



(86) Date de dépôt PCT/PCT Filing Date: 2012/05/10	(51) Cl.Int./Int.Cl. <i>C07K 14/415</i> (2006.01), <i>C12N 15/82</i> (2006.01), <i>A01H 5/00</i> (2006.01)
(87) Date publication PCT/PCT Publication Date: 2012/11/29	
(85) Entrée phase nationale/National Entry: 2013/11/19	(71) Demandeur/Applicant: E. I. DU PONT DE NEMOURS & COMPANY, US
(86) N° demande PCT/PCT Application No.: US 2012/037360	(72) Inventeurs/Inventors: BOURETT, TIMOTHY M., US; LU, ALBERT L., US; CHUN, GARY C., US
(87) N° publication PCT/PCT Publication No.: 2012/161982	
(30) Priorité/Priority: 2011/05/23 (US61/488,952)	(74) Agent: TORYS LLP

(54) Titre : PEPTIDES DE TRANSIT DE CHLOROPLASTE ET PROCEDES D'UTILISATION DE CEUX-CI
(54) Title: CHLOROPLAST TRANSIT PEPTIDES AND METHODS OF THEIR USE

Figure 1 **Strategy for CTP Development** ^{AA}

CTP framework	Uncharged N-terminus		R rich C terminus Amphipathic helix	BB
msCTP1	OsdXP	ssRUBISCO (Zm)	β Gluc	
msCTP2	NADP	OsSOD	Os soluble starch	CC
msCTP3	OsNADP-malic enzyme	OsPHD Aldolase 2	TRX	
msCTP4	NADP/OsNADP- malic enzyme	ssRUBISCO (Zm)	Os sol. starch/TRX	
msCTP5	NADP/OsNADP	OsAPx05/ ssRUBISCO	OsSOD/OsPGDK	
msCTP6	Most freq AA	Most freq AA	Most freq AA	
msCTP7	Most freq AA	Most freq AA	Most freq AA	

(57) **Abrégé/Abstract:**
Methods and compositions are provided for targeting a polypeptide of interest to a chloroplast. Recombinant polynucleotides comprising a nucleotide sequence encoding a chimeric chloroplast transit peptide (CTP) operably linked to a heterologous

(57) **Abrégé(suite)/Abstract(continued):**

polynucleotide of interest are provided. In specific embodiments, the chimeric CTP comprises an N-terminal domain operably linked to a central domain operably linked to a C-terminal domain of a CTP to form a chimeric chloroplast transit peptide having CTP activity. Recombinant polypeptides encoding the same, as well as, cells, plant cells, plants and seeds are further provided which comprise the recombinant polynucleotides. Methods of use of the various sequences are also provided.

- (51) International Patent Classification:
C07K 14/415 (2006.01) A01H 5/00 (2006.01)
C12N 15/82 (2006.01)

(21) International Application Number:
PCT/US2012/037360

(22) International Filing Date:
10 May 2012 (10.05.2012)

(25) Filing Language:
English

(26) Publication Language:
English

(30) Priority Data:
61/488,952 23 May 2011 (23.05.2011) US

(71) Applicants (for all designated States except US): E. I. DUPONT DE NEMOURS & COMPANY [US/US]; 1007 Market Street, Wilmington, Delaware 19898 (US). CHUN, Gary C. [US/US]; c/o Pioneer Hi-Bred International, Inc., 7250 N.W. 62nd Avenue, Johnston, Iowa 50131-0552 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BOURETT, Timothy M. [US/US]; c/o Pioneer Hi-Bred International, Inc., 7250 N.W. 62nd Avenue, Johnston, Iowa 50131-0552 (US). LU, Albert L. [CA/US]; c/o Pioneer Hi-Bred International, Inc., 7250 N.W. 62nd Avenue, Johnston, Iowa 50131-0552 (US).

(74) Agent: LAPPEGARD, Kathryn K.; Pioneer Hi-Bred International, Inc., 7250 N.W. 62nd Avenue, Johnston, Iowa 50131-0552 (US).

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

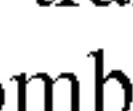
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Declarations under Rule 4.17:
— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))

[Continued on next page]

(54) Title: CHLOROPLAST TRANSIT PEPTIDES AND METHODS OF THEIR USE

Figure 1

Strategy for CTP Development

CTP framework	Uncharged N-terminus		R rich C-terminus Amphipathic helix
msCTP1	OsdXP	ssRUBISCO (Zm)	
msCTP2	NADP	OsSOD	Os soluble starch
msCTP3	OsNADP-malic enzyme	OsPHD Aldolase 2	
msCTP4	NADP/OsNADP-malic enzyme	ssRUBISCO (Zm)	Os sol. starch/TRX
msCTP5	NADP/OsNADP	OsAPx05/ssRUBISCO	OsSOD/OsPGDK
msCTP6	Most freq AA	Most freq AA	Most freq AA
msCTP7	Most freq AA	Most freq AA	Most freq AA

(57) Abstract: Methods and compositions are provided for targeting a polypeptide of interest to a chloroplast. Recombinant polynucleotides comprising a nucleotide sequence encoding a chimeric chloroplast transit peptide (CTP) operably linked to a heterologous polynucleotide of interest are provided. In specific embodiments, the chimeric CTP comprises an N-terminal domain operably linked to a central domain operably linked to a C-terminal domain of a CTP to form a chimeric chloroplast transit peptide having CTP activity. Recombinant polypeptides encoding the same, as well as, cells, plant cells, plants and seeds are further provided which comprise the recombinant polynucleotides. Methods of use of the various sequences are also provided.

WO 2012/161982 A1



Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

CHLOROPLAST TRANSIT PEPTIDES AND METHODS OF THEIR USE

REFERENCE TO A SEQUENCE LISTING SUBMITTED AS A TEXT FILE VIA EFS-WEB

The official copy of the sequence listing is submitted electronically via EFS-
5 Web as an ASCII formatted sequence listing with a file named 414530SEQLIST.txt,
created on March 26, 2012, and having a size of 34 KB and is filed concurrently with
the specification. The sequence listing contained in this ASCII formatted document is
part of the specification and is herein incorporated by reference in its entirety.

10 FIELD OF THE INVENTION

This invention is in the field of molecular biology. More specifically, this
invention pertains to targeting sequences of interest to a chloroplast by employing
novel chloroplast transit peptides.

15 BACKGROUND OF THE INVENTION

Plastids are a heterogeneous family of organelles found ubiquitously in plants
and algal cells. Most prominent are the chloroplasts, which carry out such essential
processes as photosynthesis and the biosynthesis of fatty acids as well as of amino
acids. Chloroplasts are complex organelles composed of six distinct suborganellar
20 compartments: three different membranes (the two envelope membranes and the
internal thylakoid membranes) and three compartments (the innermembrane space of
the envelope, the stroma and the thylakoid lumen). More than 98% of all plastid
proteins are translated on cytosolic ribosomes. Such proteins are posttranslationally
targeted to and imported into the organelle. For a review, see, Jarvis *et al.* (2008)
25 *New Phytologist* 179:257-285. Such translocation is mediated by multiprotein
complexes in the outer and inner envelope membranes called TOC (Translocon at the
Outer envelope membrane of Chloroplasts) and TIC (Translocon at the Inner
envelope membrane of Chloroplasts). See, Soll *et al.* (2004) *Nature Reviews.*
Molecular Cell Biology 5:198-208, Bedard *et al.* (2005) *Journal of Experimental*
30 *Botany* 56:2287-2320, Kessler *et al.* (2006) *Traffic* 7:248-257, and Smith *et al.* (2006)

Canadian Journal of Botany 84:531-542. Once the chloroplast precursor enters the stroma, the transit peptide is cleaved off, leaving the remaining part of the protein to take on its final conformation or engage one of a number of different sorting pathways. See, Keegstra *et al.* (1999) *Plant Cell* 11:557-570, Jarvis *et al.* (2004) and
5 *Gutensohn et al.* (2006) *Journal of Plant Physiology* 163:333-347.

Methods and compositions are needed to allow heterologous polypeptides to be targeted to the chloroplast.

BRIEF SUMMARY OF THE INVENTION

10 Methods and compositions are provided for targeting a polypeptide of interest to a chloroplast. Recombinant polynucleotides comprising a nucleotide sequence encoding a chimeric chloroplast transit peptide (CTP) operably linked to a heterologous polynucleotide of interest are provided. In specific embodiments, the chimeric CTP comprises an N-terminal domain operably linked to a central domain
15 operably linked to a C-terminal domain of a CTP to form a chimeric chloroplast transit peptide having CTP activity. Recombinant polypeptides encoding the same, as well as, cells, plant cells, plants and seeds are further provided which comprise the recombinant polynucleotides. Methods of use of the various sequences are also provided.

20

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 provides a strategy for developing the recombinant chloroplast transit peptides provided herein. The origin of each segment of the CTP framework for the recombinant chloroplast transit peptides is provided.

25 Figure 2 provides an amino acid alignment of chloroplast transit peptides from various monocot plants. The most frequent amino acids are highlighted

.

DETAILED DESCRIPTION OF THE INVENTION

The present inventions now will be described more fully hereinafter with
30 reference to the accompanying drawings, in which some, but not all embodiments of the inventions are shown. Indeed, these inventions may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will satisfy applicable legal requirements. Like numbers refer to like elements throughout.

Many modifications and other embodiments of the inventions set forth herein will come to mind to one skilled in the art to which these inventions pertain having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is to be understood that the inventions are not to be limited to
5 the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended claims. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation.

