamycin (20μgAB salts, 2 M glucose, 0.25 mg/ml FeSO4, 7H2O, 1 M MgSO4, 1 M CaCl2) and a single colony was used to inoculate 5 ml of AB-spectinomycin/kanamycin broth. These cultures were grown overnight at 28°C. The recombinant Agrobacterium were then used to transform Arabidopsis thaliana plants by the flower dipping method (Clough et al., 1998, Plant J., 16:735-743) as described in Example 1.

Expression Levels of Insulin in Arabidopsis thaliana

[2007] Expression levels of the fusion protein OLEO-hP1N (4409) was determined in transgenic Arabidopsis thaliana mature seed using the procedure outlined in Example 2 above. Coomassie-stained gel of oil body proteins from two representative lines (4409-6 and 4409-8) comparing the migration of Oleosin-hP1N fusion protein (as denoted by the black arrow) to non-transformed (wt) Arabidopsis (FIG. 14). The level of expression was determined by densitometry to measure on average about 0.10% of total seed protein. This level was calculated above and beyond the co-migration of an endogenous protein of the same molecular weight in the
non-transformed seed (wt) which constituted approximately 0.04% of the total seed protein.

Example 9
Transformation of Safflower

[0208] This transformation protocol is similar to that outlined by Orlikowska T. K. et al. (1995) Plant Cell, Tissue and Organ Culture 40: 85-91, but with modifications and improvements both for transforming S-317 and for using phosphinothricin as the selectable marker. Decontaminate seeds from S-317 California variety of safflower, which are not damaged, cracked or diseased, in 0.1% HCl, for 12 minutes followed by 4-5 rinses with sterile distilled water. Germinate sterile seeds in the dark on MS medium (Murashige T. & Skoog F. (1962) Physiol. Plant. 15: 473-497) with 1% sucrose and 0.25% GELRITE® agar substitute. Initiate Agrobacterium cultures from frozen glycerol stocks in 5 ml AB minimal liquid media with antibiotic selection, and grow for 48 hours at 28°C. Grow an aliquot of this culture grown overnight in 5 ml of Luria broth with selection for transformation. Wash 6-8 ml of bacterial cells twice with AB media, and make up to a final cell density of 0.4-0.5 (OD600).

[0209] Remove two-day-old cotyledons from germinated seedlings, dip in the prepared Agrobacterium cells, and plate on MS medium with 3% sucrose, 4 μM N6-benzyladenine (BA) and 0.8 μM naphthaleneacetic acid (NAA). Incubate plates at 21°C under dark conditions. After 3 days, transfer the same medium with 300 mg/L timentin. After an additional 4 days, move all cultures to the light. After 3 days, place explants on selection medium with phosphinothricin added at 0.5 mg/L. For continued bud elongation, transfer explants weekly onto MS medium without phytohormones but with twice the basal amount of KNO₃. Excise shoots that had elongated to greater than 10 mm from the initial explant and individually grow on selection. For rooting, place green shoots, representing putative transgenic tissue, on MS medium with 2% sucrose, 10 μM indolebutyric acid and 0.5 μM NAA. Transfer rooted shoots to a well drained soil-less mix and grow under high humidity and 12 hours of light.

Example 10
Flax Transformation Protocol

[0210] This transformation procedure is similar to that outlined by Dong J. and McHughen A. (Plant Cell Reports (1991) 10:555-560), Dong J. and McHughen A. (Plant Sciences (1993) 88:61-71) and Mlynarova et al. (Plant Cell Reports (1994) 13: 282-285). Decontaminate flax seeds, which are not damaged, cracked or diseased, in a 70% ethanol solution for 5 to 7 minutes, followed by 25 minutes in a 50% bleach solution with Tween 20 (3-4 drops per 100 ml) with continuous stirring. Rinse seeds 5 to 7 times with sterile distilled water. Germinate decontaminated seeds in the light on MS medium (Murashige T. & Skoog F. (1962) Physiol. Plant. 15: 473-497) with 2% sucrose and 0.3% GELRITE® agar substitute in Magenta jars. For transformation, grow Agrobacterium cultures overnight in AB broth plus the appropriate antibiotic for selection. Wash to 6 ml of overnight cells twice, and re-suspended in 5 ml of AB broth; add 2 ml of this stock to 98 ml of induction medium (MS basal medium with 3% sucrose, 5 μM 6-benzylaminopurine (BA) and 0.25 μM alpha-naphthalene acetic acid (NAA) and adjust for a final OD₆₀₀ of 1.0.

[0211] Section hypocotyl explants, and inoculate in the prepared Agrobacterium cell solution for about 4 h (stir plates gently 1-2 times during this period). After the infection period, remove explants from the liquid inoculation medium and blot on sterile filter paper. Plate 15-20 explants on 0.7% agar-solidified induction medium in tissue culture plates. Seal the plates with plastic wrap, and co-cultivate explants for 48 h under light conditions (23-24°C). After 2 days, transfer the green, meristematic explants to the same medium containing 300 mg/L TIMENTIN® (tetracyclin sodium/potassium davalanate) (pre-selection media) and wrap with plastic wrap. After 3 days, transfer the cultures to the same medium containing 10 mg/L DL PPT (Selection 1). Wrap the plates with PARAFILM® sheet and incubate at 24°C under light conditions. Transfer cultures every two weeks and keep on this media for one month. For shoot elongation, transfer the cultures two weeks on selection medium II (MS basal medium containing 2% sucrose, 500 mg/L MES buffer, 300 mg/L TIMENTIN® (tetracyclin sodium/potassium davalanate) and 10 mg/L DL PPT) in Magenta jars. Putative transformed shoots, which survived selection, are dark green and form vigorous roots in 7-10 days when planted individually on selection II media. Transfer rooted shoots to sterilized greenhouse soil mix in small pots and cover plantlets with clear plastic cups for acclimatization. For maturation, transfer actively growing plants to one-gallon pots with a well-drained soil mix and grow under greenhouse conditions.

[0212] While the present invention has been described with reference to what are presently considered to be the preferred examples, it is to be understood that the invention is not limited to the disclosed examples. To the contrary, the invention is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.

[0213] All publications, patents and patent applications are herein incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

**TABLE 1**

<table>
<thead>
<tr>
<th>SEQ.ID</th>
<th>Intron Motif</th>
<th>Nucleic Acid Sequence Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>(P01308) Human preproinsulin (Comprises genes V00555, M10039, J00265, X70558, U15140, B005255 and A009655)</td>
<td>Neohuman Native Insulin</td>
</tr>
</tbody>
</table>
## TABLE 1-continued

Examples of Known Insulin Sequences

<table>
<thead>
<tr>
<th>SEQ. ID</th>
<th>Insulin Motif</th>
<th>(Amino Acid Sequence Identifier)</th>
<th>(Nucleic Acid Sequence Identifier)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammals</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>(AAB25818)</td>
<td>Preproinsulin C-peptide Equus prezewalskii (horses, zebras, rhinoceros, and tapir (Perissodactyla))</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>(P03130)</td>
<td>Preproinsulin Equus caballus (horse)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>(P03131)</td>
<td>Preproinsulin [Encompasses genes U30610, and M61153] Oryctolagus cuniculus (domestic rabbit)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>(P03121)</td>
<td>Insulin Balaenoptera physalus (finback whale)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>(P03134)</td>
<td>Insulin Balaenoptera borealis (sperm whale)</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>(P03135)</td>
<td>Preproinsulin [Comprises genes AF304555 and AY044828] Sus scrofa (pig) swine</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>(P03136)</td>
<td>Insulin Elephas maximus (Asian elephant)</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>(P03137)</td>
<td>Preproinsulin [Gene M54979] Bos taurus (cow)</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>(P03138)</td>
<td>Preproinsulin [Gene X00659] Ovis aries (sheep)</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>(P03120)</td>
<td>Insulin Camelus dromedarius (Arabian camel)</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>(P03121)</td>
<td>Preproinsulin [Gene X00179] Canis sp. (canine)</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>(P03128)</td>
<td>Insulin Hystrix cristata (crested porcupine)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>(P03064)</td>
<td>Preproinsulin [Gene 302989] Octopus bimaculatus (duoroucouli) owl monkey</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>(P03047)</td>
<td>Preproinsulin [Gene X81992] Cercohippus cebus (African green monkey)</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>(P03041)</td>
<td>Preproinsulin [Gene X61089] Pan troglodytes (chimpanzee)</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>(Q03972)</td>
<td>Insulin Orthorhynchus analectus (platypus)</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>(AAM76564)</td>
<td>Insulin [Gene AY092024] Pongo pygmaeus (orangutan)</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>(AAO69335)</td>
<td>preproinsulin [Gene: AE011815 containing AY137498, AY137599, and AY137599] Gorilla gorilla (gorilla)</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>(INMK8Q)</td>
<td>Insulin Satyrus cf. mcarinius (common squirrel monkey)</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>(P03131)</td>
<td>Preproinsulin [Gene M26328] Cricetus cricetus (long-tailed hamster)</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>(P03132)</td>
<td>Insulin 1 precursor [Encompasses genes V01242, V01242, and M25584] Rattus norvegicus (Norway rat)</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>(P03123)</td>
<td>Insulin 2 precursor [Encompasses genes V01243, X00748, M25583, and M25585] Rattus norvegicus (Norway rat)</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>(P03124)</td>
<td>Insulin Acomys cahirinus (Egyptian spiny mouse)</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>(P03125)</td>
<td>Insulin 1 precursor [Encompasses genes X04776, and AY07574] Mus musculus (house mouse)</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>(P03126)</td>
<td>Insulin 2 precursor [Gene X04726] Mus musculus (house mouse)</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>(P03127)</td>
<td>Insulin Chirogalea breviceps (chinchilla)</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>(P03129)</td>
<td>Preproinsulin [Encompasses genes K02233, and M11713] Cavia porcellus (domestic guinea pig)</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>(P03175)</td>
<td>Preproinsulin [Gene M57671] Otis tarda (dog)</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>(P18109)</td>
<td>Insulin Delturus virginianus (North American opossum)</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>(P21963)</td>
<td>Preproinsulin [Redeleus sp.]</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>(Q23587)</td>
<td>Preproinsulin [Gene X08241] Pammomys obesus (fur sand rat)</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>(Q01513)</td>
<td>Preproinsulin [Gene AY038064] Spermophilus tridecemlineatus (nut-brown lined ground squirrel)</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>(740063A)</td>
<td>Insulin C peptide Cavia porcellus (domestic guinea pig)</td>
<td></td>
</tr>
<tr>
<td>Birds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>(P03132)</td>
<td>Preproinsulin [Encompasses genes AE012454 (containing, J00872, J00873, and J00874), V00416, V00418, and X58093] Gallus gallus (chicken)</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>(P03133)</td>
<td>Preproinsulin Anas platyrhynchos (mallard duck)</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>(P07454)</td>
<td>Insulin Anser anser (western graylag goose)</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>(P21403)</td>
<td>Preproinsulin [Gene: AE006925 containing J06611, and X56612] Selasphorus rufus (rufous hummingbird)</td>
<td></td>
</tr>
<tr>
<td>Fish</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>(O73727)</td>
<td>Preproinsulin [Gene: AE036326] Danio rerio (zebrafish)</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>(P03135)</td>
<td>Preproinsulin [Gene: X00989] Cyprinus carpio (common carp)</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>(P03137)</td>
<td>Insulin Rutilus rutilus (common carp)</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>(P03139)</td>
<td>Insulin Rhinocoris plasii (lilien)</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>(P03140)</td>
<td>Insulin Katsuwonus pelamis (skjafjock tuna)</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>(P03141)</td>
<td>Preproinsulin [Gene: V00634] Lophius piscatorius (southern dolphin fish)</td>
<td></td>
</tr>
<tr>
<td>SEQ. ID</td>
<td>Insulin Motif</td>
<td>Examples of Known Insulin Sequences</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>-----------------------------------</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>(P04667)</td>
<td>Preproinsulin (Comprises genes: X00148, J00936, K01655, and X13559) Oecophorus ketu (chum salmon)</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>(P07453)</td>
<td>Myxococcales sp. (doddy sculpin)</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>(P09476)</td>
<td>Insulin Lepisosteus platida (aligator gar)</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>(P09477)</td>
<td>Insulin Platicthys flesus (European flounder)</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>(P09536)</td>
<td>Insulin Hydrobates colliei (southern flounder)</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>(P012704)</td>
<td>Insulin Spaulus acanthias (spiny dogfish)</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>(P012705)</td>
<td>Insulin Torpedo marmoratus (marbled electric ray)</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>(P012709)</td>
<td>Preproinsulin (Gene: U82395) Callirhynchus miniatus (elephant fish)</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>(P012806)</td>
<td>Insulin Petromyzon marinus (sea lamprey)</td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>(P023187)</td>
<td>Insulin Oncorhynchus gorbuscha (pink salmon)</td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>(P039075)</td>
<td>Insulin Anima varia (bowfin)</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>(P042633)</td>
<td>Insulin Anguilla rostrata (American eel)</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>(P080525)</td>
<td>Preproinsulin (Gene: AF038123) Oreochromis niloticus (Nil tilapia)</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>(P081423)</td>
<td>Insulin Euphias pacificus (sperm whale)</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>(P091881)</td>
<td>Insulin Piaractus mesopotamicus (Pacu)</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>(Q9PV7K)</td>
<td>Preproinsulin (Gene: ABO29318) Varanus moseri (barfin flounder)</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>(160326A4)</td>
<td>Insulin C Peptide Anguilla anguilla (European Eel)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Apodemus</em></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>(P012706)</td>
<td>Insulin 1 precursor (Gene: M24443) Xenopus laevis (African clawed frog)</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>(P012707)</td>
<td>Insulin 2 precursor (Gene: M24442) Xenopus laevis (African clawed frog)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Reptiles</em></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>(P012708)</td>
<td>Insulin Trachemys scripta (red-eared slider turtle)</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>(P012709)</td>
<td>Insulin Alligator mississippiensis (American alligator)</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>(P013045)</td>
<td>Insulin Beaufortia rhomboidea (snaker)</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>(P01334)</td>
<td>Insulin Crocodylus niloticus (western diamondback rattlesnake)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Human</em></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>(AAA72172)</td>
<td>Synthetic preproinsulin (Gene: J02547)</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>(AAA72106)</td>
<td>Synthetic insulin alpha chain 3' end (Gene: ABO03171 or M38610)</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>(AAA72917)</td>
<td>Synthetic insulin beta chain 3' end (Gene: ABO03171 or M38611)</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>(CAA00712)</td>
<td>Synthetic insulin (Gene: A07765)</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>(CAA00731)</td>
<td>Synthetic insulin (Gene: A07758)</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>(CAA00741)</td>
<td>Synthetic insulin (Gene: A07761)</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>(CAA00715)</td>
<td>Unrelated protein product (Gene: A07764)</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>(CAA00736)</td>
<td>Synthetic proinsulin (Gene: A08012) (EP357163-A)</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>(CAA00738)</td>
<td>Synthetic insulin (Gene: A06468) (EP:377156-A)</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>(CAA01561)</td>
<td>Modified insulin precursor (Gene: A21563) (WO:90011261)</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>(CAA01254)</td>
<td>Synthetic insulin (Gene: A15938) (EP:214826-A)</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>(CAA01799)</td>
<td>Asp(B1), Asp(B4), Asp(B10), Asp(B16), Glu(B27) insulin synthetic construct (Gene: A20317)</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>(CAA01799)</td>
<td>Glic(B9), Glic(A12) insulin precursor synthetic construct (Gene: A20314)</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>(CAA23424)</td>
<td>Synthetic proinsulin (Gene: V00082)</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>(CAA24707)</td>
<td>Synthetic insulin C chain (Gene: V01461)</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>(CAA25151)</td>
<td>Synthetic insulin B chain (Gene: V00062)</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>(CAE00306)</td>
<td>Unrelated synthetic protein product (Gene: A2573757) (Pat. WO 02/079250)</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>(Gene: M31026)</td>
<td>Synthetic human insulin B and mini-C chains using denaturated silica gel chromatography</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>(1BVB)</td>
<td>Chain B [d-Ala26]-Des(B27-B30)-Insulin-B26-Amine A Superpotent Single-Replacement Insulin Analogue</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>(1H1UA)</td>
<td>Chain A Insulin Mutant (B1, B10, B16, B27)glu, Des-B30, Nitr</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>(1H1UB)</td>
<td>Chain B Insulin Mutant (B1, B10, B16, B27)glu, Des-B30, Nitr</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>(1H1SA)</td>
<td>Chain A Human Insulin Mutant-Hir(B16)</td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>(1H1SB)</td>
<td>Chain B Human Insulin Mutant-Hir(B16)</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>(1ICAA)</td>
<td>Chain A Non-Standard Design Or Unstable Insulin Analogue</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>(1ICAB)</td>
<td>Chain B Non-Standard Design Or Unstable Insulin Analogue</td>
<td></td>
</tr>
</tbody>
</table>