10 *I. Compositions*

A. Overview

In the production of transgenic plants it is often useful to direct foreign proteins to specific subcellular locations, e.g., the plastid, vacuole, mitochondria, or ER. When the gene is translated, the resulting protein has the transit peptide fused to
15 the amino terminus of the protein of interest, and thus the protein is directed to the desired subcellular compartment. Of particular interest is the identification of transit peptides that will direct transport to a plastid. As used herein, a “plastid” refers to an organelle present in plant cells that stores and manufactures chemical compounds used by the cell, such as starch, fatty acids, terpenes, and that has been derived from a
20 proplastid. Thus, plastids of plants typically have the same genetic content. Plastids include chloroplasts, which are responsible for photosynthesis, amyloplasts, chromoplasts, statoliths, leucoplasts, elaioplasts, and proteinoplasts. Plastids contain photosynthetic machinery and many additional biosynthetic enzymes including those leading to the production of fatty acids, amino acids, carotenoids, terpenoids, and
25 starch. Thus, there is a need for the ability to target polypeptides of interest to plastids to modulate or alter the physiological processes that occur within these organelles. In addition, some polypeptides are toxic when expressed recombinantly in the cytoplasm. Because plastids are subcompartments, it is possible to target polypeptides of interest to the plastids to sequester them from the cytoplasm, and thus
30 allow for higher expression levels. Furthermore, expression of recombinant polypeptides in plastids may facilitate isolation of the polypeptide for various applications. As discussed in further detail herein, novel chimeric chloroplast transit peptides are provided which can be used in plastid targeting.

The compositions provided herein include recombinant polynucleotides comprising a nucleotide sequence encoding a novel chloroplast transit peptide (CTP) operably linked to a nucleotide sequence encoding a polypeptide of interest. The CTP-encoding sequences disclosed herein, when assembled within a DNA construct
5 such that the CTP-encoding sequence is operably linked to a nucleotide sequence encoding the polypeptide of interest, facilitate co-translational or post-translational transport of the peptide of interest to the chloroplast of a plant cell.

B. Chloroplast Transit Peptides

10 Chloroplasts are organelles found in plant cells and eukaryotic algae that conduct photosynthesis. The chloroplast is a complex cellular organelle composed of three membranes: the inner envelope membrane, the outer envelope membrane, and the thylakoid membrane. The membranes together enclose three aqueous compartments termed the intermediate space, the stroma, and the thylakoid lumen.
15 While chloroplasts contain their own circular genome, many constituent chloroplast proteins are encoded by the nuclear genes and are cytoplasmically-synthesized as precursor forms which contain N-terminal extensions known as chloroplast transit peptides (CTPs). As used herein, the term "chloroplast transit peptide" or "CTP" refers to the N-terminal portion of a chloroplast precursor protein and influences the
20 recognition of the chloroplast surface and mediates the post-translational translocation of pre-proteins across the chloroplast envelope and into the various subcompartments within the chloroplast (e.g. stroma, thylakoid and thylakoid membrane). Thus, as used herein, a polypeptide having "CTP activity" comprises a polypeptide which when operably linked to the N-terminal region of a protein of interest facilitates
25 translocation of the polypeptide of interest to the chloroplast.

Assays to determine the efficiency by which the CTP sequences provided herein target a protein of interest to a chloroplast are known. See, for example, Mishkind *et al.* (1985) *J. of Cell. Biol.* 100:226-234, which is herein incorporated by reference in its entirety. A reporter gene such as glucuronidase (GUS),
30 chloramphenicol acetyl transferase (CAT), or green fluorescent protein (GFP) is operably linked to the CTP sequence. This fusion is placed behind the control of a suitable promoter, ligated into a transformation vector, and transformed into a plant or plant cell. Following an adequate period of time for expression and localization into the chloroplast, the chloroplast fraction is extracted and reporter activity assayed. The

ability of the CTP sequences to target and deliver the reporter protein to the chloroplast can be compared to other known CTP sequences. See, de Castro Silva Filho *et al.* (1996) *Plant Mol. Biol.* 30: 769-780. Protein import can also be verified *in vitro* through the addition of proteases to the isolated chloroplast fraction. Proteins which were successfully imported into the chloroplast are resistant to the externally added proteases whereas proteins that remain in the cytosol are susceptible to digestion. Protein import can also be verified by the presence of functional protein in the chloroplast using standard molecular techniques for detection, by evaluating the phenotype resulting from expression of a chloroplast targeted protein, or by microscopy.

a. Chimeric Chloroplast Transit Peptides

Recombinant polynucleotides encoding a chimeric CTP operably linked to a heterologous polynucleotide of interest are provided herein. The chimeric CTPs comprise heterologous domains of known or predicted CTPs which, when operably linked, have CTP activity.

CTPs have a preference for hydroxylated amino acids (S, T, P) and lack acidic residues. They share a common structural framework comprising an uncharged N-terminal region (“N-terminal domain”), a central region (“central domain”), and a basic arginine-rich amphipathic C-terminal region (“C-terminal domain”). The domain framework structure of CTPs is provided in Figure 1. Thus, the CTPs provided herein comprise 3 domains, an N-terminal domain, a central domain and a C-terminal domain.

As used herein, “N-terminal domain” refers to an N-terminal hydrophobic region of a CTP comprising uncharged amino acids. The N-terminal domain can comprise at least 5-10, 5-11, 5-12, 5-13, 5-14, 5-15, 5-16, 5-17, 5-18, 5-19, 5-20 or more amino acids from the N-terminus of a CTP sequence generally beginning with MA and ending in G/P. The N-terminal domain can comprise additional sequences, such as linker sequences, such that when operably linked to a central domain and C-terminal domain reconstitutes a CTP having CTP activity.

A “central domain” as used herein, refers to a central region of a CTP comprising an amino acid sequence lacking acidic amino acids and enriched in serine, threonine, lysine and arginine. The central domain can comprise at least 4-5, 4-6, 4-7, 4-8, 4-9, 4-10, 4-11, 4-12, 4-13, 4-14, 4-15 or more amino acids from the central

region of a CTP sequence. The central domain can also comprise additional sequences such as linker sequences, such that when operably linked to an N-terminal domain and a C-terminal domain reconstitutes a CTP having CTP activity.

As used herein, “C-terminal domain” refers to a C-terminal region of a CTP comprising an amino acid sequence which is basic, arginine-rich and predicted to form an amphiphilic beta strand. The C-terminal domain can comprise at least 5-10, 5-15, 5-16, 5-17, 5-18, 5-19, 5-20, 5-21, 5-22, 5-23, 5-24, 5-25, 5-26, 5-27, 5-28, 5-29, 5-30 or more amino acids from the C-terminal region of a CTP sequence. The C-terminal domain can comprise additional sequences such as linker sequences, such that when operably linked to an N-terminal domain and a central domain reconstitutes a CTP having CTP activity.

Non-limiting examples of domains for various CTPs are set forth in SEQ ID NOS: 24-43 and summarized in Table 4.

A “chimeric CTP” provided herein comprises an N-terminal domain, a central domain, and a C-terminal domain from any CTP in which the sequence of at least one of the domains is heterologous to the sequence of the other domains and whereby the domains, when operably linked, reconstitute a CTP with CTP activity. As used herein the term “chimeric” refers to a sequence having two or more heterologous sequences linked together. As used herein, “heterologous” in reference to a sequence is a sequence that originates from a foreign species, for example, from a different CTP, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a heterologous domain is intended at least one of the CTP domains is not from the same CTP, but could be from a different CTP of the same plant species or a different plant species. The chimeric CTPs provided herein can vary in length from about 30, 35, 40, 45, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65 or more amino acid residues in length such that it comprises an N-terminal domain, a central domain, and a C-terminal domain and retains CTP activity.

The domains of the chimeric CTPs can be from any known or predicted CTP sequence. For example, in some embodiments, the chimeric CTP can comprise, but is not limited to, an N-terminal domain, a central domain or a C-terminal domain from a CTP from *Oryza sativa* 1-deoxy-D xyulose-5-Phosphate Synthase, *Oryza sativa* - Superoxide dismutase, *Oryza sativa* -soluble starch synthase, *Oryza sativa* -NADP-dependent Malic acid enzyme, *Oryza sativa* -Phospho-2-dehydro-3-deoxyheptonate

Aldolase 2, *Oryza sativa* -L-Ascorbate peroxidase 5, *Oryza sativa* -Phosphoglucan water dikinase, *Zea Mays* ssRUBISCO, *Zea Mays* -beta-glucosidase, *Zea Mays* -Malate dehydrogenase, *Zea Mays* Thioredoxin M-type or active variants thereof.

In specific, non-limiting, embodiments, the N-terminal domain of the chimeric CTP is from a CTP from *Oryza sativa* 1-deoxy-D xyulose-5-Phosphate Synthase, *Oryza sativa* -NADP-dependent Malic acid enzyme, *Zea Mays* -Malate dehydrogenase or active variants thereof. In other non-limiting embodiments, the central domain of the chimeric CTP is from a CTP from *Oryza sativa* -Superoxide dismutase, *Oryza sativa* -Phospho-2-dehydro-3-deoxyheptonate Aldolase 2, *Oryza sativa* -L-Ascorbate peroxidase 5, *Zea Mays* ssRUBISCO or active variants thereof. In yet other non-limiting embodiments, the C-terminal domain of the chimeric CTP is from a CTP from *Oryza sativa* -soluble starch synthase, *Oryza sativa* -Superoxide dismutase, *Oryza sativa* -Phosphoglucan water dikinase, *Zea Mays* Thioredoxin M-type, *Zea Mays* -beta-glucosidase or active variants thereof. Non-limiting examples of various CTPs, N-terminal domains, central domains and C-terminal domains of CTPs are set forth in SEQ ID NOS: 13-43.

In one specific embodiment, the chimeric CTP comprises the N-terminal domain from the *Oryza sativa* 1-deoxy-D xyulose-5-Phosphate Synthase CTP or an active variant thereof, the central domain from the *Zea Mays* ssRUBISCO CTP or an active variant thereof, and the C-terminal domain of the *Zea Mays* -beta-glucosidase CTP or an active variant thereof. In another specific embodiment, the chimeric CTP comprises the N-terminal domain from the *Zea Mays* -Malate dehydrogenase CTP or an active variant thereof, the central domain from the *Oryza sativa* -Superoxide dismutase CTP or an active variant thereof, and the C-terminal domain from the *Oryza sativa* -soluble starch synthase CTP or an active variant thereof. In yet another specific embodiment, the chimeric CTP comprises the N-terminal domain from the *Oryza sativa* -NADP-dependent Malic acid enzyme CTP or active variant thereof, the central domain from the *Oryza sativa* -Phospho-2-dehydro-3-deoxyheptonate Aldolase 2 CTP or an active variant thereof, and the C-terminal domain from the *Zea Mays* Thioredoxin M-type CTP or an active variant thereof.

Examples of chimeric CTPs are set forth in the amino acid sequences of SEQ ID NO: 1 (msCTP1) or an active variant or fragment thereof, SEQ ID NO: 2 (msCTP2) or an active variant or fragment thereof and SEQ ID NO: 3 (msCTP3) or

an active variant or fragment thereof. The domain structures of the various CTPs provided herein are depicted in Figure 1.

The chimeric CTPs provided herein can also comprise chimeric domains. As used herein, a “chimeric domain” refers to an N-terminal domain, central domain, or C-terminal domain of a CTP which comprises portions of two or more heterologous N-terminal domain, central domain, or C-terminal domain sequences fused together to reconstitute a complete domain. For example, a chimeric domain (i.e. a “chimeric N-terminal domain”, “chimeric central domain” or “chimeric C-terminal domain”) provided herein can comprise at least 2, 3, 4 or more heterologous CTP sequences fused together such that the chimeric domain, when incorporated in a chimeric CTP, has CTP activity.

In some embodiments, the chimeric CTPs can comprise at least 1, 2 or 3 chimeric domains. In specific embodiments, at least one portion of the chimeric N-terminal domain is from the N-terminal domain of the *Oryza sativa* -NADP-dependent Malic CTP, *Zea Mays* -Malate dehydrogenase CTP or active variants thereof. In other embodiments at least one portion of the chimeric central domain is from the central domain of the *Oryza sativa* -NADP-dependent Malic CTP, *Zea Mays* -Malate dehydrogenase CTP or active variants thereof. In yet other embodiments, at least one portion of the chimeric C-terminal domain is from the C-terminal domain of the *Oryza sativa* -soluble starch synthase CTP, *Zea Mays* Thioredoxin M-type CTP, *Oryza sativa* -Superoxide dismutase CTP, *Oryza sativa* -Phosphoglucan water dikinase CTP or active variants thereof.

In a specific embodiment, the chimeric CTP comprises a chimeric N-terminal domain comprising a portion of the N-terminal domain from the *Zea Mays*- Malate dehydrogenase CTP fused in frame to a portion of the N-terminal domain of the *Oryza sativa*- NADP-dependent Malic acid enzyme CTP, a central domain from the *Zea Mays* ssRUBISCO CTP, and a chimeric C-terminal domain comprising a portion of the C-terminal domain from the *Oryza sativa*- soluble starch synthase CTP fused in frame to a portion of the C-terminal domain from the *Zea Mays* Thioredoxin M-type CTP, wherein the chimeric CTP has CTP activity.

In another specific embodiment, the chimeric CTP comprises a chimeric N-terminal domain comprising a portion of the *Zea Mays*- Malate dehydrogenase CTP fused in frame to a portion of the *Oryza sativa*- NADP-dependent Malic acid enzyme CTP, a chimeric central domain comprising a portion of the *Oryza sativa*-L-Ascorbate

peroxidase 5 CTP fused in frame to a portion of the *Zea Mays* ssRUBISCO CTP, and a chimeric C-terminal domain comprising a portion of the *Oryza sativa*- Superoxide dismutase CTP fused in frame to a portion of the *Oryza sativa*-Phosphoglucan water dikinase CTP, wherein the chimeric CTP has CTP activity.

5 Exemplary CTPs comprising chimeric domains are set forth in the amino acid sequences of SEQ ID NO:4 (msCTP4) or an active or fragment variant thereof and SEQ ID NO:5 (msCTP5) or an active variant or fragment thereof. Examples of chimeric CTP domain structures are provided in Figure 1.

b. Consensus Chloroplast Transit Peptides

10 While the chimeric CTPs described in the previous section employed a domain approach for CTP design, it is recognized that other approaches can be used to design chloroplast transit peptides having CTP activity. Provided herein are recombinant polynucleotides encoding CTPs with sequences based on the alignment of various known monocot CTP sequences and the most frequent amino acids at each
15 position operably linked to a heterologous polynucleotide encoding a polypeptide of interest. Figure 2 provides the alignment of the various monocot CTP sequences used to determine the most frequent amino acids. The various CTPs were aligned based on the structural framework of the different domains as described elsewhere herein and a consensus CTP sequence is provided.

20 In one embodiment, a CTP is provided comprising the following CTP consensus sequence:

MXXXVXXAAAXXXSXPRXXXGXXXXXXXXXXXXXXXXXXXXAA
XX RXXX :: (SEQ ID NO:11)

or an active variant thereof, where the X indicates any amino acid.

25 Based on the consensus sequence, various CTPs can be constructed such that the resulting CTP has CTP activity. In some cases, a dominant amino acid residue may not be apparent. In these cases, one of the more frequent amino acid residues can be chosen to be incorporated into the sequence. It is recognized that many CTP sequences can be provided from the consensus sequence disclosed herein.

30 In one non-limiting embodiment, a CTP is provided having the following sequence:

MALASVMAAAAASVVSFPAGRGSGGSSVLRSRALSLAGSRRSAAAV
RR LAL :: (SEQ ID NO:6) (msCTP6)

or an active variant or fragment thereof. In another non-limiting embodiment, a CTP sequence is provided having the following sequence:

MAVATVLAAAALAAVSPPGLRSSLGFPVRRSLPSAARGGSPAATRR

CR AA :: (SEQ ID NO:7) (msCTP7)

5 or an active variant or fragment thereof.

c. Other components of the CTPs Provided Herein

It is recognized that the various CTPs disclosed herein can be modified to improve and/or alter the translocation of the polypeptide of interest into the
10 chloroplast. For example, the CTP can contain additional regions that alter or improve the interactions with cytosolic factors that facilitate the passage of precursors from the ribosomes to the chloroplast surface. See, for example, Hiltbrunner *et al.* (2001) *Journal of Cell Biology* 154:309-316, Jackson-Constan *et al.* (2001) *Biochimica et Biophysica Acta* 1541:102-113, both of which are herein incorporated
15 by reference. Other regions can be employed to increase the efficiency of chloroplast import. See, for example, May *et al.* (2000) *Plant Cell* 12:53-64, Qbadou *et al.* (2006) *EMBO Journal* 25:1837-1837 and Sohrt *et al.* (2000) *Journal of Cell Biology* 148:1213-1221, herein incorporated by reference. Such regions may be native (derived from a region of the same chloroplast targeted polypeptide as the CTP) or
20 heterologous to the operably linked CTP provided herein.

The various CTPs disclosed herein can further comprise additional sequences which modulate the final location of the polypeptide of interest in the chloroplast. For example, the various CTPs disclosed herein could further comprise a thylakoid lumen targeting domain. Proteins to be targeted to the thylakoid lumen bear an additional
25 cleavable targeting signal, which like the transit peptide, is removed once translocation is complete. The luminal targeting peptides are extremely similar to the signal peptides that mediate inner membrane transport in bacteria. See, for example, Keegstra *et al.* (1999) *Plant Cell* 11:557-570, Jarvis (2004) *Current Biology* 14: R1064-R1077, Gutensohn *et al.* (2006) *Journal of Plant Physiology* 163:333-347, and
30 Jarvis (2008) *New Phytologist* 179:257-285, all of which are incorporated by reference in their entirety, which discuss the various sorting pathways in a chloroplast. Such regions which modulate the location of the polypeptide of interest in a chloroplast may be native (derived from a region of the same chloroplast targeted polypeptide as the CTP) or heterologous to the operably linked CTP provided herein.

The term "chloroplast transit peptide cleavage site" refers to a site between two amino acids in a chloroplast-targeting sequence at which the chloroplast processing protease acts. CTPs target the desired protein to the chloroplast and can facilitate the protein's translocation into the organelle. This is accompanied by the cleavage of the transit peptide from the mature polypeptide or protein at the appropriate transit peptide cleavage site by a chloroplast processing protease. Accordingly, a CTP can further comprise a suitable cleavage site for the correct processing of the pre-protein to the mature polypeptide contained within the chloroplast. In one non-limiting example, the CTP cleavage site is within the N-terminus of the IP2-127 protein between amino acid 15 and 16 in SEQ ID NO: 12, when msCTP4 was used in combination with IP2-127. As discussed above, the sequences beyond the cleaved fragments may be important for localization/transport efficiency and be employed with any of the CTPs disclosed herein.

d. Polynucleotide and Polypeptide Fragments and Variants of CTPs

Fragments and variants of the CTP-sequences (i.e. SEQ ID NOS: 1-7 and 13-23) are also encompassed herein. By "fragment" is intended a portion of the polynucleotide or a portion of the amino acid sequence and hence protein encoded thereby. Fragments of a polynucleotide may encode protein fragments that retain CTP activity when reconstituted in a CTP and are thus capable of facilitating the translocation of a polypeptide of interest into the chloroplast of a plant. Alternatively, fragments of a polynucleotide that are useful as a hybridization probe generally do not encode fragment proteins retaining biological activity. Thus, fragments of a nucleotide sequence may range from at least about 10, 20, 30, 40, 50, 60, 70, 80 nucleotides or up to the full length CTP.

A fragment of a polynucleotide that encodes a biologically active portion of a CTP-polypeptide will encode at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60 contiguous amino acids, or up to the total number of amino acids present in any one of SEQ ID NOS: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or of the various chimeric CTPs disclosed herein. Fragments of a CTP-encoding sequence that are useful as hybridization probes or PCR primers generally need not encode a biologically active portion of a CTP.

"Variant" CTP is intended to mean a protein derived from the CTP (i.e. SEQ ID NOS: 1-7 and 13-23) by deletion (i.e., truncation at the 5' and/or 3' end) and/or a deletion or addition of one or more amino acids at one or more internal sites in the CTP and/or substitution of one or more amino acids at one or more sites in the CTP, and/or substitution of one or more of the N-terminal, central, or C-terminal domains of the CTP and/or substitution of a portion of one or more of the N-terminal, central, or C-terminal domains of the CTP. Variant proteins encompassed are biologically active, that is they continue to possess the desired biological activity of the CTP, that is, have CTP activity when reconstituted in a CTP. Such variants may result from, for example, genetic polymorphism or from human manipulation.