With Enhanced Activity
<table>
<thead>
<tr>
<th>SEQ. NO.</th>
<th>Insulin Motif</th>
</tr>
</thead>
<tbody>
<tr>
<td>102</td>
<td>(JUCAC) Chain C Non-Standard Design Of Unstable Insulin Analogues With Enhanced Activity</td>
</tr>
<tr>
<td>103</td>
<td>(JUCAD) Chain D Non-Standard Design Of Unstable Insulin Analogues With Enhanced Activity</td>
</tr>
<tr>
<td>104</td>
<td>(JUT3A) Chain A Of An Unstable Insulin Analog With Native Activity</td>
</tr>
<tr>
<td>105</td>
<td>(JUT3B) Chain B Of An Unstable Insulin Analog With Native Activity</td>
</tr>
<tr>
<td>106</td>
<td>(JUT3C) Chain C Of An Unstable Insulin Analog With Native Activity</td>
</tr>
<tr>
<td>107</td>
<td>(JUT3D) Chain D Of An Unstable Insulin Analog With Native Activity</td>
</tr>
<tr>
<td>108</td>
<td>(JXMFA) Chain A Of Human Insulin Mutant Ile-A2-Allo-Ile, His-B10-Aasp, Pro-B28-Lys, Lys-B29-Pro</td>
</tr>
<tr>
<td>109</td>
<td>(JXMFB) Chain B Of Human Insulin Mutant Ile-A2-Allo-Ile, His-B10-Aasp, Pro-B28-Lys, Lys-B29-Pro</td>
</tr>
<tr>
<td>110</td>
<td>(JXMCA) Chain A Of Human Insulin Mutant Ile-A2-Alia, His-B10-Aasp, Pro-B28-Lys, Lys-B29-Pro</td>
</tr>
<tr>
<td>111</td>
<td>(JXMNB) Chain B Of Human Insulin Mutant Ile-A2-Alia, His-B10-Aasp, Pro-B28-Lys, Lys-B29-Pro</td>
</tr>
<tr>
<td>112</td>
<td>(JXWA) Chain A Allo-Ile2-Insulin, An Inactive Chiral Analogue</td>
</tr>
<tr>
<td>113</td>
<td>(JXWB) Chain B Allo-Ile2-Insulin, An Inactive Chiral Analogue</td>
</tr>
<tr>
<td>114</td>
<td>(JXWC) Chain C Allo-Ile2-Insulin, An Inactive Chiral Analogue</td>
</tr>
<tr>
<td>115</td>
<td>(JXWD) Chain D Allo-Ile2-Insulin, An Inactive Chiral Analogue</td>
</tr>
<tr>
<td>118</td>
<td>(JHHA) Chain A Of B9(Aasp) mutant</td>
</tr>
<tr>
<td>119</td>
<td>(JHIB) Chain B Of B9(Aasp) mutant</td>
</tr>
<tr>
<td>120</td>
<td>(JHIA) Chain A Of B25(phe) mutant</td>
</tr>
<tr>
<td>121</td>
<td>(JHIB) Chain B Of B25(phe) mutant</td>
</tr>
<tr>
<td>122</td>
<td>(WWW) Chain A, Human Insulin Two Disulfide Model</td>
</tr>
<tr>
<td>123</td>
<td>(WWW) Chain B, Human Insulin Two Disulfide Model</td>
</tr>
<tr>
<td>124</td>
<td>(WWW) Synthetic albebrin insulin [Gene: AY017185]</td>
</tr>
<tr>
<td>125</td>
<td>(WWW) Synthetic albebrin insulin [Gene: AY017184]</td>
</tr>
<tr>
<td>126</td>
<td>(WWW) N-terminal of an interleukinin 2-insulin fusion protein</td>
</tr>
<tr>
<td>127</td>
<td>(WWW) Beta-galactosidase-insulin fusion protein N-terminal</td>
</tr>
<tr>
<td>128</td>
<td>(PC7082) Epidermal growth factor/Des-B30 single-chain human insulin precursor fusion protein</td>
</tr>
<tr>
<td>129</td>
<td>(WWW) E. coli penicillinase/insulin I fusion protein 5' end [Gene: A10034998 or J02553]</td>
</tr>
<tr>
<td>130</td>
<td>(WWW) E. coli penicillinase/insulin I fusion protein 5' end [Gene: A10034998 or J02554]</td>
</tr>
<tr>
<td>131</td>
<td>(WWW) Rat insulin signal sequence/E. coli beta-galactosidase fusion protein [Gene: J02555]</td>
</tr>
<tr>
<td>132</td>
<td>(WWW) Simian virus 40 (SV40)/rat preproinsulin I fusion protein [Gene: J02559]</td>
</tr>
<tr>
<td>133</td>
<td>(IEFEA) Chain A, An Active Mini-Preinsulin, MC2p</td>
</tr>
<tr>
<td>134</td>
<td>(JCAA) Chain A, Non-Standard Design Of Unstable Insulin Analogues With Enhanced Activity</td>
</tr>
<tr>
<td>135</td>
<td>(JCAB) Chain B, Non-Standard Design Of Unstable Insulin Analogues With Enhanced Activity</td>
</tr>
<tr>
<td>136</td>
<td>(JCAM) Chain C, Non-Standard Design Of Unstable Insulin Analogues With Enhanced Activity</td>
</tr>
<tr>
<td>137</td>
<td>(JCAD) Chain D, Non-Standard Design Of Unstable Insulin Analogues With Enhanced Activity</td>
</tr>
<tr>
<td>138</td>
<td>(JCKC) Chain C, A Human Insulin Peptide-His-Dab Complex</td>
</tr>
<tr>
<td>139</td>
<td>(JU3A) Chain A, An Unstable Insulin Analog With Native Activity</td>
</tr>
<tr>
<td>140</td>
<td>(JU3B) Chain B, An Unstable Insulin Analog With Native Activity</td>
</tr>
<tr>
<td>141</td>
<td>(JU3C) Chain C, An Unstable Insulin Analog With Native Activity</td>
</tr>
<tr>
<td>142</td>
<td>(JU3D) Chain D, An Unstable Insulin Analog With Native Activity</td>
</tr>
<tr>
<td>143</td>
<td>(ISRA) Chain A, Mini-Preinsulin, Two Chain Insulin Analog Mutant: Des B30, His(B 10)asp, Pro(B 28)asp</td>
</tr>
</tbody>
</table>
### TABLE 1 continued

<table>
<thead>
<tr>
<th>SEQ ID NO:</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>144</td>
<td>(1547b) Chain B, Mini-Proinsulin, Two Chain Insulin Analog Mutant: Des B50, His(B10)ap, Pro(B28)ap. Single chain insulin analog mutant: Des B50, His(B10)ap, Pro(B28)ap and peptide bond between Lys B29 and Gly A1.</td>
</tr>
</tbody>
</table>

**SUMMARY OF SEQUENCES**

[0214] SEQ ID NO:1 and 2 set forth the nucleotide sequence and the deduced amino acid sequence, respectively, of the PRS-D9ScFv-KLIP27-MI-KDEL fusion protein in the plasmid pSBS4404.

[0215] SEQ ID NO:3 and 4 set forth the nucleotide sequence and the deduced amino acid sequence, respectively, of the Oleeo-KLIP28-KLIP27-MI fusion protein in plasmid pSBS4405.

[0216] SEQ ID NO:5 and 6 set forth the nucleotide sequence and the deduced amino acid sequence, respectively, of the PRS-MI-tetrasaccharide linker-D9ScFv-KDEL fusion protein in plasmid pSBS4414.

[0217] SEQ ID NO:7 to 145 set forth known insulin sequences which are described in Table 1.

[0218] SEQ ID NO:146 to 148 set forth the amino acid sequences of fragments of the insulin C-peptide.

[0219] SEQ ID NO:149 sets forth the amino acid sequence of the tetrasaccharide processing peptide.

[0220] SEQ ID NO:150 to 155 sets forth the amino acid sequence of the polypeptides capable of retaining the insulin polyepitope to the ER.

[0221] SEQ ID NO:156 to 160 sets forth the amino acid sequences of the polypeptides capable of retaining the insulin polyepitope to an ER derived storage organanelle.

[0222] SEQ ID NO:161 sets forth the amino acid sequence of a PRS signal sequence.

[0223] SEQ ID NO:162 to 171 set forth the amino acid sequences of yeast leader sequences and sequences derived therefrom.

[0224] SEQ ID NO:172 to 173 set forth the amino acid sequences of spacer peptides.

[0225] SEQ ID NO:174 sets forth the amino acid sequence of the KLIP28 sequence.

[0226] SEQ ID NO:175 sets forth the nucleotide sequence of the forward primer 1325 which is complementary to the 5' region of the D9ScFv cDNA and is designed to add a Splh site to the 3' region facilitate subsequent ligation.

[0227] SEQ ID NO:176 sets forth the nucleotide sequence of the reverse primer 1326 which is complementary to the 3' region of the D9ScFv cDNA and is designed to add a Xhol site to the 5' region facilitate subsequent ligation.

[0228] SEQ ID NO:177 sets forth the nucleotide sequence of the forward primer 1324 which is complementary to a 20 nucleotide region of reverse primer 1323 and is designed to form the 5' end of the Klip27-MI fusion.

[0229] SEQ ID NO:178 sets forth the nucleotide sequence of the reverse primer 1323 which is complementary to a 20 nucleotide region of forward primer 1324 and is designed to form the 5' end of the Klip27-MI fusion.

[0230] SEQ ID NO:179 sets forth the nucleotide sequence of the forward primer 1322 which is complementary to a 19 nucleotide region of reverse primer 1321 and is designed to form the 3' end of the Klip27-MI fusion.

[0231] SEQ ID NO:180 sets forth the nucleotide sequence of the reverse primer 1321 which is complementary to a 19 nucleotide region of forward primer 1322 and is designed to form the 3' end of the Klip27-MI fusion.

[0232] SEQ ID NO:181 sets forth the nucleotide sequence of the forward primer 1364 which is complementary to the 5' region of the Klip27-MI sequence and is designed to add a Xhol site to the 5' region facilitate subsequent ligation.

[0233] SEQ ID NO:182 sets forth the nucleotide sequence of the reverse primer 1334 which is complementary to the 3' region of the Klip27-MI sequence and is designed to add a HindIII site to the 3' region facilitate subsequent ligation and a 3' KDEL sequence.

[0234] SEQ ID NO:183 sets forth the nucleotide sequence of the reverse primer 1329 which is complementary to the 3' region of the Klip27-MI sequence and is designed to add a HindIII site to the 3' region facilitate subsequent ligation.

[0235] SEQ ID NO:184 sets forth the nucleotide sequence of the forward primer 1363 which is complementary to the 5' region of the Klip27-MI sequence and is designed to add a Splh site to the 5' region facilitate subsequent ligation.

[0236] SEQ ID NO:185 sets forth the nucleotide sequence of the forward primer 1515 which is complementary to the 5' region of the insulin B chain sequence and is designed to insert the intervening tetrasaccharide site between the authentic A and B chains of human insulin in conjunction with reverse primer 1518.

[0237] SEQ ID NO:186 sets forth the nucleotide sequence of the reverse primer 1518 which is complementary to the 3' region of the insulin B chain and 5' region of the insulin A chain with the intervening tetrasaccharide mini-c-peptide sequence and is designed to insert the intervening tetrasaccharide site between the authentic A and B chains of human insulin.

[0238] SEQ ID NO:187 sets forth the nucleotide sequence of the reverse primer 1517 which is complementary to the 3' region of the D9 ScFv/KDEL sequence and is designed to amplify the centre MI-tetrasaccharide linker-D9ScFv/KDEL to create the pSBS4414 insert.

[0239] SEQ ID NO:188 and 189 set forth the nucleotide sequence and the deduced amino acid sequence, respectively, of the PRS-Klipp27-MI fusion protein in plasmid pSBS4401.

[0240] SEQ ID NO:190 sets forth the nucleotide sequence of the forward primer 1457 which is complementary to the 5' region of the insulin B chain sequence and is designed to generate the human proinsulin (hRIN) fragment in conjunction with reverse primer 1591.
SEQ ID NO: 191 sets forth the nucleotide sequence of the reverse primer 1458 which is complementary to the 3' region of human proinsulin (hPIN) is designed to generate the human pro(hPIN) and add a 3' HindIII cloning site.

SEQ ID NO: 192 sets forth the nucleotide sequence of the forward primer 1455 which is complementary to the 5' region of the Sphl site of pSBS4404 and is designed amplify the Arabidopsis oleosin gene in conjunction with reverse primer 1456.

SEQ ID NO: 193 sets forth the nucleotide sequence of the reverse primer 1456 which is complementary to the 3' region of the Arabidopsis oleosin gene and is designed amplify the Arabidopsis oleosin gene in conjunction with forward primer 1455.

SEQ ID NO: 194 sets forth the nucleotide sequence of the overlapping bridging PCR primer which is complementary to the 3' region of the Arabidopsis oleosin gene and the 5' end of the human proinsulin gene and is designed to create the pSBS4409 insert in conjunction with forward primer 1455 and reverse primer 1456.

SEQ ID NO: 195 and 196 set forth the nucleotide sequence and the deduced amino acid sequence, respectively, of the OLEO-hPIN fusion protein in plasmid pSBS4409.

SEQUENCE LISTING

<160>  NUMBER OF SEQ ID NOS: 196

<210>  SEQ ID NO 1
<211>  LENGTH: 1143
<220>  ORGANISM: Artificial Sequence
<220>  FEATURE: OTHER INFORMATION: Insulin fusion protein nucleic acid sequence
<400> SEQ ID NO: 1
atgaacttcc ttaaggcttt cccctttttc gctttctttt gttctgcttc atcactctgt  60
gcgttaacgc atgctgacat tggagtgaca cacgccttcc cccctctgcg tgcgctgctg  120
ggagaacgag tcacactgag ccgacggtcg cgtgaagctc agtcagagcc ttttagagag taccatctca  180
aagactctct ttggactgta ccagcagaac cagaggacag ctctctactc ctggtgctac  240
tttgacatca ctggagaatgc tgggtctct gagctcttcag ctgagctgtg atctggagaac  300
gattcactct ttacatctgag cagttgacag acgtgaagac tggcagttca ctttgctcag  360
cacactata cacaccccctc caacggctgg gcttgagacac aagcaacgat taagggcct  420
cgaacggtgc ttggcttctg aggtcctgca ccagcagact acgtttgctgc tggtaacgc  480
gttcacaactgc agcagctgag acgtggctgtg atcgagctgtg gggcctcact gaaagatcac  540
taggcagttt ctgcagcata ctcaggtatg atgcttgatg aagatggcca  600
gaacgctgcc ttggccagat ttgagagatt ttcactggtca gttgagtcag taactacact  660
gagaagtcag acggccagcc cacattcact gcagatacac cttcgcaccc aagctcatac  720
cactcagaa gcgtcagactc tggagactt gccgctattt accttgcaac atggagattt  780
gactcctggt gcacggccag cacctccactc cttgtgagtt caccatttga tgccagcttg  840
tcggcaggt ccgctccagt gcgtcagttt cttgttagtt gacgagtttt gtctcacaacaa  900
acaatgtag ggtcatcttc gttgctgctc ttggtgctc tgggtactac agagtttttt  960
gactcctggt tttgtagttt caccatttgc ctggcagctc atgtctgctc tggcggtgag 1020
gttggttag gaaagttgtt ctattatcct cttgagcag tgaagagat ccgtgcagctc 1080
tggagaact ctatttggct acatttccaa ttggagaact atttgcaaca ggtgagactt 1140
tga 1143

<210>  SEQ ID NO 2
<211>  LENGTH: 380
<220>  ORGANISM: Artificial Sequence
<220>  FEATURE:
OTHER INFORMATION. Insulin fusion protein

SEQUENCE: 2

Met Asn Phe Leu Lys Ser Phe Pro Phe Tyr Ala Phe Leu Cys Phe Gly
1      5       10       15

Gln Tyr Phe Val Ala Val Thr His Ala Asp Ile Val Met Thr Gln Ser
20     25      30

Pro Ser Ser Leu Ala Met Ser Val Gly Gln Arg Val Thr Met Arg Cys
35     40      45

Lys Ser Ser Glu Ser Leu Leu Lys Ser Thr Asn Gln Lys Asn Tyr Leu
50     55      60

Ala Trp Tyr Gln Gln Lys Asp Gly Ser Pro Lys Leu Val Tyr
65     70      75      80

Phe Ala Ser Thr Arg Glu Ser Gly Val Pro Asp Arg Phe Ile Gly Ser
95     100     105

Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Val Gln Ala Gln
110    115     120     125

Asp Leu Ala Asp Tyr Phe Cys Gln Gin His Tyr Asn Thr Pro Pro Thr
130    135     140

Phe Gly Ala Gin Thr Lys Leu Glu Leu Lys Arg Ser Pro Asn Gly Ala
145    150     155     160

Ser His Ser Gly Ser Ala Pro Gly Thr Ser Ser Asa Ser Gly Ser Gin
170    175     180     185

Val His Leu Gin Ser Gly Ala Gin Leu Met Lys Gin Phe Ala Gin
190    195     200     205

Met Lys Ile Ser Cys Lys Ala Thr Gly Tyr Thr Phe Ser Ser Tyr Thr
210    215     220

Ile Glu Trp Val Lys Gin Arg Pro Gly His Gly Leu Glu Thr Ile Gly
225    230     235     240

Glu Ile Leu Pro Gly Ser Gly Ser Thr Tyr Asn Gln Lys Phe Lys
245    250     255     260

Gly Lys Ala Thr Phe Thr Ala Asp Thr Ser Ser Asn Thr Ala Tyr Met
270    275     280     285

Gln Leu Ser Ser Leu Thr Ser Asp Ser Ala Val Tyr Tyr Cys Ala
290    295     300     305

Arg Leu Asp Val Asp Ser Trp Gly Gin Gly Thr Thr Leu Thr Val Ser
310    315     320     325

Ser Gin Pro Ile Ser Asp Thr Glu Ser Gin Thr Ser Val Asn Leu
330    335     340     345

Met Ala Asp Thr Glu Ser Ala Thr Cys Gin Thr Thr Asn Ser Gly
350    355     360     365

Gly Leu Asp Val Gly Leu Ile Ser Met Ala Lys Arg Glu Glu Gly
370    375     380
-continued

<210> SEQ ID NO 3
<211> LENGTH: 945
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin fusion protein nucleic acid sequence

<400> SEQUENCE: 3

atggcgagata caaagtaggg aacccatcaca gatatcactg gcagagcaca gccggagag 60
atggccgacgccgccacca cacccagatg tccgagcgcg gatctgcgta ctccagctt 120
atggcagctt ccctaggtgc aactgtgtgc aagctggctg gccgctttctg ccagctttc 180
agctctaccct cagtcttcgg tggctagact tggactgttg cacaacttttt gctgtgttact 240
ttcagccggt cttctgcctgg ggtcttccgc acagttgctac gcttcatcag cgggctttctt 300
tctctggag ggttggcag tgcgcttata acgttttttt tctggatatt cggcaacgga 360
gagcaccacca agggcgccca cagcttgccag tgaatgctgg aagcaacaagct 420
caggtcgta aagacagacg tcgcttacac gcacagcgac aatccttggtg ggacacagac 480
gttgacgcta ctctggtggc ccgacacact acaatgtgcgc agatcaaccc caatctctctc 540
tacaaggtga aagatgctcag cgaaggctg aaggaacagt gatctcttga aagatcctgg 600
cagacacac ccagagacctcctcagc ccagcagccag ttccaaaccag ttcatagcgac tgaatcccgag 660
acagcttagc ttgaacctct acggcgatgt actgagacgc cgggttgata cacaacaaat 720
ttcgaggggt ttcggtgctg cgattgtgct ttgtaggtcta agagagagag caagacgcct 780
aaagttgtag attaactacat ctgcttggat cacttctttga aggtccttta cctgttgggt 840
ggagagagag gattttctta cactcttaag gcgtactagg aacaatgctg cagttgctg 900
aactctattt gtcaccttta ccaatgggag aacataggca actga 945

<210> SEQ ID NO 4
<211> LENGTH: 314
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin fusion protein

<400> SEQUENCE: 4

Met Ala Asp Thr Ala Arg Gly Thr His His Asp Ile Ile Gly Arg Arg Asp
1 5 10 15
Gln Tyr Pro Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Ser Gly
20 25 30
Arg Gly Ser Asp Tyr Ser Lys Ser Arg Gln Ile Ala Ala Ala Thr
35 40 45
Ala Val Thr Ala Gly Ser Leu Leu Val Leu Ser Leu Ser Leu Thr Leu
50 55 60
Val Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile
65 70 75 80
Phe Ser Pro Ile Leu Val Pro Ala Ile Thr Val Ala Leu Leu Ile
85 90 95
Thr Gly Phe Leu Ser Ser Gly Phe Gly Ile Ala Ile Thr Val
100 105 110
Phe Ser Trp Ile Tyr Ala Thr Gly Glu His Pro Gln Gly Ser Asp Lys
115 120 125
-continued

Leu 130
Asp Arg 145
Arg Arg 165
Arg Ile Pro 190
His Gly Leu 180
Ser Lys Phe 210
Asn Leu Met 225
Ser Gly Leu 245
Glu Gly Leu 260
Val Gly Ala 275
Pro Lys Ala 290
Ser Leu Tyr 305

<210> SEQ ID NO 5
<211> LENGTH: 1011
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: OTHER INFORMATION: Insulin fusion protein nucleic acid sequence

<400> SEQUENCE: 5

atgcatccc ttaagccttt tcccttctca gttttccttt gtttgcgctca ataccttcttt
60
gctgctcg gcgtgaatat gtaaactactctg tcttactgat tgggcctttc
tccctcggc ggtgagagag agaatctttcc taactcctct agactaag aagagagag
120
atgttaggca aagtgtgacat ctctttattg tcataatcaact aatggagaga
ttgactac gagaactcttg tctcttcactg catttttcct cattttaccctg
180
agaagaaga gagaatgctg gtagcagcag tcctttccct cctggtctat gtctctggga
cgagctgccga cagctcagtc catatgctctt taaaagcttt cattttatag
240
aacttctctt ccttgtaca ggcagatgct ctagaactct ttgtaacctt ggtatctgtt
300
gcataactca ggaatactcg ggtctcgtatt gctcctcatg gcagtgatct tcggagacat
360
ttactcttt ccatacgccaggtctcaggtt ctgactttgat cagattacct tctcagcaaa
cattattacc ctcctcctccag gtgtgcggtct gggacacagct ggtgctcctg
420
aacgctcgct ccatataggg tctttcactga ggcactatag ctcgatgtcgg atttcaggttg
480
cactgcgcag cagttggga tctgtctagt tacccgctgg cccatctgctt gatactccgt
540
aagatctcgt gcattctcct cctgatagtc ttgagtttaag ggtatagcag gaggctcctg
600
actgctgcgt gctgctggtg ctcctctcct cctgactact ctcagcactg
cctgcggctg ccagttgggt gatacttgg cccatcgctg
cctactgcct cctgctctct ggtgattgac ggtggtctgc
cgggtctcgct cgctgatcatt ggtatagtct gctgctctcg
660
aagctgtctc gattagctg cctgtctttc cccatgctctg ggtattaacag gaggctcctg
720
aacagctcg cattgactcg cttgactac gatagctctg ggttctctgc
cgtgctgctc ctcagtctcc gcctagtctg ggtatagtct
780
aagctgtctg gcctgtctgc ccagttgggt gatacttgg cccatcgctg
cctactgcct cctgctctct ggtgattgac ggtggtctgc
840
aacgctcgct ccatataggg tctttcactga ggcactatag ctcgatgtcgg atttcaggttg
900
-continued

ttcgacgac tgacatcga ggacgtgga gtcattact gttcaagatt ggtggtgac 960

tctggtgggg caagccacac tctcagcttg aagctcagagg atgagttg 1011

<210> SEQ ID NO 6
<211> LENGTH: 336
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin fusion protein

<400> SEQUENCE: 6

Met Asn Phe Leu Lys Ser Phe Pro Phe Tyr Ala Phe Leu Cys Phe Gly
1    5                      10                      15

Gln Tyr Phe Val Ala Val Thr His Ala Phe Val Asn Gln His Leu Cys
20   25                     30                      35

Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Gln Arg Gly
30   40                     45                      50

Phe Phe Tyr Thr Pro Lys Thr Arg Arg Gly Arg Ile Val Glu Gln
50   55                     60                      65

Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Gln Ala Tyr Cys Aas
65   70                     75                      80

Arg Arg Lys Arg Asp Ile Val Met Thr Gln Ser Pro Ser Ser Leu Ala
85   90                     95                      100

Met Ser Val Gly Gln Arg Val Thr Met Arg Cys Lys Ser Ser Gln Ser
100 105                     110                     115

Leu Leu Lys Ser Thr Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln
115 120                     125                     130

Lys Pro Gly Gln Ser Pro Lys Leu Leu Val Tyr Phe Ala Ser Thr Arg
130 135                     140                     145

Glu Ser Gly Val Pro Asp Arg Phe Ile Gly Ser Gly Ser Gln Thr Asp
145 150                     155                     160

Phe Thr Leu Thr Ile Ser Ser Val Gln Ala Gln Met Ala Gln Thr
165 170                     175                     180

Phe Cys Gin Gin His Tyr Asn Thr Thr Pro Thr Phe Gly Ala Gln Thr
185 190                     195                     200

Lys Leu Gln Leu Lys Arg Ser Pro Asn Gly Ala Ser His Ser Gly Ser
200 205                     210                     215

Ala Pro Gly Thr Ser Ser Ala Gln Ser Gln Val His Leu Gin Gin
215 220                     225                     230

Ser Gly Ala Gln Ser Met Lys Pro Gly Ala Ser Met Lys Ile Ser Cys
230 235                     240                     245

Lys Ala Thr Gln Tyr Thr Phe Ser Ser Tyr Thr Phe Ser Thr Met Lys
245 250                     255                     260

Gln Arg Pro Gly His Gly Leu Gln Thr Thr Ile Gly Ile Leu Gln Pro Gly
260 265                     270                     275

Ser Gly Ser Thr Tyr Asn Gln Lys Phe Lys Gin Ala Thr Phe
275 280                     285                     290

Thr Ala Asp Thr Ser Ser Thr Val Tyr Cys Ala Arg Leu Asp Val Asp
290 295                     300                     305

Thr Ser Gly Ser Ala Val Tyr Cys Ala Arg Leu Asp Val Asp
305 310                     315                     320

Ser Thr Gln Gin Gin Thr Thr Val Ser Ser Lys Asp Gin Thr
320 325                     330                     335
---continued---

<210> SEQ ID NO 7
<211> LENGTH: 110
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 7

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Leu Ala Leu 1 5 10 15
Trp Gly Pro Asp Pro Ala Ala Ala Phe Val Asn Gln His Leu Cys Gly 20 25 30
Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe 35 40 45
Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Glu Gin Val Gly 50 55 60
Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gin Pro Leu 65 70 75 80
Ala Leu Glu Gly Ser Leu Glu Lys Arg Gly Ile Val Glu Gin Cys Cys 95 96
Thr Ser Ile Cys Ser Leu Tyr Gin Leu Glu Asn Tyr Cys Asn 100 105 110

<210> SEQ ID NO 8
<211> LENGTH: 31
<212> TYPE: PRT
<213> ORGANISM: Equus przewalskii

<400> SEQUENCE: 8

Glu Ala Glu Asp Pro Gln Val Gly Glu Val Glu Leu Gly Gly Gly Pro 1 5 10 15
Gly Leu Gly Gly Leu Gin Pro Leu Ala Leu Ala Gly Pro Gin Gin 20 25 30

<210> SEQ ID NO 9
<211> LENGTH: 86
<212> TYPE: PRT
<213> ORGANISM: Equus caballus

<400> SEQUENCE: 9

Phe Val Asn Gin His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr 1 5 10 15
Leu Val Cys Gly Arg Gly Phe Phe Tyr Thr Pro Lys Ala Xaa Xaa 20 25 30
Glu Ala Glu Asp Pro Gln Val Gly Glu Val Glu Leu Gly Gly Gly Pro 35 40 45
Gly Leu Gly Gly Leu Gin Pro Leu Ala Leu Ala Gly Pro Gin Gin Xaa 50 55 60
Xaa Gly Ile Val Glu Gin Cys Cys Thr Gly Ile Cys Ser Leu Tyr Gin 65 70 75 80
Leu Glu Asn Tyr Cys Asn
<210> SEQ ID NO 10
<211> LENGTH: 110
<212> TYPE: PRT
<213> ORGANISM: Oryctolagus cuniculus

<400> SEQUENCE: 10
Met Ala Ser Leu Ala Ala Leu Leu Pro Leu Leu Ala Leu Leu Val Leu 1 5 10 15
Cys Arg Leu Arg Pro Ala Gin Ala Phe Val Asn Gin His Leu Cys Gly 20 25 30
Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe 35 40 45
Phe Tyr Thr Pro Lys Ser Arg Arg Gin Val Glu Gin Val Gly 50 55 60
Gln Ala Gin Leu Gly Gly Gin Pro Gly Ala Gin Gly Gin Gin Gin Ser 65 70 75 80
Ala Leu Gin Leu Leu Ala Gin Lys Arg Gin Gin Gin Gin Gin Gin Gin 85 90 95 100
Thr Ser Ile Cys Ser Leu Tyr Gin Leu Glu Gin Tyr Cys Asn 105 110

---continued---
-continued

Phe Tyr Thr Pro Lys Ala Arg Arg Glu Ala Glu Asn Pro Gin Ala Gly
50 55 60
Ala Val Glu Leu Gly Gly Leu Gly Gly Leu Gin Ala Leu Ala Leu
65 70 75 80
Glu Gly Pro Pro Gin Lys Arg Gly Ile Val Glu Gin Cys Thr Ser
90 90 95
Ile Cys Ser Leu Tyr Gin Leu Glu Asn Tyr Cys Asn
100 105

SEQ ID NO: 14
LENGTH: 21
TYPE: PRT
ORGANISM: Elephas maximus

SEQUENCE: 14
Gly Ile Val Glu Gin Cys Thr Gly Val Cys Ser Leu Tyr Gin Leu
1 5 10 15
Glu Asn Tyr Cys Asn
20

SEQ ID NO: 15
LENGTH: 105
TYPE: PRT
ORGANISM: Bos taurus

SEQUENCE: 15
Met Ala Leu Trp Thr Arg Leu Arg Pro Leu Leu Ala Leu Ala Leu
1 5 10 15
Trp Pro Pro Pro Pro Ala Arg Ala Phe Val Asn Gin His Leu Cys Gly
20 25 30
Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45
Phe Tyr Thr Pro Lys Ala Arg Arg Glu Val Glu Gly Pro Gin Val Gly
50 55 60
Ala Leu Glu Leu Ala Gly Gly Pro Gly Ala Gly Leu Glu Gly Pro
65 70 75 80
Pro Gin Lys Arg Gly Ile Val Glu Gin Cys Ala Ser Val Cys Ser
85 90 95
Leu Tyr Gin Leu Glu Asn Tyr Cys Asn
100 105

SEQ ID NO: 16
LENGTH: 105
TYPE: PRT
ORGANISM: Ovis aries

SEQUENCE: 16
Met Ala Leu Trp Thr Arg Leu Val Leu Ala Leu Ala Leu
1 5 10 15
Trp Ala Pro Ala Pro Ala His Ala Phe Val Asn Gin His Leu Cys Gly
20 25 30
Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45
Phe Tyr Thr Pro Lys Ala Arg Arg Glu Val Glu Gly Pro Gin Val Gly
50 55 60
Ala Leu Glu Leu Ala Gly Gly Pro Gly Ala Gly Leu Glu Gly Pro
65 70 75 80
Pro Gln Lys Arg Gly Ile Val Glu Gln Cys Cys Ala Gly Val Cys Ser
89 90
Leu Tyr Gin Leu Glu Asn Tyr Cys Asn
100 105

<210> SEQ ID NO 17
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Camelus dromedaries
<400> SEQUENCE: 17
Gly Ile Val Glu Gln Cys Ala Ser Val Cys Ser Leu Tyr Gin Leu
1 5 10 15
Glu Asn Tyr Cys Asn
20

<210> SEQ ID NO 18
<211> LENGTH: 110
<212> TYPE: PRT
<213> ORGANISM: Canis sp.
<400> SEQUENCE: 18
Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15
Trp Ala Pro Ala Pro Thr Arg Ala Phe Val Asn Gin His Leu Cys Gly
20 25 30
Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45
Phe Tyr Thr Pro Lys Ala Arg Arg Glu Val Glu Asp Leu Gin Val Arg
50 55 60
Asp Val Glu Leu Ala Gly Ala Pro Gly Glu Gly Leu Gin Pro Leu
65 70 75 80
Ala Leu Glu Gly Ala Leu Gin Lys Arg Gly Ile Val Glu Gin Cys Cys
85 90 95
Thr Ser Ile Cys Ser Leu Tyr Gin Leu Glu Asn Tyr Cys Asn
100 105 110

<210> SEQ ID NO 19
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Hystrix cristata
<400> SEQUENCE: 19
Gly Ile Val Asp Gin Cys Thr Gly Val Cys Ser Leu Tyr Gin Leu
1 5 10 15
Gln Asn Tyr Cys Asn
20