For polynucleotides encoding a CTP, a variant comprises a polynucleotide having a deletion (i.e., truncations) at the 5' and/or 3' end and/or a deletion and/or addition of one or more nucleotides at one or more internal sites within the polynucleotide and/or a substitution of one or more nucleotides at one or more sites in the polynucleotide and/or substitution of one or more of the N-terminal, central, or C-terminal domains of the polynucleotide encoding the CTP and/or substitution of a portion of one or more of the N-terminal, central, or C-terminal domains of the polynucleotide encoding the CTP. Variant polynucleotides also include synthetically derived polynucleotides, such as those generated, for example, by using site-directed mutagenesis or gene synthesis but which still encode a CTP.

Biologically active variants of a CTP provided herein (and the polynucleotide encoding the same) will have at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the polypeptide of any one of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or to any N-terminal domain or portion thereof, any central domain or portion thereof or any C-terminal domain or portion thereof from any one of SEQ ID NOS: 1- 7, 13-43 or any of the other CTPs disclosed herein.

The CTP-sequences and the active variants and fragments thereof may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants and fragments of the CTPs can be prepared by mutations in the DNA. Methods for mutagenesis and polynucleotide alterations are well known in the art. See, for example, Kunkel (1985) *Proc. Natl. Acad. Sci. USA*

82:488-492; Kunkel *et al.* (1987) *Methods in Enzymol.* 154:367-382; U.S. Patent No. 4,873,192; Walker and Gastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York) and the references cited therein.

Guidance as to appropriate amino acid substitutions that do not affect biological
5 activity of the protein of interest may be found in the model of Dayhoff *et al.* (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, D.C.), herein incorporated by reference. Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be optimal.

Obviously, the mutations that will be made in the DNA encoding the variant
10 must not place the sequence out of reading frame and optimally will not create complementary regions that could produce secondary mRNA structure. See, EP Patent Application Publication No. 75,444.

Variant polynucleotides and proteins also encompass sequences and proteins derived from a mutagenic and recombinogenic procedure such as DNA shuffling.
15 With such a procedure, one or more different CTP-sequences can be manipulated to create a new CTP possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*. For example, using this
20 approach, sequence motifs encoding a domain of interest may be shuffled between the CTP sequences disclosed herein and other known CTPs to obtain a new polynucleotide coding for a polypeptide with an improved property of interest, such as an improved efficiency of transport to the chloroplast. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751; Stemmer (1994) *Nature* 370:389-391; Crameri *et al.* (1997) *Nature Biotech.* 15:436-438; Moore *et al.* (1997) *J. Mol. Biol.* 272:336-347; Zhang *et al.* (1997) *Proc. Natl. Acad. Sci. USA* 94:4504-4509; Crameri *et al.* (1998) *Nature* 391:288-291; and U.S. Patent Nos. 5,605,793 and 5,837,458.

30

e. Sequence Comparisons

The following terms are used to describe the sequence relationships between two or more polynucleotides or polypeptides: (a) "reference sequence", (b) "comparison window", (c) "sequence identity", and, (d) "percent sequence identity."

5 (a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence or protein sequence.

10 (b) As used herein, "comparison window" makes reference to a contiguous and specified segment of a polypeptide sequence, wherein the polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two polypeptides. Generally, the comparison window is at least 5, 10, 15, or 20 contiguous amino acids in length, or it can be 30, 40, 50, 100, or longer.
15 Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polypeptide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent sequence identity between any two sequences can
20 be accomplished using a mathematical algorithm. Non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (1988) *CABIOS* 4:11-17; the local alignment algorithm of Smith *et al.* (1981) *Adv. Appl. Math.* 2:482; the global alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443-453; the search-for-local alignment method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci.* 85:2444-2448; the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877.

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations
30 include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the GCG Wisconsin Genetics Software Package, Version 10 (available from Accelrys Inc., 9685 Scranton Road, San Diego, California, USA). Alignments using these programs can be performed

using the default parameters. The CLUSTAL program is well described by Higgins *et al.* (1988) *Gene* 73:237-244 (1988); Higgins *et al.* (1989) *CABIOS* 5:151-153; Corpet *et al.* (1988) *Nucleic Acids Res.* 16:10881-90; Huang *et al.* (1992) *CABIOS* 8:155-65; and Pearson *et al.* (1994) *Meth. Mol. Biol.* 24:307-331. The ALIGN program is based
5 on the algorithm of Myers and Miller (1988) *supra*. A PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used with the ALIGN program when comparing amino acid sequences. The BLAST programs of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403 are based on the algorithm of Karlin and Altschul (1990) *supra*. BLAST nucleotide searches can be performed with the BLASTN
10 program, score = 100, wordlength = 12, to obtain nucleotide sequences homologous to a nucleotide sequence encoding a protein of the invention. BLAST protein searches can be performed with the BLASTX program, score = 50, wordlength = 3, to obtain amino acid sequences homologous to a protein or polypeptide of the invention. BLASTP protein searches can be performed using default parameters. See,
15 blast.ncbi.nlm.nih.gov/Blast.cgi.

To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul *et al.* (1997) *supra*. When utilizing BLAST, Gapped BLAST, or PSI-BLAST, the
20 default parameters of the respective programs (e.g., BLASTN for nucleotide sequences, BLASTP for proteins) can be used. See www.ncbi.nlm.nih.gov. Alignment may also be performed manually by inspection.

In one embodiment, sequence identity/similarity values provided herein refer
25 to the value obtained using GAP Version 10 using the following parameters: % identity and % similarity for an amino acid sequence using GAP Weight of 8 and Length Weight of 2, and the BLOSUM62 scoring matrix; or any equivalent program thereof. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide
30 or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by GAP Version 10.

GAP uses the algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443-453, to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible

alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps. It allows for the provision of a gap creation penalty and a gap extension penalty in units of matched bases. GAP must make a profit of gap creation penalty number of matches for each gap it inserts. If a gap extension penalty greater than zero is chosen, GAP must, in addition, make a profit for each gap inserted of the length of the gap times the gap extension penalty. Default gap creation penalty values and gap extension penalty values in Version 10 of the GCG Wisconsin Genetics Software Package for protein sequences are 8 and 2, respectively. For nucleotide sequences the default gap creation penalty is 50 while the default gap extension penalty is 3. The gap creation and gap extension penalties can be expressed as an integer selected from the group of integers consisting of from 0 to 200. Thus, for example, the gap creation and gap extension penalties can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 or greater.

GAP presents one member of the family of best alignments. There may be many members of this family, but no other member has a better quality. GAP displays four figures of merit for alignments: Quality, Ratio, Identity, and Similarity. The Quality is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment. Percent Identity is the percent of the symbols that actually match. Percent Similarity is the percent of the symbols that are similar. Symbols that are across from gaps are ignored. A similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the similarity threshold. The scoring matrix used in Version 10 of the GCG Wisconsin Genetics Software Package is BLOSUM62 (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

(c) As used herein, "sequence identity" or "identity" in the context of two polynucleotides or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity). When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to

have "sequence similarity" or "similarity". Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percent sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

(d) As used herein, "percent sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percent sequence identity.

(e) Two sequences are "optimally aligned" when they are aligned for similarity scoring using a defined amino acid substitution matrix (e.g., BLOSUM62), gap existence penalty and gap extension penalty so as to arrive at the highest score possible for that pair of sequences. Amino acids substitution matrices and their use in quantifying the similarity between two sequences are well-known in the art and described, e.g., in Dayhoff et al. (1978) "A model of evolutionary change in proteins." In "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3 (ed. M.O. Dayhoff), pp. 345-352. Natl. Biomed. Res. Found., Washington, DC and Henikoff et al. (1992) Proc. Natl. Acad. Sci. USA 89:10915-10919. The BLOSUM62 matrix is often used as a default scoring substitution matrix in sequence alignment protocols such as Gapped BLAST 2.0. The gap existence penalty is imposed for the introduction of a single amino acid gap in one of the aligned sequences, and the gap extension penalty is imposed for each additional empty amino acid position inserted into an already opened gap. The gap existence penalty is imposed for the introduction of a single amino acid gap in one of the aligned sequences, and the gap extension penalty is imposed for each additional empty amino acid position inserted into an already

opened gap. The alignment is defined by the amino acids positions of each sequence at which the alignment begins and ends, and optionally by the insertion of a gap or multiple gaps in one or both sequences, so as to arrive at the highest possible score. While optimal alignment and scoring can be accomplished manually, the process is
5 facilitated by the use of a computer-implemented alignment algorithm, e.g., gapped BLAST 2.0, described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402, and made available to the public at the National Center for Biotechnology Information Website (<http://www.ncbi.nlm.nih.gov>). Optimal alignments, including multiple alignments, can be prepared using, e.g., PSI-BLAST, available through
10 <http://www.ncbi.nlm.nih.gov> and described by Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402.

As used herein, similarity score and bit score is determined employing the BLAST alignment used the BLOSUM62 substitution matrix, a gap existence penalty of 11, and a gap extension penalty of 1. For the same pair of sequences, if there is a
15 numerical difference between the scores obtained when using one or the other sequence as query sequences, a greater value of similarity score is selected.

C. Polynucleotides/Polypeptides of Interest

Any heterologous polynucleotide of interest (i.e., the "polypeptide of interest")
20 may be used with the CTP-encoding sequences disclosed herein (i.e. the various chimeric CTPs disclosed herein and/or SEQ ID NOS: 1, 2, 3, 4, 5, 6, 7 or active variants or fragments thereof). It is recognized that any polypeptides of interest can be operably linked to the CTP-encoding sequences provided herein and expressed in a plant.

25 Such polynucleotides/polypeptides of interest include, but are not limited to, herbicide-tolerance coding sequences, insecticidal coding sequences, nematocidal coding sequences, antimicrobial coding sequences, antifungal coding sequences, antiviral coding sequences, abiotic and biotic stress tolerance coding sequences, or sequences modifying plant traits such as yield, grain quality, nutrient content, starch
30 quality and quantity, nitrogen fixation and/or utilization, and oil content and/or composition. More specific polynucleotides of interest include, but are not limited to, genes that improve crop yield, polypeptides that improve desirability of crops, genes encoding proteins conferring resistance to abiotic stress, such as drought, nitrogen, temperature, salinity, toxic metals or trace elements, or those conferring resistance to

toxins such as pesticides and herbicides, or to biotic stress, such as attacks by fungi, viruses, bacteria, insects, and nematodes, and development of diseases associated with these organisms.