<210> SEQ ID NO 20
<211> LENGTH: 108
<212> TYPE: PRT
<213> ORGANISM: Aotus trivirgatus
<400> SEQUENCE: 20
Met Ala Leu Trp Met His Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15
-continued

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trp Gly Pro Glu Pro Ala Pro Ala Phe Val Am Glu His Leu Cys Gly</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>Pro His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>Phe Tyr Ala Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln Val Gly</td>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>Gln Val Glu Leu Gly Gly Ser Ile Thr Gly Ser Leu Pro Pro Leu</td>
<td>69</td>
<td>75</td>
</tr>
<tr>
<td>Glu Gly Pro Met Gln Lys Arg Gly Val Val Asp Gln Cys Cys Thr Ser</td>
<td>85</td>
<td>90</td>
</tr>
<tr>
<td>Ile Cys Ser Leu Tyr Glu Leu Glu Amn Tyr Cys Amn</td>
<td>100</td>
<td>105</td>
</tr>
</tbody>
</table>

SEQ ID NO: 21
LENGTH: 110
TYPE: PRF
ORGANISM: Macaca fascicularis

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Ala Leu Trp Met Arg Leu Pro Leu Leu Ala Leu Ala Leu Ala Leu</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Trp Gly Pro Asp Pro Ala Pro Ala Phe Val Am Glu His Leu Cys Gly</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Pro Gln Val Gly</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu</td>
<td>65</td>
<td>70</td>
</tr>
<tr>
<td>Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys</td>
<td>85</td>
<td>90</td>
</tr>
<tr>
<td>Thr Ser Ile Cys Ser Leu Tyr Glu Leu Glu Amn Tyr Cys Amn</td>
<td>100</td>
<td>105</td>
</tr>
</tbody>
</table>

SEQ ID NO: 22
LENGTH: 110
TYPE: PRF
ORGANISM: Cercoptethus aethiops

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Ala Leu Trp Met Arg Leu Pro Leu Leu Ala Leu Ala Leu Ala Leu</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Trp Gly Pro Asp Pro Ala Pro Ala Phe Val Am Glu His Leu Cys Gly</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Pro Gln Val Gly</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu</td>
<td>65</td>
<td>70</td>
</tr>
<tr>
<td>Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys</td>
<td>85</td>
<td>90</td>
</tr>
<tr>
<td>Thr Ser Ile Cys Ser Leu Tyr Glu Leu Glu Amn Tyr Cys Amn</td>
<td>100</td>
<td>105</td>
</tr>
</tbody>
</table>
<210> SEQ ID NO 23
<211> LENGTH: 110
<212> TYPE: PRT
<213> ORGANISM: Pan troglodytes

<400> SEQUENCE: 23

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Val Leu Leu Ala Leu
1 5 10 15
Trp Gly Pro Asp Pro Ala Ser Ala Phe Val Asn Glu His Leu Cys Glu
20 25 30
Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45
Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln Val Gly
50 55 60
Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu
65 70 75 80
Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95
Thr Ser Ile Cys Ser Leu Tyr Gin Leu Glu Asn Tyr Cys Asn
100 105 110

<210> SEQ ID NO 24
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Ornithorhynchus anatinus

<400> SEQUENCE: 24

Gly Ile Val Glu Glu Cys Lys Gly Val Cys Ser Met Tyr Gin Leu
1 5 10 15
Glu Asn Tyr Cys Asn
20

<210> SEQ ID NO 25
<211> LENGTH: 65
<212> TYPE: PRT
<213> ORGANISM: Pongo pygmaeus

<400> SEQUENCE: 25

Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg
1 5 10 16
Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln
20 25 30
Val Gly Gin Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln
35 40 45
Pro Leu Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln
50 55 60
Cys
65

<210> SEQ ID NO 26
<211> LENGTH: 110
<212> TYPE: PRT
<213> ORGANISM: Gorilla gorilla

<400> SEQUENCE: 26

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15
<table>
<thead>
<tr>
<th>Residue</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trp</td>
<td>Gly</td>
<td>Pro</td>
<td>Asp</td>
<td>Pro</td>
<td>Ala</td>
<td>Ala</td>
<td>Phe</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;210&gt; SEQ ID NO 27 &lt;211&gt; LENGTH: 51 &lt;212&gt; TYPE: PRT &lt;213&gt; ORGANISM: Saimiri sciureus &lt;400&gt; SEQUENCE: 27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
-continued

Met Ala Leu Trp Met Arg Phe Leu Pro Leu Leu Ala Leu Leu Leu Val Leu 1 5 10 15
Trp Glu Pro Lys Pro Ala Gln Ala Phe Val Lys Gln His Leu Cys Gly 20 25 30
Pro His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe 35 40 45
Phe Tyr Thr Pro Lys Ser Arg Arg Glu Val Glu Asp Pro Glu Val Pro 50 55 60
Gln Leu Glu Leu Gly Gly Gly Pro Gln Ala Gln Gly Asp Leu Gln Thr Leu 65 70 75 80
Ala Leu Glu Val Ala Arg Gln Gly Ile Val Asp Gln Cys Cys 85 90 95
Thr Ser Ile Cys Ser Leu Tyr Glu Leu Glu Asn Tyr Cys Asn 100 105 110

<210> SEQ ID NO 30
<211> LENGTH: 110
<212> TYPE: PRT
<213> ORGANISM: Rattus norvegicus

<400> SEQUENCE: 30
Met Ala Leu Trp Ile Arg Phe Leu Pro Leu Leu Ala Leu Leu Ile Leu 1 5 10 15
Trp Glu Pro Arg Pro Ala Gln Ala Phe Val Lys Gln His Leu Cys Gly 20 25 30
Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe 35 40 45
Phe Tyr Thr Pro Met Ser Arg Arg Glu Val Glu Asp Pro Glu Val Ala 50 55 60
Gln Leu Glu Leu Gly Gly Gly Pro Gln Ala Gln Gly Asp Leu Gln Thr Leu 65 70 75 80
Ala Leu Glu Val Ala Arg Gln Gly Ile Val Asp Gln Cys Cys 85 90 95
Thr Ser Ile Cys Ser Leu Tyr Glu Leu Glu Asn Tyr Cys Asn 100 105 110

<210> SEQ ID NO 31
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Acomys cahirinus

<400> SEQUENCE: 31
Gly Ile Val Asp Gln Cys Thr Ser Ile Cys Ser Leu Tyr Glu Leu 1 5 10 15
Glu Asn Tyr Cys Asn 20

<210> SEQ ID NO 32
<211> LENGTH: 108
<212> TYPE: PRT
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 32
Met Ala Leu Leu Val His Phe Leu Pro Leu Leu Ala Leu Leu Leu Ala Leu 1 5 10 15
Trp Glu Pro Lys Pro Thr Gln Ala Phe Val Lys Gln His Leu Cys Gly
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro</td>
<td>His</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td>Pro</td>
<td>His</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Glu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Val</td>
<td>Cys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Cys</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Arg</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Tyr</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Lys</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Arg</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Glu</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Arg</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Pro</td>
<td>Gln</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Leu</td>
<td>Cys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>Cys</td>
<td>Cys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>Asn</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Cys</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Leu</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Cys</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Gly</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Gln</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Phe</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Pro</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Met</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Phe</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Pro</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Phe</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Trp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trp</td>
<td>Gln</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Phe</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Phe</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td>His</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>His</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Tyr</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Pro</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Val</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Val</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Pro</td>
<td>Gln</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Ala</td>
<td>Gln</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Gln</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Gly</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Cys</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Ser</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Cys</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Gln</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Tyr</td>
<td>Cys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>Asn</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Cys</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Leu</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
</tbody>
</table>
-continued

Phe Tyr Ile Pro Lys Asp Arg Arg Glu Leu Glu Asp Pro Gin Val Glu
50 55 60

Gln Thr Glu Leu Gly Met Gly Leu Gly Ala Gly Gly Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Met Ala Leu Gin Lys Arg Gly Ile Val Asp Gin Cys Cys
85 90 95

Thr Gly Thr Cys Thr Arg His Gin Leu Gin Ser Tyr Cys Asn
100 105

110

<210> SEQ ID NO 36
<211> LENGTH: 109
<212> TYPE: PRT
<213> ORGANISM: Octodon degus

<400> SEQUENCE: 36

Met Ala Pro Trp Met His Leu Leu Thr Val Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asn Ser Val Gin Ala Tyr Ser Ser Gin His Leu Cys Gly
20 25 30

Ser Asn Leu Val Glu Ala Leu Tyr Met Thr Cys Gly Arg Ser Gly Phe
35 40 45

Tyr Arg Pro His Asp Arg Arg Glu Leu Glu Asp Leu Gln Val Glu Gin
50 55 60

Ala Glu Leu Gly Leu Ala Gly Leu Gln Pro Ser Ala Leu Glu
65 70 75 80

Met Ile Leu Gin Lys Arg Gly Ile Val Asp Gin Cys Cys Asn Asn Ile
85 90 95

Cys Thr Phe Asn Gin Leu Gin Asn Tyr Cys Asn Val Pro
100 105

<210> SEQ ID NO 37
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Didelphis virginiana

<400> SEQUENCE: 37

Gly Ile Val Glu Gin Cys Cys Asn Ser Ile Cys Ser Leu Tyr Gin Leu
1 5 10 15

Glu Thr Tyr Cys Asn
20

<210> SEQ ID NO 38
<211> LENGTH: 108
<212> TYPE: PRT
<213> ORGANISM: Rodentia sp.

<400> SEQUENCE: 38

Met Ala Leu Trp Ile Leu Leu Pro Leu Leu Ala Leu Leu Ile Leu Trp
1 5 10 15

Gly Pro Asp Pro Ala Gin Ala Phe Val Asn Gin His Leu Cys Gly Ser
20 25 30

His Leu Val Glu Ala Leu Tyr Ile Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Met Ser Arg Arg Glu Val Glu Asp Pro Gin Val Gly
50 55 60
-continued

Gln Val Glu Leu Gly Ala Gly Pro Gly Ala Gly Ser Glu Gin Thr Leu
65 70 75 80

Ala Leu Glu Val Ala Arg Gin Ala Arg Ile Val Gin Gin Cys Thr Ser
95 90 95

Gly Ile Cys Ser Leu Tyr Gin Glu Asn Tyr Cys Asn
100 105

<210> SEQ ID NO: 39
<211> LENGTH: 110
<212> TYPE: PRT
<213> ORGANISM: Psammomys obesus

<400> SEQUENCE: 39

Met Ala Leu Thr Met Arg Leu Leu Pro Leu Leu Ala Phe Leu Ile Leu
1 5 10 15

Trp Glu Pro Ser Pro Ala His Ala Phe Val Asn Gin His Leu Cys Gly
20 25 30

His Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Phe Arg Arg Gly Val Asp Asp Pro Gin Met Pro
50 55 60

Gln Leu Glu Leu Gly Gln Ser Pro Gly Ala Gly Asp Leu Arg Ala Leu
65 70 75 80

Ala Leu Glu Val Ala Arg Gin Lys Arg Gly Ile Val Glu Gin Cys Cys
95 90 95

Thr Gly Ile Cys Ser Leu Tyr Gin Leu Glu Asn Tyr Cys Asn
100 105 110

<210> SEQ ID NO: 40
<211> LENGTH: 110
<212> TYPE: PRT
<213> ORGANISM: Spermophilus tridecemlineatus

<400> SEQUENCE: 40

Met Ala Leu Thr Arg Leu Leu Pro Leu Leu Ala Leu Ala Leu
1 5 10 15

Leu Gly Pro Asp Pro Ala Gin Ala Phe Val Asn Gin His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ser Arg Arg Glu Val Glu Glu Gin Gly Gly
50 55 60

Gln Val Glu Leu Gly Gln Ser Pro Gly Ala Gly Leu Pro Gin Pro Leu
65 70 75 80

Ala Leu Glu Met Ala Leu Gin Lys Arg Gly Ile Val Glu Gin Cys Cys
95 90 95

Thr Ser Ile Cys Ser Leu Tyr Gin Leu Glu Asn Tyr Cys Asn
100 105 110

<210> SEQ ID NO: 41
<211> LENGTH: 29
<212> TYPE: PRT
<213> ORGANISM: Cavia porcellus

<400> SEQUENCE: 41

Glu Leu Glu Asp Pro Gin Val Glu Gin Thr Glu Leu Gly Met Gly Leu
-continued

1 10 15
Gly Ala Gly Gly Leu Gln Pro Leu Gln Gly Ala Leu Gln
 20 25

<210> SEQ ID NO 42
<211> LENGTH: 107
<212> TYPE: PRT
<213> ORGANISM: Gallus gallus
<406> SEQUENCE: 42

Met Ala Leu Trp Ile Arg Ser Leu Pro Leu Leu Ala Leu Leu Val Phe
  1   5   10   15
Ser Gly Pro Gly Thr Ser Tyr Ala Ala Ala Asp Gln His Leu Cys Gly
  20  25  30
Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
  35  40  45
Phe Tyr Ser Pro Lys Ala Arg Arg Asp Val Glu Gln Pro Leu Val Ser
  50  55  60
Ser Pro Leu Arg Gly Glu Ala Gly Val Leu Pro Phe Gln Gln Glu Glu
  65  70  75  80
Tyr Glu Lys Val Lys Arg Gly Ile Val Glu Gln Cys Cys His Asn Thr
  85  90
Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
  90 100 105

<210> SEQ ID NO 43
<211> LENGTH: 81
<212> TYPE: PRT
<213> ORGANISM: Anas platyrhynchos
<406> SEQUENCE: 43

Ala Ala Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
  1   5   10   15
Leu Val Cys Gly Leu Arg Gly Phe Phe Tyr Ser Pro Lys Thr Xaa Xaa
  20  25  30
Asp Val Glu Gin Pro Leu Val Asn Gin Pro Leu His Gln Glu Val Gln
  35  40  45
Glu Leu Pro Phe Gin His Glu Glu Tyr Gin Xaa Xaa Gly Ile Val Glu
  50  55  60
Gln Cys Cys Gin Gin Gin Pro Cys Ser Leu Tyr Gin Leu Glu Asn Cys
  65  70  75  80
Asn

<210> SEQ ID NO 44
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Anser anser
<406> SEQUENCE: 44

Gly Ile Val Glu Gin Cys Cys Gin Gin Gin Pro Cys Gin Cys Ser Leu Tyr Gin Leu
Glu Asn Tyr Cys Asn
20

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 45
<211> LENGTH: 103
<212> TYPE: PRT
<213> ORGANISM: Selasphorus rufus

| Ile Gln Ser Leu Pro Leu Leu Ala Leu Leu Ala Leu Ser Gly Pro Gly |
|---|---|---|---|---|---|---|
|   | 1 | 5 | 10 | 15 |

Thr Ser His Ala Ala Val Asn Glu His Leu Cys Gly Ser His Leu Val
20 25 30

| Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Ser Pro |
|---|---|---|---|---|---|---|---|---|
|   | 35 | 40 | 45 |

Lys Ala Arg Arg Asp Ala Glu His Pro Leu Val Asn Gly Pro Leu His
50 55 60

| Gly Glu Val Gly Asp Leu Pro Phe Gin Gin Glu Glu Gly Phe Lys Val |
|---|---|---|---|---|---|---|---|
|   | 65 | 70 | 75 | 80 |

Lys Arg Gly Ile Val Glu Gin Cys Cys His Asn Thr Cys Ser Leu Tyr
95 90 95

Gln Leu Glu Asn Tyr Cys Asn
100

<210> SEQ ID NO 46
<211> LENGTH: 109
<212> TYPE: PRT
<213> ORGANISM: Danio rerio

| Met Ala Val Trp Leu Gin Ala Gly Ala Leu Leu Leu Val Leu Val Val |
|---|---|---|---|---|---|---|---|---|---|
|   | 1 | 5 | 10 | 15 |

Ser Ser Val Ser Thr Asn Pro Gly Thr Pro Gin His Leu Cys Gly Ser
20 25 30

| His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Pro Thr Gly Phe Phe |
|---|---|---|---|---|---|---|---|
|   | 35 | 40 | 45 |

Tyr Asn Pro Lys Arg Asp Gin Gin Leu Gin Gly Leu Gly Phe Leu Pro Pro
50 55 60

| Lys Ser Ala Gin Thr Gin Val Ala Asp Phe Ala Phe Lys Asp His |
|---|---|---|---|---|---|---|---|
|   | 65 | 70 | 75 | 80 |

| Ala Glu Leu Ile Arg Gly Ile Val Glu Gin Cys Cys His Lys |
|---|---|---|---|---|---|---|
|   | 95 | 90 | 95 |

| Pro Cys Ser Ile Phe Gin Leu Gin Asn Tyr Cys Asn |
| ---|---|---|---|---|---|---|
|   | 100 |

<210> SEQ ID NO 47
<211> LENGTH: 108
<212> TYPE: PRT
<213> ORGANISM: Cyprinus carpio

| Met Ala Val Trp Ile Gin Ala Gly Ala Leu Leu Phe Leu Leu Ala Val |
|---|---|---|---|---|---|---|---|
|   | 1 | 5 | 10 | 15 |

Ser Ser Val Asn Ala Asn Ala Gly Ala Pro Gin His Leu Cys Gly Ser
20 25 30 30
His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Pro Thr Gly Phe Phe
35
40
45

Tyr Asp Pro Lys Arg Asp Val Asp Pro Leu Gly Phe Leu Pro Pro
50
55
60

Lys Ser Ala Gln Glu Thr Glu Val Ala Asp Phe Ala Phe Lys Asp His
65
70
75
80

Ala Glu Val Ile Arg Lys Arg Gly Ile Val Glu Gln Cys Cys His Lys
85
90
95

Pro Cys Ser Ile Phe Glu Leu Gln Asn Tyr Cys Asn
100
105

<G10> SEQ ID NO: 48
<G11> LENGTH: 21
<G12> TYPE: PRT
<G13> ORGANISM: Batrachoididae gen. sp.

<G40> SEQUENCE: 49

Gly Ile Val Glu Gln Cys Cys His Arg Pro Cys Asp Ile Phe Asp Leu
1
5
10
15

Gln Ser Tyr Cys Asn
20

<G10> SEQ ID NO: 49
<G11> LENGTH: 21
<G12> TYPE: PRT
<G13> ORGANISM: Thunnus thynnus

<G40> SEQUENCE: 49

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Asn Ile Phe Asp Leu
1
5
10
15

Gln Asn Tyr Cys Asn
20

<G10> SEQ ID NO: 50
<G11> LENGTH: 21
<G12> TYPE: PRT
<G13> ORGANISM: Katamwonus pelamis

<G40> SEQUENCE: 50

Gly Ile His Glx Glx Cys Cys His Lys Pro Cys Asx Ile Phe Glx Leu
1
5
10
15

Glx Asx Tyr Cys Asn
20

<G10> SEQ ID NO: 51
<G11> LENGTH: 116
<G12> TYPE: PRT
<G13> ORGANISM: Lophius piscatorius

<G40> SEQUENCE: 51

Met Ala Ala Leu Trp Leu Gln Ser Phe Ser Leu Leu Val Leu Leu Val
1
5
10
15

Val Ser Trp Pro Gly Ser Gln Ala Val Ala Pro Ala Gln His Leu Cys
20
25
30

Gly Ser His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Asp Arg Gly
35
40
45

Phe Phe Tyr Asn Pro Lys Arg Asp Val Asp Gln Leu Leu Gln Leu Gly Phe Leu
<table>
<thead>
<tr>
<th></th>
<th>50</th>
<th>55</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro</td>
<td>Pro Lys Ser Gly Gly Ala Ala Ala Gly Ala Ala Asp</td>
<td></td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Lys Glu Val</td>
<td>75</td>
<td>70</td>
</tr>
<tr>
<td>Ala</td>
<td>Ala Glu Phe Ala Phe Lys Asp Gln Met Glu Met</td>
<td></td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Met Val Lys Arg Gly</td>
<td>90</td>
<td>50</td>
</tr>
<tr>
<td>Ile</td>
<td>Ile Glu Glu Cys Cys His Arg Pro Cys Asn</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>Val</td>
<td>Ile Phe Asp Leu Gln</td>
<td>105</td>
<td>95</td>
</tr>
<tr>
<td>Asn</td>
<td>Asn Tyr Cys Asn</td>
<td></td>
<td>110</td>
</tr>
<tr>
<td></td>
<td></td>
<td>115</td>
<td>115</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 52
<211> LENGTH: 115
<212> TYPE: PRT
<213> ORGANISM: Myxine glutinosa

<400> SEQUENCE: 52

Met Ala Leu Ser Pro Phe Leu Ala Ala Val Ile Pro Leu Val Leu Leu 1 5 10 15
Leu Ser Arg Ala Pro Pro Ser Ala Asp Thr Arg Thr Gly His Leu 20 25 30
Cys Gly Lys Asp Leu Val Asn Ala Ala Tyr Ile Ala Cys Gly Val Arg 35 40 45
Gly Phe Phe Tyr Asp Pro Thr Lys Met Lys Arg Asp Thr Gly Ala Leu 50 55 60
Ala Ala Phe Leu Pro Leu Ala Tyr Ala Glu Asn Glu Ser Gin Asp 65 70 75 80
Asp Glu Ser Ile Gly Ile Asn Glu Val Leu Lys Ser Lys Arg Gly Ile 85 90 95
Val Glu Glu Cys Cys His Arg Cys Ser Ile Tyr Asp Leu Glu Asn 100 105 110
Tyr Cys Asn 115

<210> SEQ ID NO 53
<211> LENGTH: 105
<212> TYPE: PRT
<213> ORGANISM: Oncorhynchus keta

<400> SEQUENCE: 53

Met Ala Phe Tyr Leu Gln Ala Ala Ala Ser Leu Leu Val Leu Leu Ala Leu 1 5 10 15
Ser Pro Gly Val Asp Ala Ala Ala Ala Gin His Leu Cys Gly Ser His 20 25 30
Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Glu Gly Lys Phe Phe Tyr 35 40 45
Thr Pro Lys Arg Asp Val Asp Pro Leu Ile Gly Phe Leu Ser Pro Lys 50 55 60
Ser Ala Lys Glu Asn Glu Gly Tyr Pro Phe Lys Asp Gln Thr Glu Met 65 70 75 80
Met Val Lys Arg Gly Ile Val Glu Glu Cys Cys His Lys Pro Cys Asn 85 90 95
Ile Phe Asp Leu Gln Asn Tyr Cys Asn 100 105 105
-continued

<210> SEQ ID NO 54
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Myxocephalus scorpius

<400> SEQUENCE: 54

Gly Ile Val Glu Gln Cys Cys His Arg Pro Cys Asn Ile Arg Val Leu
  1  5  10  15
  Glu Asn Tyr Cys Asn
  20

<210> SEQ ID NO 55
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Lepisosteus spatula

<400> SEQUENCE: 55

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Thr Ile Tyr Glu Leu
  1  5  10  15
  Glu Asn Tyr Cys Asn
  20

<210> SEQ ID NO 56
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Platichthys flesus

<400> SEQUENCE: 56

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Asn Ile Phe Asp Leu
  1  5  10  15
  Gln Asn Tyr Cys Asn
  20

<210> SEQ ID NO 57
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Hydrologus collisi

<400> SEQUENCE: 57

Gly Ile Val Glu Gln Cys Cys His Asn Thr Cys Ser Leu Ala Asn Leu
  1  5  10  15
  Glu Gly Tyr Cys Asn
  20

<210> SEQ ID NO 58
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Squalus acantias

<400> SEQUENCE: 59

Gly Ile Val Glu His Cys Cys His Asn Thr Cys Ser Leu Tyr Asp Leu
  1  5  10  15
  Glu Gly Tyr Cys Asn Gln
  20

<210> SEQ ID NO 59
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Torpedo marmorata

<400> SEQUENCE: 59
Gly Ile Val Glu His Cys His Asn Thr Cys Ser Leu Phe Asp Leu
1     5       10       15
Glu Gly Tyr Cys Asn
20

<210> SEQ ID NO 60
<211> LENGTH: 89
<212> TYPE: PRT
<213> ORGANISM: Callorhinchus milii

<400> SEQUENCE: 60
Val Pro Thr Gin Arg Leu Cys Gin Ser Gin Leu Val Asp Ala Leu Tyr
1     5       10       15
Phe Val Cys Gly Gin Gly Phe Phe Tyr Ser Pro Lys Gin Ile Arg
20    25      30
Asp Val Gly Pro Leu Ser Ala Phe Arg Asp Leu Glu Pro Pro Leu Asp
35    40      45
Thr Gin Met Gin Gin Gin Leu Ala Gly Ser
50    55      60
Lys Met Lys Arg Gin Ile Val Gin Cys Cys His Asn Thr Cys Ser
65    70      75      80
Leu Val Asn Leu Glu Gly Tyr Cys Asn
85

<210> SEQ ID NO 61
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Petromyzon marinus

<400> SEQUENCE: 61
Gly Ile Val Glu Gin Cys Cys His Arg Lys Cys Ser Ile Tyr Asp Met
1     5       10       15
Glu Asn Tyr Cys Asn
20

<210> SEQ ID NO 62
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Oncorhynchus gorbuscha

<400> SEQUENCE: 62
Gly Ile Val Glu Gin Cys Cys His Arg Lys Cys Ser Ile Phe Asp Leu
1     5       10       15
Gln Asn Tyr Cys Asn
20

<210> SEQ ID NO 63
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Amia calva

<400> SEQUENCE: 63
Gly Ile Val Glu Gin Cys Leu Lys Pro Cys Thr Ile Tyr Glu Met
1     5       10       15
Glu Lys Tyr Cys Asn
20
-continued

<210> SEQ ID NO 64
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Anguilla rostrata

<400> SEQUENCE: 64

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Ser Ile Phe Asp Leu
  1  5  10  15

Gln Asn Tyr Cys Asn
  20

<210> SEQ ID NO 65
<211> LENGTH: 113
<212> TYPE: PRT
<213> ORGANISM: Osteochromis niloticus

<400> SEQUENCE: 65

Met Ala Ala Leu Trp Leu Gln Ala Phe Ser Leu Leu Val Leu Met Met
  1  5  10  15
Val Ser Trp Pro Gly Ser Gln Ala Val Gly Gly Pro Gln His Leu Cys
  20  25  30
Gly Ser His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Asp Arg Gly
  35  40  45
Phe Phe Tyr Asn Pro Arg Arg Asp Val Asp Pro Leu Gly Phe Leu
  50  55  60
Pro Pro Lys Ala Gly Gly Ala Val Val Gly Gly Gly Gln Asn Glu Val
  65  70  75  80
Thr Phe Lys Asp Gln Met Glu Met Met Val Lys Arg Gly Ile Val Glu
  85  90  95
Glu Cys Cys His Lys Pro Cys Thr Ile Phe Asp Leu Gln Asn Tyr Cys
 100 105 110

Asn

<210> SEQ ID NO 66
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Acipenser gueldenstaedti

<400> SEQUENCE: 66

Gly Ile Val Glu Gln Cys Cys His Ser Pro Cys Ser Leu Tyr Asp Leu
  1  6  10  15

Glu Asn Tyr Cys Asn
  20

<210> SEQ ID NO 67
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Piaractus mesopotamicus

<400> SEQUENCE: 67

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Ser Ile Phe Asp Leu
  1  5  10  15

Gln Asn Tyr Cys Asn
  20

<210> SEQ ID NO 68
<211> LENGTH: 115
<212> TYPE: PRT
-continued

<213> ORGANISM: Verasper moseri

<400> SEQUENCE: 69

Met Ala Ala Leu Trp Leu Gln Ser Val Ser Leu Leu Val Leu Val Leu Met Leu
1  5  10  15
Val Ser Trp Ser Gly Ser Gln Ala Val Leu Pro Pro Gln His Leu Cys
20  25  30
Gly Ala His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly
35  40  45
Phe Phe Tyr Thr Pro Lys Arg Asp Val Asp Pro Leu Leu Gly Phe Leu
50  55  60
Pro Ala Lys Ser Gly Ala Ala Gly Gly Glu Asn Glu Val Ala
65  70  75  80
Glu Phe Ala Phe Lys Asp Glu Met Glu Met Met Val Lys Arg Gly Ile
85  90  95
Val Glu Glu Cys Cys His Lys Pro Cys Asn Ile Phe Asp Leu Gin Asn
100 105 110

Tyr Cys Asn
115

<210> SEQ ID NO 69
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Anquilla anguilla

<400> SEQUENCE: 69

Amp Val Glu Pro Leu Leu Gly Phe Leu Ser Pro Lys Ser Gly Glu Glu
1  5  10  15
Asn Glu Val Asp Asp Phe Pro Tyr Lys Gly Glu Gly Glu Leu
20  25  30

<210> SEQ ID NO 70
<211> LENGTH: 106
<212> TYPE: PRT
<213> ORGANISM: Xenopus laevis

<400> SEQUENCE: 70

Met Ala Leu Trp Met Gln Cys Leu Pro Leu Val Leu Val Leu Phe Phe
1  5  10  15
Ser Thr Pro Asn Thr Glu Ala Leu Val Asn Glu His Leu Cys Gly Ser
20  25  30
His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Asp Arg Gly Phe Phe
35  40  45
Tyr Tyr Pro Lys Val Lys Arg Asp Met Gln Ala Leu Val Ser Gly
50  55  60
Pro Glu Asn Asp Leu Asp Gly Met Glu Leu Pro Gln Glu Tyr
65  70  75  80
Gln Lys Met Lys Arg Gly Ile Val Glu Glu Cys Cys His Ser Thr Cys
85  90  95
Ser Leu Phe Glu Leu Glu Ser Tyr Cys Asn
100 105

<210> SEQ ID NO 71
<211> LENGTH: 106
<212> TYPE: PRT
<213> ORGANISM: Xenopus laevis
<400> SEQUENCE: 71
Met Ala Leu Trp Met Gln Cys Leu Pro Leu Val Leu Val Leu Leu Phe
5  15
Ser Thr Pro Asn Thr Glu Ala Leu Ala Asn Glu His Leu Cys Gly Ser
20 25 30
His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Asp Arg Gly Phe Phe
35 40 45
Tyr Tyr Pro Lys Ile Lys Arg Asp Ile Glu Gln Ala Gln Val Asn Gly
50 55 60
Pro Gln Asp Asn Glu Leu Asp Gly Met Gln Phe Gln Pro Gln Glu Tyr
65 70 75 80
Gln Lys Met Lys Arg Gly Ile Val Glu Gin Cys Cys His Ser Thr Cys
85 90 95
Ser Leu Phe Gin Leu Glu Asn Tyr Cys Asn
100 105

<210> SEQ ID NO: 72
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Trachemys scripta

<400> SEQUENCE: 72
Gly Ile Val Glu Gin Cys Cys His Asn Thr Cys Ser Leu Tyr Gin Leu
1  5  10  15
Glu Asn Tyr Cys Asn
20

<210> SEQ ID NO: 73
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Alligator mississippiensis

<400> SEQUENCE: 73
Gly Ile Val Glu Gin Cys Cys His Asn Thr Cys Ser Leu Tyr Gin Leu
1  5  10  15
Glu Asn Tyr Cys Asn
20

<210> SEQ ID NO: 74
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Zacyscch rhynenades

<400> SEQUENCE: 74
Gly Ile Val Glu Gin Cys Cys Glu Asn Thr Cys Ser Leu Tyr Gin Leu
1  5  10  15
Glu Asn Tyr Cys Asn
20

<210> SEQ ID NO: 75
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Crotalus atrox

<400> SEQUENCE: 75
Gly Ile Val Glu Gin Cys Cys Glu Asn Thr Cys Ser Leu Tyr Gin Leu
1  5  10  15
-continued-

Glu Asn Tyr Cys Asn

20

<210> SEQ ID NO: 76
<211> LENGTH: 114
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Preproinsulin

<400> SEQUENCE: 76

Met Gly Leu Trp Ile Arg Leu Leu Pro Leu Ile Ala Leu Leu Ile Leu
1  5       10  15
Trp Gly Pro Asp Pro Ala Ala Ala Glu Phe Arg Met Phe Val Asn Gin
20 25 30
His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Val Cys Gly
35 40 45
Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp
50 55 60
65 70 75 80
Leu Gin Pro Leu Ala Leu Glu Gly Ser Leu Gin Lys Arg Gly Ile Val
85 90 95
Glu Gin Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gin Leu Asn Tyr
100 105 110

Cys Asn

<210> SEQ ID NO: 77
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin

<400> SEQUENCE: 77

Gly Ile Val Glu Gin Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gin Leu
1  5       10  15
Glu Asn Tyr Cys Asn
20

<210> SEQ ID NO: 78
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin

<400> SEQUENCE: 78

Phe Val Asn Gin His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1  5       10  16
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
20 25 30

<210> SEQ ID NO: 79
<211> LENGTH: 53
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin
<400> SEQUENCE: 79
Phe Val Asp Gin His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
  1  5       10       15
Leu Val Cys Gly Glu Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys
  20   25     30
Gly Ile Val Glu Gln Cys Thr Ser Ile Cys Ser Leu Tyr Glu Leu
  35   40     45
Glu Asp Tyr Cys Ann
  50

<415> SEQ ID NO: 90
<421> LENGTH: 59
<422> TYPE: PRT
<423> ORGANISM: Artificial Sequence
<424> FEATURE:
<425> OTHER INFORMATION: Insulin

<400> SEQUENCE: 80
Phe Val Glu Gin His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
  1  5       10       15
Leu Val Cys Gly Glu Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys
  20   25     30
Gly Ile Val Glu Gln Cys Thr Ser Ile Cys Ser Leu Tyr Glu Leu
  35   40     45
Glu Glu Tyr Cys Ann
  50