An "herbicide resistance protein" or a protein resulting from expression of an
5 "herbicide resistance-encoding nucleic acid molecule" includes proteins that confer upon a cell the ability to tolerate a higher concentration of an herbicide than cells that do not express the protein, or to tolerate a certain concentration of an herbicide for a longer period of time than cells that do not express the protein. Herbicide resistance traits may be introduced into plants by genes coding for resistance to herbicides that
10 act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides, genes coding for resistance to herbicides that act to inhibit the action of glutamine synthase, such as phosphinothricin or basta (e.g., the bar gene), glyphosate (e.g., the EPSP synthase gene and the GAT gene), HPPD inhibitors (e.g., the HPPD gene) or other such genes known in the art. See, for example, US Patent
15 Nos. 7,626,077, 5,310,667, 5,866,775, 6,225,114, 6,248,876, 7,169,970, 6,867,293, and US Provisional Application No. 61/401,456, each of which is herein incorporated by reference.

Polynucleotides that improve crop yield include dwarfing genes, such as Rht1 and Rht2 (Peng *et al.* (1999) *Nature* 400:256-261), and those that increase plant
20 growth, such as ammonium-inducible glutamate dehydrogenase. Polynucleotides that improve desirability of crops include, for example, those that allow plants to have a reduced saturated fat content, those that boost the nutritional value of plants, and those that increase grain protein. Polynucleotides that improve salt tolerance are those that increase or allow plant growth in an environment of higher salinity than the
25 native environment of the plant into which the salt-tolerant gene(s) has been introduced.

Polynucleotides/polypeptides that influence amino acid biosynthesis include, for example, anthranilate synthase (AS; EC 4.1.3.27) which catalyzes the first reaction branching from the aromatic amino acid pathway to the biosynthesis of
30 tryptophan in plants, fungi, and bacteria. In plants, the chemical processes for the biosynthesis of tryptophan are compartmentalized in the chloroplast. See, for example, US Pub. 20080050506, herein incorporated by reference. Additional sequences of interest include Chorismate Pyruvate Lyase (CPL) which refers to a gene encoding an enzyme which catalyzes the conversion of chorismate to pyruvate

and pHBA. The most well characterized CPL gene has been isolated from *E. coli* and bears the GenBank accession number M96268. See, US Patent No. 7,361,811, herein incorporated by reference.

These polynucleotide sequences of interest may encode proteins involved in providing disease or pest resistance. By "disease resistance" or "pest resistance" is intended that the plants avoid the harmful symptoms that are the outcome of the plant-pathogen interactions. Disease resistance and insect resistance genes such as lysozymes or cecropins for antibacterial protection, or proteins such as defensins, glucanases or chitinases for antifungal protection, or *Bacillus thuringiensis* endotoxins, protease inhibitors, collagenases, lectins, or glycosidases for controlling nematodes or insects are all examples of useful gene products.

In some embodiments, a CTP provided herein is operably linked to a heterologous polypeptide of interest comprising an insecticidal protein and expression of the polypeptide controls a pest (i.e. insecticidal activity). As used herein, by "controlling a pest" or "controls a pest" is intended any affect on a pest that results in limiting the damage that the pest causes. Controlling a pest includes, but is not limited to, killing the pest, inhibiting development of the pest, altering fertility or growth of the pest in such a manner that the pest provides less damage to the plant, decreasing the number of offspring produced, producing less fit pests, producing pests more susceptible to predator attack, or deterring the pests from eating the plant.

"Pest" includes, but is not limited to, insects, fungi, bacteria, viruses, nematodes, mites, ticks, and the like. Insect pests include insects selected from the orders Coleoptera, Diptera, Hymenoptera, Lepidoptera, Mallophaga, Homoptera, Hemiptera, Orthoptera, Thysanoptera, Dermaptera, Isoptera, Anoplura, Siphonaptera, Trichoptera, etc., particularly Coleoptera, Lepidoptera, and Diptera. Viruses include but are not limited to tobacco or cucumber mosaic virus, ringspot virus, necrosis virus, maize dwarf mosaic virus, etc. Nematodes include but are not limited to parasitic nematodes such as root knot, cyst, and lesion nematodes, including *Heterodera spp.*, *Meloidogyne spp.*, and *Globodera spp.*; particularly members of the cyst nematodes, including, but not limited to, *Heterodera glycines* (soybean cyst nematode); *Heterodera schachtii* (beet cyst nematode); *Heterodera avenae* (cereal cyst nematode); and *Globodera rostochiensis* and *Globodera pallida* (potato cyst nematodes). Lesion nematodes include but are not limited to *Pratylenchus spp.* Fungal pests include those that cause leaf, yellow, stripe and stem rusts.

In other embodiments, a polypeptide of interest comprises a *Bacillus thuringiensis* polypeptide having insecticidal activity (i.e. controls a pest). Some examples of *Bacillus thuringiensis* toxic proteins include the Cry proteins. Other *Bacillus thuringiensis* toxic proteins are described in U.S. Patent Nos. 5,366,892; 5,747,450; 5,737,514; 5,723,756; 5,593,881; and Geiser *et al.* (1986) *Gene* 48:109, herein incorporated by reference. In a specific embodiment, the *Bacillus thuringiensis* polypeptide is IP2-127 (SEQ ID NO: 12) or an active variant or fragment thereof. IP2-127 is a Cry2 protein of *Bacillus thuringiensis* with insecticidal activity. The IP2-127 protein may be modified to comprise, for example, a short linker sequence or a reporter gene in order to allow detection of the protein. An Example of a modified IP2-127 protein sequence is set forth in SEQ ID NO: 8 or an active variant or fragment thereof and is encoded by the polynucleotide sequence set forth in SEQ ID NO:9 or an active variant or fragment thereof and comprises an IP2-127- AcGFP fusion protein.

It is recognized that any polypeptide of interest may be modified to comprise, for example, a short linker sequence or a reporter gene in order to allow detection of the protein in the chloroplast.

Active variants or fragments of polynucleotides/polypeptides of interest (i.e. SEQ ID NO:12) are also provided. Such active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the native polynucleotide/polypeptide of interest, wherein the active variants retain biological activity and are functional in chloroplasts. Active fragments can comprise nucleic acid/amino acid sequences having at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 100, 150, or more consecutive nucleic acids/amino acids of the native polynucleotide/polypeptide of interest, where the active fragments retain biological activity and are functional in chloroplasts. As used herein, a "native" polynucleotide or polypeptide comprises a naturally occurring nucleotide sequence or amino acid sequence, respectively. Methods to determine sequence identity/sequence similarity are described in detail elsewhere herein.

D. Plants

Compositions comprising a cell, a transgenic plant cell, a transgenic plant, transgenic plant parts and seeds, plant explants and grain having the recombinant polynucleotide encoding a CTP operably linked to a heterologous polynucleotide

encoding a polypeptide of interest are further provided. In one embodiment, a cell, a plant cell, a plant, plant parts and seeds, plant explants and grain comprise at least one polynucleotide encoding a CTP provided herein (i.e. The various chimeric CTPs disclosed herein and/or SEQ ID NOS: 1, 2, 3, 4, 5, 6, 7 or active variants or fragments thereof) operably linked to a polypeptide of interest. The CTP may comprise a chimeric CTP, a chimeric CTP comprising chimeric domains, or a CTP comprising a consensus sequence as described in detail elsewhere herein. In some cases, the polynucleotide encoding the polypeptide of interest can comprise an insecticidal protein that controls a pest, a *Bacillus thuringiensis* protein having insecticidal activity, or an IP2-127 protein (i.e. SEQ ID NO:12) or active variant or fragment thereof.

As used herein, the term plant includes whole plants, plant organs, plant tissues, seeds and plant cells and progeny of the same, plant protoplasts, plant cell tissue cultures from which plants can be regenerated, plant calli, plant clumps, and plant cells that are intact in plants or parts of plants such as embryos, pollen, ovules, seeds, leaves, flowers, branches, fruit, kernels, ears, cobs, husks, stalks, roots, root tips, anthers, and the like. Grain is intended to mean the mature seed produced by commercial growers for purposes other than growing or reproducing the species. Progeny, variants, and mutants of the regenerated plants are also included, provided that these parts comprise the introduced recombinant polynucleotides.

A transformed plant or transformed plant cell provided herein is one in which genetic alteration, such as transformation, has been affected as to a gene of interest, or is a plant or plant cell which is descended from a plant or cell so altered and which comprises the alteration. A “transgene” is a gene that has been introduced into the genome by a transformation procedure. Accordingly, a “transgenic plant” is a plant that contains a transgene, whether the transgene was introduced into that particular plant by transformation or by breeding; thus, descendants of an originally-transformed plant are encompassed by the definition. A “subject plant or plant cell” is one in which genetic alteration, such as transformation, has been affected as to a gene of interest, or is a plant or plant cell which is descended from a plant or cell so altered and which comprises the alteration. A “control” or “control plant” or “control plant cell” provides a reference point for measuring changes in phenotype of the subject plant or plant cell. A control plant or plant cell may comprise, for example: (a) a wild-type plant or cell, i.e., of the same genotype as the starting material for the

genetic alteration which resulted in the subject plant or cell; (b) a plant or plant cell of the same genotype as the starting material but which has been transformed with a null construct (i.e., with a construct which does not express the CTP operably linked to a polypeptide of interest, such as a construct comprising a marker gene); (c) a plant or
5 plant cell which is a non-transformed segregant among progeny of a subject plant or plant cell; (d) a plant or plant cell genetically identical to the subject plant or plant cell but which is not exposed to conditions or stimuli that would induce expression of the recombinant polynucleotide; or (e) the subject plant or plant cell itself, under conditions in which the recombinant polynucleotide is not expressed.