<415> SEQ ID NO: 91
<421> LENGTH: 53
<422> TYPE: PRT
<423> ORGANISM: Artificial Sequence
<424> FEATURE:
<425> OTHER INFORMATION: Insulin

<400> SEQUENCE: 91
Phe Val Gin Gin His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
  1  5       10       15
Leu Val Cys Gly Glu Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys
  20   25     30
Gly Ile Val Glu Gln Cys Thr Ser Ile Cys Ser Leu Tyr Glu Leu
  35   40     45
Glu Ann Tyr Cys Gly
  50

<415> SEQ ID NO: 92
<421> LENGTH: 53
<422> TYPE: PRT
<423> ORGANISM: Artificial Sequence
<424> FEATURE:
<425> OTHER INFORMATION: Unnamed protein product with insulin homology

<400> SEQUENCE: 92
Phe Val Thr Gin His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
  1  5       10       15
Leu Val Cys Gly Glu Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys
  20   25     30
Gly Ile Val Glu Gln Cys Thr Ser Ile Cys Ser Leu Tyr Glu Leu
  35   40     45
<table>
<thead>
<tr>
<th>Residue</th>
<th>Sequence</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu</td>
<td>His</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Ser</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Val</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Cys</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Leu</td>
<td>Val</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>Tyr</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>Arg</td>
<td>Glu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Lys</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Cys</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Tyr</td>
<td>Cys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Ser</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Val</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Cys</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Val</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Glu</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Lys</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Cys</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Tyr</td>
<td>Cys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Val</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Cys</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Val</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Thr</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Lys</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Cys</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Tyr</td>
<td>Cys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>Asn</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>
<200> FEATURE:
<223> OTHER INFORMATION: Insulin

<400> SEQUENCE: 86

Arg Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu
  1      5     10  15

Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala Ala
  20     25     30

Lys Gly Ile Val Glu Gln Cys Thr Ser Ile Cys Ser Leu Tyr Gln
  35     40     45

Leu Glu Asn Tyr Cys Asn
  50

<210> SEQ ID NO: 87
<211> LENGTH: 61
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin

<400> SEQUENCE: 87

Lys Glu Thr Leu Thr Ile Thr Cys Ala Val Pro Thr Trp Leu Lys Leu
  1  5  10  15

Trp Thr Trp Phe Ala Val Lys Glu Val Ser Ser Thr Asn Leu Arg Leu
  20  25  30

Leu Arg Val Leu Ser Asn Asn Ala Val Pro Pro Ser Ala Pro Cys Thr
  35  40  45

Asn Trp Lys Thr Thr Ala Thr Arg Ser Ser Pro Gln Ala
  50  55  60

<210> SEQ ID NO: 88
<211> LENGTH: 61
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Preproinsulin

<400> SEQUENCE: 88

Lys Asp Ser Leu Thr Asn Thr Cys Ala Val Ser Thr Trp Leu Lys Leu
  1  5  10  15

Cys Thr Trp Phe Ala Val Lys Glu Val Ser Ser Thr Leu Leu Arg Leu
  20  25  30

Leu Arg Val Leu Ser Asn Asn Ala Val Pro Pro Ser Ala Asn Tyr Thr
  35  40  45

Asn Trp Lys Thr Thr Ala Thr Arg Ser Ser Pro Gln Ala
  50  55  60

<210> SEQ ID NO: 89
<211> LENGTH: 87
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Proinsulin

<400> SEQUENCE: 89

Met Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu
  1  5  10  15

Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr Arg
  20  25  30
Arg Glu Ala Glu Asp Leu Gln Val Gly Gln Val Glu Leu Gly Gly Gly
35 40 45
Pro Gly Ala Gly Ser Leu Gln Pro Leu Ala Leu Glu Gly Ser Leu Gln
50 55 60
Lys Arg Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr
65 70 75 80
Gln Leu Glu Asn Tyr Cys Asn
95

<210> SEQ ID NO 90
<211> LENGTH: 35
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin
<400> SEQUENCE: 90
Arg Arg Glu Ala Ala Asp Leu Gln Val Gly Gln Val Glu Leu Gly Gly
1  5 10 15
Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu Ala Leu Glu Gly Ser Leu
20 25 30
Gln Lys Arg
35

<210> SEQ ID NO 91
<211> LENGTH: 23
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin
<400> SEQUENCE: 91
Gly Pro Glu Thr Leu Cys Ala Ala Glu Leu Val Asp Ala Leu Gln Phe
1  5 10 15
Val Cys Gly Asp Arg Gly Phe
20

<210> SEQ ID NO 92
<211> LENGTH: 124
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin
<400> SEQUENCE: 92
Met Lys Leu Lys Thr Val Arg Ser Ala Val Leu Ser Ser Leu Phe Ala
1  5 10 15
Ser Gln Val Leu Gly Gln Pro Ile Asp Thr Glu Ser Gln Thr Thr
20 25 30
Ser Val Asn Leu Met Ala Asp Thr Glu Ser Ala Phe Ala Thr Gln
35 40 45
Thr Asn Ser Gly Leu Asp Val Val Gly Leu Ile Ser Met Ala Lys
50 55 60
Arg Glu Gly Glu Pro Lys Phe Val Asn Gln His Leu Cys Gly Ser
65 70 75 80
His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe
85 90 95
-continued

Tyr Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gin Cys Thr Ser
100 105
Ile Cys Ser Leu Tyr Gin Leu Glu Asn Tyr Cys Asn
115 120

<210> SEQ ID NO 93
<211> LENGTH: 171
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin

<400> SEQUENCE: 93

ttgcatcgc agaccctgg cttcttcac cttggccgg ctttgccccg cttgctgccc ggg 60
gaagcgcgg cttcttcac accaaaccc cctgtaaac ctaagcgccc cattgccgg 120
cactgctca cccagatcct cctccccac ccaactggca actactgcac c 171

<210> SEQ ID NO 94
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin analogue

<400> SEQUENCE: 94

Gly Ile Val Glu Gin Cys Thr Ser Ile Cys Ser Leu Tyr Gin Leu
1   5  10  15
Glu Asn Tyr Cys Asn
20

<210> SEQ ID NO 95
<211> LENGTH: 26
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin analogue
<222> LOCATION: (26)...(26)
<223> OTHER INFORMATION: X = any amino acid

<400> SEQUENCE: 95

Phe Val Asn Gin His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1   6  10  15
Leu Val Cys Gin Arg Gly Phe Phe Xaa
20  25

<210> SEQ ID NO 96
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 96

Gly Ile Val Glu Gin Cys Thr Ser Ile Cys Ser Leu Tyr Gin Leu
1   5  10  15
Glu Asn Tyr Cys Asn
20

<210> SEQ ID NO 97
<211> LENGTH: 29
<212> TYPE: PRT
-continued

<213> ORGANISM: *Homo sapiens*

<400> SEQUENCE: 97

Glu Val Asn Gin His Leu Cys Gly Ser Glu Leu Val Glu Ala Leu Glu
1  5  10  15
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Glu Pro Lys
20  25

<210> SEQ ID NO 98
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: *Homo sapiens*

<400> SEQUENCE: 98

Gly Ile Val Glu Gin Cys Thr Ser Ile Cys Ser Leu Tyr Gin Leu
1  5  10  15
Glu Asn Tyr Cys Asn
20

<210> SEQ ID NO 99
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: *Homo sapiens*

<400> SEQUENCE: 99

Phe Val Asn Gin His Leu Cys Gly Ser His Leu Val Glu Ala Leu His
1  5  10  15
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
20  25  30

<210> SEQ ID NO 100
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: *Homo sapiens*

<400> SEQUENCE: 100

Gly Ile Val Glu Gin Cys Lys Ser Ile Cys Ser Leu Tyr Gin Leu
1  5  10  15
Glu Asn Tyr Cys Asn
20

<210> SEQ ID NO 101
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: *Homo sapiens*

<400> SEQUENCE: 101

Phe Val Asn Gin His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1  5  10  15
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
20  25  30

<210> SEQ ID NO 102
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: *Homo sapiens*

<400> SEQUENCE: 102

Gly Ile Val Glu Gin Cys Lys Ser Ile Cys Ser Leu Tyr Gin Leu
1  5  10  15
Glu Asn Tyr Cys Asn
  20

<210> SEQ ID NO 103
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 103
Phe Val Asn Glu His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
  1   5   10  15
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
  20  25  30

<210> SEQ ID NO 104
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 104
Gly Ile Val Glu Glu Cys Xaa Ser Ile Cys Ser Leu Tyr Glu Leu
  1   5   10  15
Glu Asn Tyr Cys Asn
  20

<210> SEQ ID NO 105
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 105
Phe Val Asn Glu His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
  1   5   10  15
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
  20  25  30

<210> SEQ ID NO 106
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 106
Gly Ile Val Glu Glu Cys Xaa Ser Ile Cys Ser Leu Tyr Glu Leu
  1   5   10  15
Glu Asn Tyr Cys Asn
  20

<210> SEQ ID NO 107
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 107
-continued

Fhe Val Aen Gin His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
  1     5      10     15

Leu Val Cys Gly Glu Arg Gly Fhe Fhe Tyr Thr Pro Lys Thr
 20    25     30

<210> SEQ ID NO 108
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: [1]...[2]
<223> OTHER INFORMATION: X = any amino acids

<400> SEQUENCE: 108

Gly Xaa Val Glu Gin Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gin Leu
  1     5      10     15

Glu Aen Tyr Cys Aen
  20

<210> SEQ ID NO 109
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 109

Fhe Val Aen Gin His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
  1     5      10     15

Leu Val Cys Gly Glu Arg Gly Fhe Fhe Tyr Thr Lys Pro Thr
 20    25     30

<210> SEQ ID NO 110
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 110

Gly Ala Val Glu Gin Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gin Leu
  1     5      10     15

Glu Aen Tyr Cys Aen
  20

<210> SEQ ID NO 111
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 111

Fhe Val Aen Gin His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
  1     5      10     15

Leu Val Cys Gly Glu Arg Gly Fhe Fhe Tyr Thr Lys Pro Thr
 20    25     30

<210> SEQ ID NO 112
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence

<400> SEQUENCE: 112

Phe Val Aen Gin His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
  1     5      10     15

Leu Val Cys Gly Glu Arg Gly Fhe Fhe Tyr Thr Lys Pro Thr
 20    25     30

<222> LOCATION: [21]...[2]
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly Xaa Val Glu Gln Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu</td>
<td>X = any amino acid</td>
</tr>
<tr>
<td></td>
<td><strong>SEQUENCE:</strong> 112</td>
</tr>
<tr>
<td></td>
<td>Glu Asn Tyr Cys Asn</td>
</tr>
<tr>
<td></td>
<td><strong>SEQUENCE:</strong> 113</td>
</tr>
<tr>
<td></td>
<td>Phe Val Asn Gln His Leu Gly Ser His Leu Val Gln Ala Leu Tyr</td>
</tr>
<tr>
<td></td>
<td>Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr</td>
</tr>
<tr>
<td></td>
<td><strong>SEQUENCE:</strong> 114</td>
</tr>
<tr>
<td></td>
<td>Gly Xaa Val Glu Gln Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu</td>
</tr>
<tr>
<td></td>
<td>Glu Asn Tyr Cys Asn</td>
</tr>
<tr>
<td></td>
<td><strong>SEQUENCE:</strong> 115</td>
</tr>
<tr>
<td></td>
<td>Phe Val Asn Gln His Leu Gly Ser His Leu Val Gln Ala Leu Tyr</td>
</tr>
<tr>
<td></td>
<td>Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr</td>
</tr>
<tr>
<td></td>
<td><strong>SEQUENCE:</strong> 116</td>
</tr>
<tr>
<td></td>
<td>Gly Gly Gly Glu Gln Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu</td>
</tr>
<tr>
<td></td>
<td>Glu Asn Tyr Cys Asn</td>
</tr>
</tbody>
</table>
-continued

<210> SEQ ID NO 117
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 117

Phe Val Asn Gin His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
1  5  10  15
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Lys Pro Thr
20 25 30

<210> SEQ ID NO 118
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Homo sapiens insulin mutant

<400> SEQUENCE: 118

Xaa Ile Val Glu Gin Cys Thr Ser Ile Cys Ser Leu Tyr Gin Leu
1  5  10  15
Glu Gin Tyr Cys Asn
20

<210> SEQ ID NO 119
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Homo sapiens insulin mutant

<400> SEQUENCE: 119

Xaa Val Asn Gin His Leu Cys Gly Asp His Leu Val Glu Ala Leu Tyr
1  5  10  15
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
20 25 30

<210> SEQ ID NO 120
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Homo sapiens insulin mutant

<400> SEQUENCE: 120

Gly Ile Val Glu Gin Cys Thr Ser Ile Cys Ser Leu Tyr Gin Leu
1  5  10  15
Glu Gin Tyr Cys Asn
20

<210> SEQ ID NO 121
<211> LENGTH: 29
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
OTHER INFORMATION: Homo sapiens insulin mutant

SEQUENCE: 121

Phe Val Asn Gin His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
  1      5      10    15
Leu Val Cys Gly Glu Arg Gly Phe Tyr Thr Pro Lys Thr
  20    25

SEQ ID NO 122
LENGTH: 21
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Homo sapiens insulin

SEQUENCE: 122

Gly Ile Val Glu Gin Ser Thr Ser Ile Ser Ser Leu Tyr Gin Leu
  1      5      10    15
Glu Asn Tyr Cys Asn
  20

SEQ ID NO 123
LENGTH: 30
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Homo sapiens insulin

SEQUENCE: 123

Phe Val Asn Gin His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
  1      5      10    15
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Lys Pro Thr
  20    25    30

SEQ ID NO 124
LENGTH: 79
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Insulin

SEQUENCE: 124

Met Asp Pro Gly Asp Pro Glu Cys Leu Gin Leu Leu Arg Arg Leu
  1      5      10    15
Gly Gly Ser Val Glu Val Glu Val Thr Gly Thr Val His Val Glu
  20    25    30
Val Ser Pro Glu Asp Pro Gly Asp Pro Glu Cys Leu Gin Leu Leu
  35    40    45
Arg Arg Leu Gly Ser Val Glu Val Glu Thr Gly Thr Val
  50    55    60
His Val Glu Val Ser Pro Gly Glu Arg Gly Phe Phe Tyr Cys Asn
  65    70    75

SEQ ID NO 125
LENGTH: 87
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Insulin

SEQUENCE: 125
-continued

```plaintext
Net Leu Lys Glu Lys Lys Tyr Ser Pro Asp Pro Gly Asp Pro Glu Cys
  1 5 10 15
Leu Gln Leu Leu Arg Arg Leu Gly Gly Ser Val Glu Val Glu Val
  20 25 30
Thr Gly Gly Thr Val His Val Glu Val Ser Pro Asp Pro Gly Asp
  35 40 45
Pro Gly Cys Leu Gln Leu Leu Arg Arg Leu Gly Gly Ser Val Glu
  50 55 60
Val Gln Val Thr Gly Thr Val Gly Thr Val His Val Glu Val Ser Pro Gly Glu
  65 70 75 80
Arg Gly Phe Phe Tyr Cys Asn
  85
```

SEQ ID NO 126
LENGTH: 23
TYPE: PRT
ORGANISM: Artificial Sequence

OTHER INFORMATION: Insulin fusion protein
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (5)...
OTHER INFORMATION: X = any amino acid

SEQ ID NO 127
LENGTH: 23
TYPE: PRT
ORGANISM: Escherichia coli

FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (14)...
OTHER INFORMATION: X = any amino acid

SEQ ID NO 128
LENGTH: 96
TYPE: PRT
ORGANISM: Brevibacillus brevis

SEQ ID NO 129
LENGTH: 15
-continued

Phe Ala Aen Ser Asp Ser Glu Ser Pro Leu Ser His Asp Gly Tyr Ser
20 25 30
Leu His Asp Gly Val Ser Met Tyr Ile Glu Ala Leu Asp Lys Phe Val
35 40 45
Asn Gin His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val
50 55 60
Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Gly Ile Val Glu Gin
65 70 75 80
Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gin Leu Glu Asn Tyr Cys Asn
90 95 95

<210> SEQ ID NO 129
<211> LENGTH: 25
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin fusion protein

<400> SEQUENCE: 129

Asp Thr Thr Met Pro Ala Gly Gly Gly Gly Gly Gly Gly Gly His Leu Cys
1 5 10 15
Gly Pro His Leu Val Glu Ala Leu Tyr
20 25

<210> SEQ ID NO 130
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin fusion protein

<400> SEQUENCE: 130

Leu Glu Aen Tyr Cys Asn
1 5

<210> SEQ ID NO 131
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin fusion protein