10 Plant cells that have been transformed to have a recombinant polynucleotide encoding a CTP operably linked to a polypeptide of interest provided herein can be grown into whole plants. The regeneration, development, and cultivation of plants from single plant protoplast transformants or from various transformed explants is well known in the art. See, for example, McCormick *et al.* (1986) *Plant Cell Reports*
15 5:81-84; Weissbach and Weissbach, In: *Methods for Plant Molecular Biology*, (Eds.), Academic Press, Inc. San Diego, Calif., (1988). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those individualized cells through the usual stages of embryonic development through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The
20 resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil. Preferably, the regenerated plants are self-pollinated to provide homozygous transgenic plants. Otherwise, pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important lines. Conversely, pollen from plants of these important lines is used to pollinate regenerated plants.
25 Two or more generations may be grown to ensure that expression of the desired phenotypic characteristic is stably maintained and inherited and then seeds harvested to ensure expression of the desired phenotypic characteristic has been achieved. In this manner, the compositions presented herein provide transformed seed (also referred to as “transgenic seed”) having a polynucleotide provided herein, for
30 example, a recombinant polynucleotide encoding a CTP operably linked to a polypeptide of interest, stably incorporated into their genome.

The recombinant polynucleotides disclosed herein may be used for transformation of any plant species, including, but not limited to, monocots (e.g., maize, sugarcane, wheat, rice, barley, sorghum, or rye) and dicots (e.g., soybean, *Brassica*,

sunflower, cotton, or alfalfa). Examples of plant species of interest include, but are not limited to, corn (*Zea mays*), *Brassica* sp. (e.g., *B. napus*, *B. rapa*, *B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*),
5 millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato
10 (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Coffea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia
15 (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, barley, vegetables, ornamentals, and conifers.

Vegetables include, but not limited to, tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber
20 (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include, but not limited to, azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum.

25 Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*); Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir
30 (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*), and Poplar and Eucalyptus. In specific embodiments, plants of the present invention are crop plants (for example, corn, alfalfa, sunflower, *Brassica*, soybean, cotton, safflower, peanut, sorghum,

wheat, millet, tobacco, etc.). In other embodiments, corn and soybean plants are optimal, and in yet other embodiments corn plants are optimal.

Other plants of interest include grain plants that provide seeds of interest, oil-seed plants, and leguminous plants. Seeds of interest include grain seeds, such as
5 corn, wheat, barley, rice, sorghum, rye, etc. Oil-seed plants include cotton, soybean, safflower, sunflower, *Brassica*, maize, alfalfa, palm, coconut, etc. Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, etc.

In some embodiments, the recombinant polynucleotides comprising the CTP-
10 encoding sequence operably linked to the polynucleotide encoding the polypeptide of interest are engineered into a molecular stack. Thus, the various plants, plant cells and seeds disclosed herein can further comprise one or more traits of interest, and in more specific embodiments, the plant, plant part or plant cell is stacked with any combination of polynucleotide sequences of interest in order to create plants with a
15 desired combination of traits. As used herein, the term “stacked” includes having the multiple traits present in the same plant.

These stacked combinations can be created by any method including, but not limited to, breeding plants by any conventional methodology, or genetic transformation. If the sequences are stacked by genetically transforming the plants,
20 the polynucleotide sequences of interest can be combined at any time and in any order. The traits can be introduced simultaneously in a co-transformation protocol with the polynucleotides of interest provided by any combination of transformation cassettes. For example, if two sequences will be introduced, the two sequences can be contained in separate transformation cassettes (trans) or contained on the same
25 transformation cassette (cis). Expression of the sequences can be driven by the same promoter or by different promoters. In certain cases, it may be desirable to introduce a transformation cassette that will suppress the expression of the polynucleotide of interest. This may be combined with any combination of other suppression cassettes or overexpression cassettes to generate the desired combination of traits in the plant.
30 It is further recognized that polynucleotide sequences can be stacked at a desired genomic location using a site-specific recombination system. See, for example, WO99/25821, WO99/25854, WO99/25840, WO99/25855, and WO99/25853, all of which are herein incorporated by reference.

Depending on the polypeptide of interest, the transgenic plants, plant cells or seeds expressing a recombinant polynucleotide provided herein may have a change in phenotype, including but not limited to, an altered pathogen or insect defense mechanism, an increased resistance to one or more herbicides, an increased ability to
5 withstand stressful environmental conditions, a modified ability to produce starch, a modified level of starch production, a modified oil content and/or composition, a modified carbohydrate content and/or composition, a modified ability to utilize, partition and/or store nitrogen, and the like.

10 *E. Polynucleotide Constructs*

Also provided are isolated or recombinant polynucleotides and nucleic acid constructs that encode the CTPs disclosed herein (i.e. the various chimeric CTPs disclosed herein and/or SEQ ID NOS: 1, 2, 3, 4, 5, 6, 7 or active variants or fragments thereof) operably linked to a polynucleotide encoding a polypeptide of interest. As
15 used herein, “encodes” or “encoding” refers to a DNA sequence which can be processed to generate an RNA and/or polypeptide.

The terms “polynucleotide,” “polynucleotide sequence,” “nucleic acid sequence,” and “nucleic acid fragment” are used interchangeably herein. These terms encompass nucleotide sequences and the like. A polynucleotide may be a polymer of
20 RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof. The use of the term “polynucleotide” is not intended to limit the present invention to polynucleotides comprising DNA. Those of ordinary
25 skill in the art will recognize that polynucleotides, can comprise ribonucleotides and combinations of ribonucleotides and deoxyribonucleotides. Such deoxyribonucleotides and ribonucleotides include both naturally occurring molecules and synthetic analogues. The polynucleotides provided herein also encompass all forms of sequences including, but not limited to, single-stranded forms, double-
30 stranded forms, hairpins, stem-and-loop structures, and the like.

The compositions provided herein can comprise an isolated or substantially purified polynucleotide. An “isolated” or “purified” polynucleotide is substantially or essentially free from components that normally accompany or interact with the polynucleotide as found in its naturally occurring environment. Thus, an isolated or

purified polynucleotide is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. Optimally, an “isolated” polynucleotide is free of sequences (optimally protein encoding sequences) that naturally flank the polynucleotide (i.e., sequences located at the 5' and 3' ends of the polynucleotide) in the genomic DNA of the organism from which the polynucleotide is derived. For example, in various embodiments, the isolated polynucleotide can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequence that naturally flank the polynucleotide in genomic DNA of the cell from which the polynucleotide is derived.

Further provided are recombinant polynucleotides comprising the CTP sequences and polynucleotide sequences encoding the polypeptides of interest. The terms “recombinant polynucleotide” and “recombinant DNA construct” are used interchangeably herein. A recombinant construct comprises an artificial or heterologous combination of nucleic acid sequences, e.g., regulatory and coding sequences that are not found together in nature. For example, a recombinant polynucleotide can comprise a chimeric CTP operably linked to a heterologous polynucleotide encoding a polypeptide of interest. In other embodiments, a recombinant construct may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. Such a construct may be used by itself or may be used in conjunction with a vector. If a vector is used, then the choice of vector is dependent upon the method that will be used to transform host cells as is well known to those skilled in the art. For example, a plasmid vector can be used. The skilled artisan is well aware of the genetic elements that must be present on the vector in order to successfully transform, select and propagate host cells comprising any of the isolated nucleic acid fragments of the invention. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones *et al.*, *EMBO J.* 4:2411-2418 (1985); De Almeida *et al.*, *Mol. Gen. Genetics* 218:78-86 (1989)), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, immunoblotting analysis of protein expression, or phenotypic analysis, among others.

The recombinant polynucleotides disclosed herein can be provided in expression cassettes for expression in a plant or other organism or cell type of interest. The cassette can include 5' and 3' regulatory sequences operably linked to the recombinant polynucleotide or active variant or fragment thereof. "Operably linked" is intended to mean a functional linkage between two or more elements. For example, an operable linkage between a polynucleotide of interest and a regulatory sequence (i.e., a promoter) is a functional link that allows for expression of the polynucleotide of interest. Operably linked elements may be contiguous or non-contiguous. When used to refer to the joining of two protein coding regions, by operably linked is intended that the coding regions are in the same reading frame. The cassette may additionally contain at least one additional gene to be cotransformed into the organism. Alternatively, the additional gene(s) can be provided on multiple expression cassettes. Such an expression cassette is provided with a plurality of restriction sites and/or recombination sites for insertion of the recombinant polynucleotide or active variant or fragment thereof to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selectable marker genes.

The expression cassette can include in the 5'-3' direction of transcription, a transcriptional and translational initiation region (i.e., a promoter), a CTP-encoding sequence or active variant or fragment thereof operably linked to a polynucleotide encoding a polypeptide of interest and a transcriptional and translational termination region (i.e., termination region) functional in plants. The regulatory regions (i.e., promoters, transcriptional regulatory regions, and translational termination regions) and/or the CTP-encoding sequence and/or the polynucleotide encoding the polypeptide of interest may be native/analogous to the host cell or to each other. Alternatively, the regulatory regions and/or the CTP-encoding sequence and/or the polynucleotide encoding the polypeptide of interest may be heterologous to the host cell or to each other. In specific embodiments, the CTP-encoding sequence is operably linked to the 5' end of the polynucleotide of interest, such that, in the resulting recombinant polypeptide, the CTP is operably linked to the N-terminal region of the polypeptide of interest.

As used herein, "heterologous" in reference to a sequence is a sequence that originates from a foreign species, for example, from a different CTP, or, if from the same species, is substantially modified from its native form in composition and/or

genomic locus by deliberate human intervention. For example, a heterologous domain is intended at least one of the CTP domains is not from the same CTP, but could be from a different CTP of the same plant species or a different plant species. In another example, a promoter operably linked to a heterologous polynucleotide is
5 from a species different from the species from which the polynucleotide was derived, or, if from the same/analogous species, one or both are substantially modified from their original form and/or genomic locus, or the promoter is not the native promoter for the operably linked polynucleotide.

The termination region may be native with the transcriptional initiation region,
10 may be native with the operably linked polynucleotide sequence of interest, may be native with the plant host, or may be derived from another source (i.e., foreign or heterologous) to the promoter, the CTP, the polynucleotide sequence of interest, the plant host, or any combination thereof. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline
15 synthase termination regions. See also Guerineau *et al.* (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot (1991) *Cell* 64:671-674; Sanfacon *et al.* (1991) *Genes Dev.* 5:141-149; Mogen *et al.* (1990) *Plant Cell* 2:1261-1272; Munroe *et al.* (1990) *Gene* 91:151-158; Ballas *et al.* (1989) *Nucleic Acids Res.* 17:7891-7903; and Joshi *et al.* (1987) *Nucleic Acids Res.* 15:9627-9639.

20 In preparing the expression cassette, the various DNA fragments may be manipulated, so as to provide for the DNA sequences in the proper orientation. Toward this end, adapters or linkers may be employed to join the DNA fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this
25 purpose, *in vitro* mutagenesis, primer repair, restriction, annealing, resubstitutions, e.g., transitions and transversions, may be involved.

Where appropriate, the polynucleotides may be optimized for increased expression in the transformed plant. That is, the polynucleotides can be synthesized using plant-preferred codons for improved expression. See, for example, Campbell
30 and Gowri (1990) *Plant Physiol.* 92:1-11 for a discussion of host-preferred codon usage. Methods are available in the art for synthesizing plant-preferred genes. See, for example, U.S. Patent Nos. 5,380,831, and 5,436,391, and Murray *et al.* (1989) *Nucleic Acids Res.* 17:477-498, herein incorporated by reference.

Additional sequence modifications are known to enhance gene expression in a cellular host. These include elimination of sequences encoding spurious polyadenylation signals, exon-intron splice site signals, transposon-like repeats, and other such well-characterized sequences that may be deleterious to gene expression.

5 The G-C content of the sequence may be adjusted to levels average for a given cellular host, as calculated by reference to known genes expressed in the host cell. When possible, the sequence is modified to avoid predicted hairpin secondary mRNA structures.

The expression cassettes may additionally contain 5' leader sequences. Such
10 leader sequences can act to enhance translation. Translation leaders are known in the art and include: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region) (Elroy-Stein *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6126-6130); potyvirus leaders, for example, TEV leader (Tobacco Etch Virus) (Gallie *et al.* (1995) *Gene* 165(2):233-238), MDMV leader (Maize Dwarf
15 Mosaic Virus) (*Virology* 154:9-20), and human immunoglobulin heavy-chain binding protein (BiP) (Macejak *et al.* (1991) *Nature* 353:90-94); untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4) (Jobling *et al.* (1987) *Nature* 325:622-625); tobacco mosaic virus leader (TMV) (Gallie *et al.* (1989) in *Molecular Biology of RNA*, ed. Cech (Liss, New York), pp. 237-256); and maize
20 chlorotic mottle virus leader (MCMV) (Lommel *et al.* (1991) *Virology* 81:382-385. See also, Della-Cioppa *et al.* (1987) *Plant Physiol.* 84:965-968.

A number of promoters can be used to express the recombinant polynucleotides provided herein. The promoters can be selected based on the desired outcome. It is recognized that different applications can be enhanced by the use of
25 different promoters in the expression constructs to modulate the timing, location and/or level of expression of the recombinant polynucleotide. Such expression constructs may also contain, if desired, a promoter regulatory region (e.g. one conferring inducible, constitutive, environmentally- or developmentally regulated, or cell- or tissue-specific/selective expression), a transcription initiation start site, a
30 ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal. including the native promoter of the polynucleotide sequence of interest.

In some embodiments, an expression construct provided herein can be combined with constitutive, tissue-preferred, or other promoters for expression in

plants. Examples of constitutive promoters include, for example, the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'-promoter derived from T-DNA of *Agrobacterium tumefaciens*, the ubiquitin 1 promoter, the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (U.S. Pat. No. 5,683,439), the Nos promoter, the pEmu promoter, the rubisco promoter, the GRP1-8 promoter and other transcription initiation regions from various plant genes known to those of skill. If low level expression is desired, weak promoter(s) may be used. Weak constitutive promoters include, for example, the core promoter of the Rsyn7 promoter (WO 99/43838 and U.S. Pat. No. 6,072,050), the core 35S CaMV promoter, and the like. Other constitutive promoters include, for example, U.S. Pat. Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; and 5,608,142. See also, U.S. Pat. No. 6,177,611, herein incorporated by reference.

Examples of inducible promoters are the Adh1 promoter which is inducible by hypoxia or cold stress, the Hsp70 promoter which is inducible by heat stress, the PPDK promoter and the pepcarboxylase promoter which are both inducible by light. Also useful are promoters which are chemically inducible, such as the In2-2 promoter which is safener induced (U.S. Pat. No. 5,364,780), the ERE promoter which is estrogen induced, and the Axig1 promoter which is auxin induced and tapetum specific but also active in callus (PCT US01/22169).

Examples of promoters under developmental control include promoters that initiate transcription preferentially in certain tissues, such as leaves, roots, fruit, seeds, or flowers. An exemplary promoter is the anther specific promoter 5126 (U.S. Pat. Nos. 5,689,049 and 5,689,051). Examples of seed-preferred promoters include, but are not limited to, 27 kD gamma zein promoter and waxy promoter, Boronat, A. *et al.* (1986) *Plant Sci.* 47:95-102; Reina, M. *et al.* *Nucl. Acids Res.* 18(21):6426; and Kloesgen, R. B. *et al.* (1986) *Mol. Gen. Genet.* 203:237-244. Promoters that express in the embryo, pericarp, and endosperm are disclosed in U.S. Pat. No. 6,225,529 and PCT publication WO 00/12733. The disclosures for each of these are incorporated herein by reference in their entirety.

Chemical-regulated promoters can be used to modulate the expression of a gene in a plant through the application of an exogenous chemical regulator. Depending upon the objective, the promoter may be a chemical-inducible promoter, where application of the chemical induces gene expression, or a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-

inducible promoters are known in the art and include, but are not limited to, the maize In2-2 promoter, which is activated by benzenesulfonamide herbicide safeners, the maize GST promoter, which is activated by hydrophobic electrophilic compounds that are used as pre-emergent herbicides, and the tobacco PR-1a promoter, which is

5 activated by salicylic acid. Other chemical-regulated promoters of interest include steroid-responsive promoters (see, for example, the glucocorticoid-inducible promoter in Schena *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-10425 and McNellis *et al.* (1998) *Plant J.* 14(2):247-257) and tetracycline-inducible and tetracycline-repressible promoters (see, for example, Gatz *et al.* (1991) *Mol. Gen. Genet.* 227:229-

10 237, and U.S. Pat. Nos. 5,814,618 and 5,789,156), herein incorporated by reference.

Tissue-preferred promoters can be utilized to target enhanced expression or a recombinant polynucleotide within a particular plant tissue. Tissue-preferred promoters are known in the art. See, for example, Yamamoto *et al.* (1997) *Plant J.* 12(2):255-265; Kawamata *et al.* (1997) *Plant Cell Physiol.* 38(7):792-803; Hansen *et al.* (1997) *Mol. Gen. Genet.* 254(3):337-343; Russell *et al.* (1997) *Transgenic Res.* 6(2):157-168; Rinehart *et al.* (1996) *Plant Physiol.* 112(3):1331-1341; Van Camp *et al.* (1996) *Plant Physiol.* 112(2):525-535; Canevascini *et al.* (1996) *Plant Physiol.* 112(2):513-524; Yamamoto *et al.* (1994) *Plant Cell Physiol.* 35(5):773-778; Lam (1994) *Results Probl. Cell Differ.* 20:181-196; Orozco *et al.* (1993) *Plant Mol Biol.*

15 23(6):1129-1138; Matsuoka *et al.* (1993) *Proc Natl. Acad. Sci. USA* 90(20):9586-9590; and Guevara-Garcia *et al.* (1993) *Plant J.* 4(3):495-505. Such promoters can be modified, if necessary, for weak expression.

Leaf-preferred promoters are known in the art. See, for example, Yamamoto *et al.* (1997) *Plant J.* 12(2):255-265; Kwon *et al.* (1994) *Plant Physiol.* 105:357-67; Yamamoto *et al.* (1994) *Plant Cell Physiol.* 35(5):773-778; Gotor *et al.* (1993) *Plant J.* 3:509-18; Orozco *et al.* (1993) *Plant Mol. Biol.* 23(6):1129-1138; and Matsuoka *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90(20):9586-9590. In addition, the promoters of cab and rubisco can also be used. See, for example, Simpson *et al.* (1958) *EMBO J* 4:2723-2729 and Timko *et al.* (1988) *Nature* 318:57-58.

25

The expression cassette can also comprise a selectable marker gene for the selection of transformed cells. Selectable marker genes are utilized for the selection of transformed cells or tissues. Marker genes include genes encoding antibiotic resistance, such as those encoding neomycin phosphotransferase II (NEO) and hygromycin phosphotransferase (HPT), as well as genes conferring resistance to herbicidal

30

compounds, such as glyphosate, glufosinate ammonium, bromoxynil, sulfonyleureas, dicamba, and 2,4-dichlorophenoxyacetate (2,4-D). Additional selectable markers include phenotypic markers such as β -galactosidase and fluorescent proteins such as green fluorescent protein (GFP) (Su *et al.* (2004) *Biotechnol Bioeng* 85:610-9 and Fetter *et al.* (2004) *Plant Cell* 16:215-28), cyan florescent protein (CYP) (Bolte *et al.* (2004) *J. Cell Science* 117:943-54 and Kato *et al.* (2002) *Plant Physiol* 129:913-42), and yellow florescent protein (PhiYFP™ from Evrogen, see, Bolte *et al.* (2004) *J. Cell Science* 117:943-54). For additional selectable markers, see generally, Yarranton (1992) *Curr. Opin. Biotech.* 3:506-511; Christopherson *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6314-6318; Yao *et al.* (1992) *Cell* 71:63-72; Reznikoff (1992) *Mol. Microbiol.* 6:2419-2422; Barkley *et al.* (1980) in *The Operon*, pp. 177-220; Hu *et al.* (1987) *Cell* 48:555-566; Brown *et al.* (1987) *Cell* 49:603-612; Figge *et al.* (1988) *Cell* 52:713-722; Deuschle *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:5400-5404; Fuerst *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:2549-2553; Deuschle *et al.* (1990) *Science* 248:480-483; Gossen (1993) Ph.D. Thesis, University of Heidelberg; Reines *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:1917-1921; Labow *et al.* (1990) *Mol. Cell. Biol.* 10:3343-3356; Zambretti *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3952-3956; Baim *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:5072-5076; Wyborski *et al.* (1991) *Nucleic Acids Res.* 19:4647-4653; Hillenand-Wissman (1989) *Topics Mol. Struc. Biol.* 10:143-162; Degenkolb *et al.* (1991) *Antimicrob. Agents Chemother.* 35:1591-1595; Kleinschmidt *et al.* (1988) *Biochemistry* 27:1094-1104; Bonin (1993) Ph.D. Thesis, University of Heidelberg; Gossen *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Oliva *et al.* (1992) *Antimicrob. Agents Chemother.* 36:913-919; Hlavka *et al.* (1985) *Handbook of Experimental Pharmacology*, Vol. 78 (Springer-Verlag, Berlin); Gill *et al.* (1988) *Nature* 334:721-724. Such disclosures are herein incorporated by reference. The above list of selectable marker genes is not meant to be limiting. Any selectable marker gene can be employed herein, including for example, Ac-GFP as described in Examples 2 and 3.