<400> SEQUENCE: 131

Met Thr Met Ile Thr Asp Ser Leu Glu Phe Gin Ala Trp Gly Gly Gly
1 5 10 15
Gly Gly Trp Met Arg Phe
20

<210> SEQ ID NO 132
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin fusion protein

<400> SEQUENCE: 132

Met Val Leu Arg Phe Leu Pro Leu Leu Ala Leu Leu Val Leu Trp Glu
1 5 10 16
Pro Lys Pro Ala Gin Ala
20
-continued

**SEQ ID NO 133**
**LENGTH: 60**
**TYPE: PRT**
**ORGANISM: Artificial Sequence**
**FEATURE:**
**OTHER INFORMATION: Mini-proinsulin**

**SEQUENCE:**

```
Fhe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr  
1  5  10  15
Leu Val Cys Gly Glu Arg Gly Fhe Phe Tyr Thr Pro Lys Thr Arg Arg  
20  25  30
Tyr Pro Gly Asp Val Lys Arg Gly Ile Val Glu Cys Cys Thr Ser  
35  40  45
Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn  
50  55  60
```

**SEQ ID NO 134**
**LENGTH: 21**
**TYPE: PRT**
**ORGANISM: Homo sapiens**

**SEQUENCE:**

```
Gly Ile Val Glu Cys Cys Lys Ser Ile Cys Ser Leu Tyr Gln Leu  
1  5  10  15
Glu Asn Tyr Cys Asn  
20
```

**SEQ ID NO 135**
**LENGTH: 30**
**TYPE: PRT**
**ORGANISM: Homo sapiens**

**SEQUENCE:**

```
Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr  
1  5  10  15
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr  
20  25  30
```

**SEQ ID NO 136**
**LENGTH: 21**
**TYPE: PRT**
**ORGANISM: Homo sapiens**

**SEQUENCE:**

```
Gly Ile Val Glu Cys Cys Lys Ser Ile Cys Ser Leu Tyr Gln Leu  
1  5  10  15
Glu Asn Tyr Cys Asn  
20
```

**SEQ ID NO 137**
**LENGTH: 30**
**TYPE: PRT**
**ORGANISM: Homo sapiens**

**SEQUENCE:**

```
Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr  
1  5  10  15
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr  
```
-continued

<table>
<thead>
<tr>
<th></th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Leu</td>
<td>Val</td>
<td>Glu</td>
</tr>
<tr>
<td>2</td>
<td>Ala</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td>3</td>
<td>Val</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td>4</td>
<td>Cys</td>
<td>Gly</td>
<td>Glu</td>
</tr>
<tr>
<td>5</td>
<td>Arg</td>
<td>Gly</td>
<td>Gly</td>
</tr>
</tbody>
</table>

SEQ ID NO: 138
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Homo sapiens insulin

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gly</td>
<td>Ile</td>
</tr>
<tr>
<td>2</td>
<td>Val</td>
<td>Glu</td>
</tr>
<tr>
<td>3</td>
<td>Gln</td>
<td>Cys</td>
</tr>
<tr>
<td>4</td>
<td>Xaa</td>
<td>Ser</td>
</tr>
<tr>
<td>5</td>
<td>Ile</td>
<td>Cys</td>
</tr>
<tr>
<td>6</td>
<td>Ser</td>
<td>Leu</td>
</tr>
<tr>
<td>7</td>
<td>Tyr</td>
<td>Gln</td>
</tr>
<tr>
<td>8</td>
<td>Leu</td>
<td></td>
</tr>
</tbody>
</table>

SEQ ID NO: 139
LENGTH: 21
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: [8]..[8]
OTHER INFORMATION: X = any amino acid

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Glu</td>
<td>Asn</td>
</tr>
<tr>
<td>2</td>
<td>Tyr</td>
<td>Cys</td>
</tr>
<tr>
<td>3</td>
<td>Asn</td>
<td></td>
</tr>
</tbody>
</table>

SEQ ID NO: 140
LENGTH: 30
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Homo sapiens insulin

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phe</td>
<td>Val</td>
</tr>
<tr>
<td>2</td>
<td>Asn</td>
<td>Gln</td>
</tr>
<tr>
<td>3</td>
<td>His</td>
<td>Leu</td>
</tr>
<tr>
<td>4</td>
<td>Cys</td>
<td>Gly</td>
</tr>
<tr>
<td>5</td>
<td>Ser</td>
<td>His</td>
</tr>
<tr>
<td>6</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td>7</td>
<td>Glu</td>
<td>Ala</td>
</tr>
<tr>
<td>8</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
</tbody>
</table>

SEQ ID NO: 141
LENGTH: 21
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Homo sapiens insulin
NAME/KEY: MISC_FEATURE
LOCATION: [8]..[8]
OTHER INFORMATION: X = any amino acid

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gly</td>
<td>Ile</td>
</tr>
<tr>
<td>2</td>
<td>Val</td>
<td>Glu</td>
</tr>
<tr>
<td>3</td>
<td>Gln</td>
<td>Cys</td>
</tr>
<tr>
<td>4</td>
<td>Xaa</td>
<td>Ser</td>
</tr>
<tr>
<td>5</td>
<td>Ile</td>
<td>Cys</td>
</tr>
<tr>
<td>6</td>
<td>Ser</td>
<td>Leu</td>
</tr>
<tr>
<td>7</td>
<td>Tyr</td>
<td>Gln</td>
</tr>
<tr>
<td>8</td>
<td>Leu</td>
<td></td>
</tr>
</tbody>
</table>

SEQ ID NO: 142
LENGTH: 30
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Homo sapiens insulin

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Glu</td>
<td>Asn</td>
</tr>
<tr>
<td>2</td>
<td>Tyr</td>
<td>Cys</td>
</tr>
<tr>
<td>3</td>
<td>Asn</td>
<td></td>
</tr>
</tbody>
</table>
-continued

<400> SEQUENCE: 142
Phe Val Asn Gin His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
  1    5    10   15
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
  20   25   30

<210> SEQ ID NO 143
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Mini-proinsulin mutant
<400> SEQUENCE: 143
Gly Ile Val Glu Gin Cys Thr Ser Ile Cys Ser Leu Tyr Gin Leu
  1    5    10   15
Glu Asn Tyr Cys Asn
  20

<210> SEQ ID NO 144
<211> LENGTH: 29
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Mini-proinsulin mutant
<400> SEQUENCE: 144
Phe Val Asn Gin His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
  1    5    10   15
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Asp Lys
  20   25

<210> SEQ ID NO 145
<211> LENGTH: 50
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Mini-proinsulin mutant
<400> SEQUENCE: 145
Phe Val Asn Gin His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
  1    5    10   15
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Asp Lys Gin Ile Val
  20   25   30
Glu Gin Cys Thr Ser Ile Cys Ser Leu Tyr Gin Leu Glu Asn Tyr
  35   40   45
Cys Asn
  50

<210> SEQ ID NO 146
<211> LENGTH: 3
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin C-peptide
<400> SEQUENCE: 146
Ala Ala Lys
  1
Asn Lys Arg
1

Arg Arg Lys Gln Lys Arg
1 5

Arg Arg Lys Arg
1

Lys Arg Glu Leu
1

His Arg Glu Leu
1

Asp Arg Glu Leu
<210> SEQ ID NO 153
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE: OTHER INFORMATION: ER retention sequence
<400> SEQUENCE: 193

Ala Asp Glu Leu

<210> SEQ ID NO 154
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE: OTHER INFORMATION: ER retention sequence
<400> SEQUENCE: 154

Ser Asp Glu Leu

<210> SEQ ID NO 155
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Lycopersicon esculentum Mill.
<400> SEQUENCE: 155

His Asp Glu Phe

<210> SEQ ID NO 156
<211> LENGTH: 118
<212> TYPE: PRT
<213> ORGANISM: Arabidopsis thaliana
<400> SEQUENCE: 156

Met Ala Asp Thr Ala Arg Gly Thr His His Asp Ile Ile Gly Arg Asp
  1   5   10   15
Gln Tyr Pro Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Ser Gly
  20  25   30
Arg Gly Ser Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Ala Thr
  35  40  45
Ala Val Thr Ala Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu
  50  55  60
Val Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile
  65  70  75  80
Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile
  85  90  95
Thr Gly Phe Leu Ser Ser Gly Gln Phe Gly Ile Ala Ala Ile Thr Val
  100 105 110
Phe Ser Trp Ile Tyr Lys
  115

<210> SEQ ID NO 157
<211> LENGTH: 187
<212> TYPE: PRT
<213> ORGANISM: Brassica napus
<400> SEQUENCE: 157
Met Ala Asp Thr Ala Arg Thr His His Asp Val Thr Ser Arg Asp Glu
1  5  10  15
Tyr Pro Arg Asp Arg Glu Tyr Ser Met Ile Gly Arg Arg Asp
20  25  30
Gln Tyr Ser Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Tyr Gly
35  40  45
Arg Asp Tyr Ser Lys Ser Arg Glu Ile Ala Lys Ala Val Thr Ala Val
50  55  60
Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu Val Gly
65  70  75  80
Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile Phe Ser
85  90  95
Pro Ile Leu Val Pro Ala Leu Ile Thr Val Leu Ala Leu Ile Thr Gly
100 105 110
Phe Leu Ser Ser Gly Gly Phe Ala Ile Ala Ala Ile Thr Val Phe Ser
115 120 125
Trp Ile Tyr Lys Tyr Ala Thr Gly His Pro Glu Gly Ser Asp Lys
130 135 140
Leu Asp Ser Ala Arg Met Lys Leu Gly Thr Lys Ala Glu Arg Ile Gly
145 150 155 160
Asp Arg Ala Glu Tyr Gly Glu His Thr Gly Gly Glu His Asp
165 170 175
Arg Asp Arg Thr Arg Gly Gly Glu His Thr Thr
180 185

<210> SEQ ID NO 158
<211> LENGTH: 748
<212> TYPE: DNA
<213> ORGANISM: Arabidopsis thaliana

<400> SEQUENCE: 158
taccatggg tcaagaacgg agatgatgga gagaacgcc atggctacgg tgctcctca 60
tgcccggct acttabatct gcgcgctctcg tgtgacttgg gatgatagac ttctaaaccc 120
tatatgcga agacacccca cagagaacac cggtaagggaa cttcaggccca 180
taaggaattac ggaccttgtg tgttcttacca gcatacctgc ttctgata tcgatgataaa 240
tggcattcat tacccttggg agacctactc tgtcaagctga atgtctttat tcaatacat 300
tggtctgttt ataagacccgc cttgattctct acgtattgcct ctttccgccgg 360
tgtggtaacct tcaatttttt tcccttata catacacaac atacacaagt caagacagcttg 420
asagtaatta aaaaacatag acaatgaagg aagttttattg cggtaaatc tgggttgat 480
atttagaattataggaaaa cttgccacga caagttgatgt cttggaacac tatgaggagat 540
gcagaagggaa aaccctagct cttggcaccat ttctgattgc atgcgcgaac aataagatgtg 600
ggacagtttg tcttgctctg acgaggttaa agagggtttt tttgctaaag aagctattag 660
gggttcttttt gcattgaagt ctttgctagtt cttgcacaaaa atactccagtg ataccgatga 720
agcacaacagc cctattaagc ccatggat 748
-continued

<210> TYPE: DNA
<213> ORGANISM: Arabidopsis thaliana

<400> SEQUENCE: 159
atggggctaa agacgagat gatggagaga gacgcaatgg ctacggtgcc tcctatatg 60
cggtcaacct acacacgccc tcgccgcttg gacctgctgg atagaactcc taaactatat 120
atgcaagag cattgcaac accagaacga gaaactcag cagagaacct ggacgctagtaag 180
aatcgcggcg tgttaacttc tcaacagcat gtctcctctg tggatatacgg atgatagccc 240
aatcatttac ccctggagac ctacgttga atggcagttgc ctggttttaa tataacttggg 300
tcgccttaaa tagacctggt tataacagct acctcttagct atgcaactct ttcgggggttg 360
ttcctccct ctttttcccc tatacatca cacacatcc acaagtttcc cgaatggaattg 420
gatccaaaa cattgacaa tcagggccag tttatgcctgg tgaattctttt ggtgatatctt 480
agcaaacatt gcagacaacag tttatttcgg gatactgg tggagtattgg aagagaatc 540
gagaaagcc gtaagcctttt ggagatgtcg cagcagatatt agagttgagaa 600
tctttgtctc tttttgctgg cagagagaa ggttttttctc cagagagctg ttagagggg 660
tgttccagtt gaaggttttt catgactcatt gccaatatct aagcgtggtatat cgtgaaaac 720
agacgctactaac 738

<210> SEQ ID NO: 160
<211> LENGTH: 1047
<212> TYPE: DNA
<213> ORGANISM: Secanum indicum

<400> SEQUENCE: 160
atggatctaa ttcacccctt cttcagcattta atagcctccc ctctcccttt ctcttctct 60
tcttttttct tgtcaacccct tcaagtttcc atgctctcct ttcacccctt gggacccctt 120
ttcagccagag ttgctgcttg aaaaacctcg gtcataacg ggcctcctcc cggcatgccc 180
gaagagtcct cttagcgtag tggagagaga gggcggtgct tgggtgcttg tggcaaggggg 240
gaaagagtttt cttgagcttt ttgcccagag ggcggcatcg ggcgttgctct 300
ggctgcgc gcggctcttt cggagcgaga gatcagatgg cagcctttct tcaagcctag 360
ttcgctgttta gaagatgttg tttccctggct cttacagctg gaattctgg tgaattcttt 420
ttggagagag tggagttcata tctgtcttcag gagaatactag ccctcactgg 480
tttgatgtaa ttcacccctt ttcagcctccc cttcagcattta atagcctccc ctctcccttt 540
tcgccctca ccgagagctg tctggttac cttgcacagcct tggagttcata tctgtcttg 600
ggtcgagac gggggtgggg aagagtttctt cggagcgaga gatcagatgg cagcctttct 660
ttcagcctccc cttcagcattta atagcctccc ctctcccttt ctcttctctt 720
gtgatcagac ggtgaattctt ttcagcctccc cttcagcattta atagcctccc ctctcccttt 780
ttcgctgttta gaagatgttg tttccctggct cttacagctg gaattctgg tgaattcttt 840
tttgatgtaa ttcacccctt ttcagcctccc cttcagcattta atagcctccc ctctcccttt 900
ggtcgagac gggggtgggg aagagtttctt cggagcgaga gatcagatgg cagcctttct 960
cttgcacagcct tggagttcata tctgtcttcag gagaatactag ccctcactgg 1020
cagcctgcttta gggggtgggg aagagtttctt cggagcgaga gatcagatgg cagcctttct 1087
S. manilensis P16837

<210> SEQ ID NO 161
<211> LENGTH: 25
<212> TYPE: PRT
<213> ORGANISM: Tobacco, pathogenesis related protein (PR-5) signal sequence

<400> SEQUENCE: 161

Met Asn Phe Leu Lys Ser Phe Pro Phe Tyr Ala Phe Leu Cys Phe Gly
1  5  10  15

Gln Tyr Phe Val Ala Val Thr His Ala
20  25

<210> SEQ ID NO 162
<211> LENGTH: 64
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Alpha factor leader sequence

<400> SEQUENCE: 162

 Ala Pro Val Asn Thr Thr Glu Asp Glu Thr Ala Gln Ala Gln Ala Val
1  5  10  15

Ile Gly Tyr Ser Asp Leu Glu Gly Asp Phe Asp Val Ala Val Leu Pro
20  25  30

Phe Ser Asn Ser Thr Asn Gly Leu Leu Phe Ile Aax Thr Thr Ile
35  40  45

Ala Ser Ile Ala Ala Lys Glu Gly Val Ser Leu Met Ala Lys Arg
50  55  60

<210> SEQ ID NO 163
<211> LENGTH: 63
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Alpha factor leader sequence

<400> SEQUENCE: 163

 Ala Pro Val Asn Thr Thr Glu Asp Glu Thr Ala Gln Ala Gln Ala Val
1  5  10  15

Ile Gly Tyr Ser Asp Leu Glu Gly Asp Phe Asp Val Ala Val Leu Pro
20  25  30

Phe Ser Asn Ser Thr Asn Gly Leu Leu Phe Ile Aax Thr Thr Ile
35  40  45

Ala Ser Ile Ala Ala Lys Glu Gly Val Ser Met Ala Lys Arg
50  55  60

<210> SEQ ID NO 164
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Alpha factor leader sequence

<400> SEQUENCE: 164

Gln Pro Ile Asp Glu Asp Asp Thr Ser Ser Met Ala Lys Arg
1  5  10  15

<210> SEQ ID NO 165
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Alpha factor leader sequence

SEQ: 165
Gln Pro Ile Asp Asp Thr Glu Ser Asn Thr Thr Ser Val Asn Leu Met
   1   5  10  15

Ala Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Thr Leu Ala Leu
   20  25  30

Asp Val Val Asn Leu Ile Ser Met Ala Lys Arg
   35  40

SEQ: 166
Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val Asn Leu Met
   1   5  10  15

Ala Asp Thr Glu Asp Arg Phe Ala Thr Gln Thr Thr Leu Ala Leu
   20  25  30

Asp Val Val Asn Leu Ile Ser Met Ala Lys Arg
   35  40

SEQ: 167
Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val Asn Leu Met
   1   5  10  15

Ala Asp Thr Glu Asp Arg Phe Ala Thr Gln Thr Thr Leu Ala Leu
   20  25  30

Asp Val Val Asn Leu Ile Ser Met Ala Ala Ala
   35  40

SEQ: 168
Gln Pro Ile Asp Asp Thr Glu Ser Asn Thr Thr Ser Val Asn Leu Met
   1   5  10  15

Ala Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Thr Leu Ala Leu
   20  25  30

Asp Val Val Asn Leu Ile Ser Met Ala Ala Ala
   35  40

SEQ: 169
Gln Pro Ile Asp Asp Thr Glu Ser Asn Thr Thr Ser Val Asn Leu Met
   1   5  10  15

Ala Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Thr Leu Ala Leu
   20  25  30

Asp Val Val Asn Leu Ile Ser Met Ala Ala Ala
   35  40

SEQ: 170
Gln Pro Ile Asp Asp Thr Glu Ser Asn Thr Thr Ser Val Asn Leu Met
   1   5  10  15

Ala Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Thr Leu Ala Leu
   20  25  30

Asp Val Val Asn Leu Ile Ser Met Ala Ala Ala
   35  40
-continued

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
<th>Type</th>
<th>Organism</th>
<th>Feature</th>
<th>Other Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gln Pro Ile Asp Asp Thr Glu Ser Asn Thr Ser Val Aen Leu Met</td>
<td>44</td>
<td>PRT</td>
<td>Artificial</td>
<td>Sequence</td>
<td>Alpha factor leader sequence</td>
</tr>
<tr>
<td>Gln Pro Ile Asp Asp Thr Glu Ser Asn Thr Ser Val Aen Leu Met</td>
<td>44</td>
<td>PRT</td>
<td>Artificial</td>
<td>Sequence</td>
<td>Alpha factor leader sequence</td>
</tr>
<tr>
<td>Gln Pro Ile Asp Asp Thr Glu Ser Asp Ala Thr Ser Val Aen Leu Met</td>
<td>44</td>
<td>PRT</td>
<td>Artificial</td>
<td>Sequence</td>
<td>Alpha factor leader sequence</td>
</tr>
<tr>
<td>Gln Pro Ile Asp Asp Thr Glu Ser Asp Ala Thr Ser Val Aen Leu Met</td>
<td>44</td>
<td>PRT</td>
<td>Artificial</td>
<td>Sequence</td>
<td>Alpha factor leader sequence</td>
</tr>
<tr>
<td>Gln Pro Ile Asp Asp Thr Glu Ser Asp Ala Thr Ser Val Aen Leu Met</td>
<td>44</td>
<td>PRT</td>
<td>Artificial</td>
<td>Sequence</td>
<td>Alpha factor leader sequence</td>
</tr>
<tr>
<td>Gln Pro Ile Asp Asp Thr Glu Ser Asp Ala Thr Ser Val Aen Leu Met</td>
<td>44</td>
<td>PRT</td>
<td>Artificial</td>
<td>Sequence</td>
<td>Alpha factor leader sequence</td>
</tr>
<tr>
<td>Gln Pro Ile Asp Asp Thr Glu Ser Asp Ala Thr Ser Val Aen Leu Met</td>
<td>44</td>
<td>PRT</td>
<td>Artificial</td>
<td>Sequence</td>
<td>Alpha factor leader sequence</td>
</tr>
<tr>
<td>Gln Pro Ile Asp Asp Thr Glu Ser Asp Ala Thr Ser Val Aen Leu Met</td>
<td>44</td>
<td>PRT</td>
<td>Artificial</td>
<td>Sequence</td>
<td>Alpha factor leader sequence</td>
</tr>
<tr>
<td>Ala Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Ser Leu Ala Gly</td>
<td>40</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Ser Leu Ala Gly</td>
<td>40</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Ser Leu Ala Gly</td>
<td>40</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Ser Leu Ala Gly</td>
<td>40</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Ser Leu Ala Gly</td>
<td>40</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Ser Leu Ala Gly</td>
<td>40</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Ser Leu Ala Gly</td>
<td>40</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Ser Leu Ala Gly</td>
<td>40</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
<210> SEQ ID NO 174
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Cleavage site

<400> SEQUENCE: 174

Met Ala Glu Ile Thr Arg Ile Pro Leu Tyr Lys Gly Ser Leu Arg
1   5   10  15
Lys Ala Leu Lys Glu His Gly Leu Leu Glu Asp Phe Leu Gln Lys Gln
20  25
Gln Tyr Gly Ile Ser Ser Lys Phe
30  35  40

<210> SEQ ID NO 175
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 175

gcagtctgac attggtgtga cacagtgc

27

<210> SEQ ID NO 176
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 176

aagcttgcat ttaaatactc gagaacctgta gagggtggtgcc ttg

43

<210> SEQ ID NO 177
<211> LENGTH: 70
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 177

gagaagagag aagotaagt tggtaaatcaaa catctttttggt gatctcaatct tgttgaggct

cctctacctg

60  70

<210> SEQ ID NO 178
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 178

cctctagagaat cctcattttct cacacacaag gtagagasgcc tcaaca

56

<210> SEQ ID NO 179
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer
<400> SEQUENCE: 179
ctasgctgc taagggatt g  
21

<210> SEQ ID NO 180
<211> LENGTH: 83
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 180
aagctcagc tgcataagtctcccaatgg taaaatgcagc aataagtaaagt gcaacatgt ccctccacttc ccctagcagc ctt  
60
83

<210> SEQ ID NO 181
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 181
tctgagctca acaatgtgc acaatgac  
30

<210> SEQ ID NO 182
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 182
aagcttccaa gtctcattct gttgcataag tcctccacat g  
41

<210> SEQ ID NO 183
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 183
aagcttccag tgcataagtt c  
21

<210> SEQ ID NO 184
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 184
gcatgctac ccaattgac acctg  
26

<210> SEQ ID NO 185
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 185
-continued

gcatgcagc cttggttaat cacatccttt gtgg

34

<210> SEQ ID NO: 106
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<221> OTHER INFORMATION: Primer
<400> SEQUENCE: 106

acatggtcga cacaatccttc ttcatctotto atgctttgaggtgtgaaatcct

54

<210> SEQ ID NO: 107
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<221> OTHER INFORMATION: Primer
<400> SEQUENCE: 107
gcatagcct ccaagcctct cctttgagc

29

<210> SEQ ID NO: 108
<211> LENGTH: 387
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<221> OTHER INFORMATION: Insulin fusion protein nucleic acid sequence
<400> SEQUENCE: 108

atgcaactc ttaatcccttt ccccttccac ccctttctttt gtttctgtca tatactttct

60
gctttcaccg atgcgccaccc aatgtagacg ctcgaatcccc agaccaagctc agtgaaacct

120
atgccccgag atactgagag cgcctttggt gcaccaacacc atctggaaggt tctttgaggt

180
gctggattga ttcctcagggc agagagagac ctaatgctttgt taactcaacat

240
ccttttgtcc ctcttctcttc tgttcatcttc gccgattttgc ttttttctttt tggcagagag

300
taacacccctttagctgca gaggattttc gccacaactttt ctctttctctct tggctctctt

360
tcccaatgagcacttattg cacactg

387

<210> SEQ ID NO: 109
<211> LENGTH: 128
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<221> OTHER INFORMATION: Insulin factor protein
<400> SEQUENCE: 109

Met Asn Phe Leu Lys Ser Phe Pro Phe Tyr Ala Phe Leu Cys Phe Gly
1   5   10   15
Gln Tyr Phe Val Ala Val Thr His Ala Gln Pro Ile Asp Asp Thr Glu
20  25  30
Ser Gin Thr Thr Ser Val Asn Leu Met Ala Asp Thr Glu Ser Ala
35  40  45
Phe Ala Thr Gin Thr Arg Ser Gly Gly Leu Asp Val Val Gly Leu Ile
50  55  60
Ser Met Ala Lys Arg Glu Glu Gly Gly Pro Lys Phe Val Asn Gin His
65  70  75  80
Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu
95    90    95
Arg Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gin
100   105   110
Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gin Leu Glu Asn Tyr Cys Asn
115   120   125

-continued

<210> SEQ ID NO: 190
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer
<400> SEQUENCE: 190

ttctgtgacct acaccttg 18

<210> SEQ ID NO: 191
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer
<400> SEQUENCE: 191

aagcctgccag ttcacgtagt 20

<210> SEQ ID NO: 192
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer
<400> SEQUENCE: 192

gcatgcagct gttgggc 17

<210> SEQ ID NO: 193
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer
<400> SEQUENCE: 193

gttggtggtc tggccaa 16

<210> SEQ ID NO: 194
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer
<400> SEQUENCE: 194

gttggtgacg cacactccct cgtaaacccaa cacttgctg 38

<210> SEQ ID NO: 195
<211> LENGTH: 1020
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Insulin fusion protein nucleic acid sequence
<406> SEQUENCE: 196

<410> SEQ ID NO: 196
<411> LENGTH: 257
<412> TYPE: PRT
<413> ORGANISM: Artificial Sequence
<414> FEATURE:
<415> OTHER INFORMATION: Insulin fusion protein

<420> SEQUENCE: 196

Met Ala Asp Thr Ala Arg Gly Thr His His Asp Ile Ile Gly Arg Arg
1   5   10  15
Gln Tyr Pro Met Met Gly Arg Arg Gln Tyr Gln Met Ser Gly
20  25  30
Arg Gly Ser Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Ala Thr
35  40  45
Ala Val Thr Ala Gly Gly Ser Leu Val Leu Ser Ser Leu Thr Leu
50  55  60
Val Gly Thr Val Ile Ala Leu Thr Thr Val Ala Thr Val Leu Val Ile
65  70  75  80
Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Ile
90  95
Thr Gly Phe Leu Ser Ser Gly Gly Ile Ala Ala Ile Thr Val
100 105 110
Phe Ser Trp Ile Tyr Ala Thr Gly Glu His Pro Gly Gly Ser Asp Lys
115 120 125
Leu Asp Ser Ala Arg Met Lys Leu Gly Ser Lys Ala Gln Asp Leu Lys
130 135 140
Asp Arg Ala Gln Tyrr Gly Gln Gly His Thr Gly Gly Glu His Asp
145 150 155 160

-t-continued-
1-40. (canceled)

41. A method for the expression of insulin in plant seeds comprising:
(a) providing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:
(i) a nucleic acid sequence capable of controlling expression in plant seed cells; and
(ii) a nucleic acid sequence encoding an insulin polypeptide;
(b) introducing the chimeric nucleic acid construct into a plant cell; and
(c) growing the plant cell into a mature plant capable of setting seed wherein the seed expresses insulin.

42. A method according to claim 41 wherein the insulin polypeptide accumulates within a membrane enclosed intracellular compartment within the plant cell.

43. A method according to claim 42 wherein said membrane enclosed intracellular compartment is the endoplasmic reticulum (ER) or an ER derived storage vesicle.

44. A method according to claim 41 wherein said chimeric nucleic acid sequence additionally comprises a nucleic acid sequence encoding a polypeptide which is capable of retaining the insulin polypeptide in a membrane enclosed intracellular compartment.

45. A method according to claim 44 wherein said membrane enclosed intracellular compartment is the endoplasmic reticulum (ER) or an ER derived storage organelle.

46. A method according to claim 45 wherein said polypeptide retaining the insulin polypeptide in the ER is selected from the group consisting of KDEL, HDEL, DDEL, ADEL and SDEL.

47. A method according to claim 45 wherein said polypeptide retaining the insulin polypeptide in the ER is selected from the group consisting of SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153 and SEQ ID NO:154.

48. A method according to claim 46 wherein said insulin polypeptide additionally comprises a nucleic acid sequence encoding a signal peptide.

49. A method according to claim 48 wherein said signal peptide is a tobacco pathogenesis related protein (PR-S) signal sequence.

50. A method according to claim 48 wherein said signal sequence is SEQ ID NO:161.

51. A method according to claim 45 wherein said ER-derived storage organelle is an oil body.

52. A method according to claim 45 wherein said polypeptide retaining the insulin polypeptide in an ER derived storage organelle is an oil body protein.

53. A method according to claim 42 wherein oil body protein is selected from the group of oil body proteins consisting of oleosin, caloesin and stereoleosin.

54. A method according to claim 42 wherein said oil body protein is selected from the group consisting of SEQ ID NO:156, SEQ ID NO:157 and SEQ ID NO:158, SEQ ID NO:159, and SEQ ID NO:160.

55. A method according to claim 44 wherein said chimeric nucleic acid additionally contains a nucleic acid sequence encoding a stabilizing protein fused in reading frame to the nucleic acid sequence encoding insulin.

56. A method according to claim 55 wherein said chimeric nucleic acid additionally contains a nucleic acid sequence encoding a signal peptide sequence fused in reading frame to the nucleic acid sequence encoding insulin.

57. A method according to claim 55 wherein said signal peptide is a tobacco pathogenesis related protein (PR-S) signal sequence.

58. A method according to claim 57 wherein said signal peptide is SEQ ID NO:161.

59. A method according to claim 55 wherein said nucleic acid encoding said stabilizing protein permits association of the insulin polypeptide with the oil body upon harvesting and grinding of the seed.

60. A method according to claim 59 wherein said stabilizing protein encodes a single chain antibody with specificity to an oil body.

61. A method according to claim 55 wherein the nucleic acid sequence encoding a stabilizing protein fused in reading frame to the nucleic acid sequence encoding insulin is selected from the group of polypeptides consisting of a single chain antibody and cholera toxin B subunit.

62. A method according to claim 43 wherein the chimeric nucleic acid sequence is introduced in into the plant cell under nuclear genomic integration conditions.
63. A method according to claim 41 wherein said nucleic acid sequence capable of controlling expression in plant seeds is a seed-preferred promoter.

64. A method according to claim 63 wherein the seed-preferred promoter is a phaseolin promoter.

65. A method according to claim 41 wherein the nucleic acid sequence encoding insulin is selected from the group of nucleic acid sequences consisting of human insulin, porcine insulin and bovine insulin.

66. A method according to claim 41 wherein the nucleic acid encoding insulin is a mini-insulin.

67. A method to claim 41 wherein the nucleic acid sequence encoding insulin is optimized for plant codon usage.

68. A method for obtaining plant seeds comprising insulin comprising:
   (a) providing a chimeric nucleic acid construct comprising
       in the 5' to 3' direction of transcription as operably linked components:
       (i) a nucleic acid sequence capable of controlling expression in plant seed cells; and
       (ii) a nucleic acid sequence encoding an insulin polypeptide;
   (b) introducing the chimeric nucleic acid construct into a plant cell;
   (c) growing the plant cell into a mature plant capable of setting seed; and
   (d) obtaining seeds from said plant wherein the seed comprises insulin.

69. A method according to claim 68 wherein at least 0.1% of the total soluble protein present in the seed is insulin.

70. A plant capable of setting seed comprising a chimeric nucleic acid sequence comprising in the 5' to 3' direction of transcription:
   (a) a first nucleic acid sequence capable of controlling expression in a plant seed cell operatively linked to;
   (b) a second nucleic acid sequence encoding an insulin polypeptide, wherein the seed contains insulin.

71. A plant according to claim 68 wherein the chimeric nucleic acid sequence is integrated in the plant's nuclear genome.

72. A plant according to claim 60 wherein the plant is an Arabidopsis, flax or saflower plant.

73. A plant seed comprising a chimeric nucleic acid sequence comprising in the 5' to 3' direction of transcription:
   (a) a first nucleic acid sequence capable of controlling expression in a plant seed cell operatively linked to;
   (b) a second nucleic acid sequence encoding an insulin polypeptide.

74. Plant seed according to claim 73 wherein at least 0.1% of the total soluble protein present in the seed is insulin.

75. A nucleic acid sequence encoding insulin linked to nucleic acid sequence comprising a promoter capable of controlling expression in a plant seed cell.

76. A nucleic acid sequence according to claim 75 wherein said promoter is a seed preferred promoter.

77. A nucleic acid sequence according to claim 76 wherein said seed preferred promoter is a phaseolin promoter.

78. A nucleic acid sequence according to claim 75 wherein said nucleic acid sequence additionally contains a sequence capable of retaining the insulin polypeptide in a membrane enclosed intracellular compartment.

79. A nucleic acid sequence according to claim 75 wherein said nucleic acid sequence additionally contains a sequence capable of retaining the insulin polypeptide in the ER or an ER derived storage organelle.

80. Use of a plant seed prepared according to claim 40 to obtain substantially pure insulin.