30 II. Methods of Introducing

The methods provided herein comprise introducing into a cell, plant cell, plant or seed a recombinant polynucleotide or nucleic acid construct encoding a CTP provided herein (i.e. Any of the chimeric CTPs provided herein and/or SEQ ID NOS:

1, 2, 3, 4, 5, 6, 7 or active variants or fragments thereof) operably linked to a heterologous polynucleotide encoding a polypeptide of interest.

In some embodiments, the CTP introduced in the recombinant polynucleotide can be a chimeric CTP, a chimeric CTP comprising at least one chimeric domain, or a CTP comprising a consensus sequence as described in detail elsewhere herein. The CTP may be linked to any polypeptide of interest. For example, the polypeptide of interest can comprise an insecticidal protein whose expression controls a pest, a *Bacillus thuringiensis* polypeptide having insecticidal activity, or an IP2-127 polypeptide (i.e. SEQ ID NO:12 or an active variant or fragment thereof).

The methods provided herein do not depend on a particular method for introducing a sequence into the host cell, only that the polynucleotide gains access to the interior of a least one cell of the host. Methods for introducing polynucleotides into host cells (i.e. plants) are known in the art and include, but are not limited to, stable transformation methods, transient transformation methods, and virus-mediated methods.

The terms "introducing" and "introduced" are intended to mean providing a nucleic acid (e.g., recombinant polynucleotide) or protein into a cell. Introduced includes reference to the incorporation of a nucleic acid into a eukaryotic or prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell, and includes reference to the transient provision of a nucleic acid or protein to the cell. Introduced includes reference to stable or transient transformation methods, as well as sexually crossing. Thus, "introduced" in the context of inserting a nucleic acid fragment (e.g., a recombinant polynucleotide) into a cell, means "transfection" or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid fragment into a eukaryotic or prokaryotic cell where the nucleic acid fragment may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid, or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (e.g., transfected mRNA).

"Stable transformation" is intended to mean that the nucleotide construct introduced into a host (i.e., a plant) integrates into the genome of the plant and is capable of being inherited by the progeny thereof. "Transient transformation" is intended to mean that a polynucleotide is introduced into the host (i.e., a plant) and expressed temporally.

Transformation protocols as well as protocols for introducing polynucleotide sequences into plants may vary depending on the type of plant or plant cell, i.e., monocot or dicot, targeted for transformation. Suitable methods of introducing polynucleotides into plant cells include microinjection (Crossway *et al.* (1986) *Biotechniques* 4:320-334), electroporation (Riggs *et al.* (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-5606, *Agrobacterium*-mediated transformation (Townsend *et al.*, U.S. Patent No. 5,563,055; Zhao *et al.*, U.S. Patent No. 5,981,840), direct gene transfer (Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722), and ballistic particle acceleration (see, for example, Sanford *et al.*, U.S. Patent No. 4,945,050; Tomes *et al.*, U.S. Patent No. 5,879,918; Tomes *et al.*, U.S. Patent No. 5,886,244; Bidney *et al.*, U.S. Patent No. 5,932,782; Tomes *et al.* (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment," in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg and Phillips (Springer-Verlag, Berlin); McCabe *et al.* (1988) *Biotechnology* 6:923-926); and *Lec1* transformation (WO 00/28058).

Also see Weissinger *et al.* (1988) *Ann. Rev. Genet.* 22:421-477; Sanford *et al.* (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou *et al.* (1988) *Plant Physiol.* 87:671-674 (soybean); McCabe *et al.* (1988) *Bio/Technology* 6:923-926 (soybean); Finer and McMullen (1991) *In Vitro Cell Dev. Biol.* 27P:175-182 (soybean); Singh *et al.* (1998) *Theor. Appl. Genet.* 96:319-324 (soybean); Datta *et al.* (1990) *Biotechnology* 8:736-740 (rice); Klein *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:4305-4309 (maize); Klein *et al.* (1988) *Biotechnology* 6:559-563 (maize); Tomes, U.S. Patent No. 5,240,855; Buising *et al.*, U.S. Patent Nos. 5,322,783 and 5,324,646; Tomes *et al.* (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment," in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg (Springer-Verlag, Berlin) (maize); Klein *et al.* (1988) *Plant Physiol.* 91:440-444 (maize); Fromm *et al.* (1990) *Biotechnology* 8:833-839 (maize); Hooykaas-Van Slogteren *et al.* (1984) *Nature (London)* 311:763-764; Bowen *et al.*, U.S. Patent No. 5,736,369 (cereals); Bytebier *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:5345-5349 (Liliaceae); De Wet *et al.* (1985) in *The Experimental Manipulation of Ovule Tissues*, ed. Chapman *et al.* (Longman, New York), pp. 197-209 (pollen); Kaeppler *et al.* (1990) *Plant Cell Reports* 9:415-418 and Kaeppler *et al.* (1992) *Theor. Appl. Genet.* 84:560-566 (whisker-mediated transformation); D'Halluin *et al.* (1992) *Plant Cell* 4:1495-1505 (electroporation); Li *et al.* (1993) *Plant Cell Reports* 12:250-255 and Christou and Ford (1995) *Annals of Botany* 75:407-413 (rice); Osjoda *et al.*

(1996) *Nature Biotechnology* 14:745-750 (maize via *Agrobacterium tumefaciens*); all of which are herein incorporated by reference.

In specific embodiments, the recombinant polynucleotides disclosed herein can be provided to a plant using a variety of transient transformation methods. Such transient transformation methods include, but are not limited to, the introduction of the recombinant polynucleotide or variants thereof directly into the plant. Such methods include, for example, microinjection or particle bombardment. See, for example, Crossway *et al.* (1986) *Mol Gen. Genet.* 202:179-185; Nomura *et al.* (1986) *Plant Sci.* 44:53-58; Hepler *et al.* (1994) *Proc. Natl. Acad. Sci.* 91: 2176-2180 and Hush *et al.* (1994) *The Journal of Cell Science* 107:775-784, all of which are herein incorporated by reference. Alternatively, the polynucleotides can be transiently transformed into the plant using techniques known in the art. Such techniques include viral vector system and the precipitation of the polynucleotide in a manner that precludes subsequent release of the DNA. Thus, the transcription from the particle-bound DNA can occur, but the frequency with which it is released to become integrated into the genome is greatly reduced. Such methods include the use of particles coated with polyethylimine (PEI; Sigma #P3143).

In other embodiments, recombinant polynucleotides disclosed herein may be introduced into plants by contacting plants with a virus or viral nucleic acids. Generally, such methods involve incorporating a nucleotide construct provided herein within a viral DNA or RNA molecule. Methods for introducing polynucleotides into plants and expressing a protein encoded therein, involving viral DNA or RNA molecules, are known in the art. See, for example, U.S. Patent Nos. 5,889,191, 5,889,190, 5,866,785, 5,589,367, 5,316,931, and Porta *et al.* (1996) *Molecular Biotechnology* 5:209-221; herein incorporated by reference.

Methods are known in the art for the targeted insertion of a polynucleotide at a specific location in the plant genome. In one embodiment, the insertion of the polynucleotide at a desired genomic location is achieved using a site-specific recombination system. See, for example, WO99/25821, WO99/25854, WO99/25840, WO99/25855, and WO99/25853, all of which are herein incorporated by reference. Briefly, the recombinant polynucleotides provided herein can be contained in a transfer cassette flanked by two non-identical recombination sites. The transfer cassette is introduced into a plant having stably incorporated into its genome a target site which is flanked by two non-identical recombination sites that correspond to the

sites of the transfer cassette. An appropriate recombinase is provided and the transfer cassette is integrated at the target site. The recombinant polynucleotide is thereby integrated at a specific chromosomal position in the plant genome.

The cells that have been transformed may be grown into plants in accordance with conventional ways. See, for example, McCormick *et al.* (1986) *Plant Cell Reports* 5:81-84. These plants may then be grown, and either pollinated with the same transformed strain or different strains, and the resulting progeny having constitutive expression of the desired phenotypic characteristic identified. Two or more generations may be grown to ensure that expression of the desired phenotypic characteristic is stably maintained and inherited and then seeds harvested to ensure expression of the desired phenotypic characteristic has been achieved. In this manner, transformed seed (also referred to as "transgenic seed") having a recombinant polynucleotide disclosed herein, for example, an expression cassette provided herein, stably incorporated into their genome is provided.

III. Methods of Use

Provided herein is a method of targeting a polypeptide of interest to a chloroplast comprising expressing a recombinant polynucleotide encoding a CTP provided herein (i.e. Any of the chimeric CTPs provided herein and/or SEQ ID NOS: 1, 2, 3, 4, 5, 6, 7 or active variants or fragments thereof) operably linked to a heterologous polynucleotide encoding a polypeptide of interest in a cell, plant cell, plant, plant part or seed.

The methods further provide a CTP comprising a chimeric CTP, a chimeric CTP comprising at least one chimeric domain, or a CTP comprising a consensus sequence as described in detail elsewhere herein. The recombinant polynucleotide provided in the methods can comprise a CTP provided herein linked to any polypeptide of interest. For example, the polypeptide of interest can comprise an insecticidal protein whose expression controls a pest, a *Bacillus thuringiensis* polypeptide having insecticidal activity, or an IP2-127 polypeptide (i.e. SEQ ID NO:12 or an active variant or fragment thereof).

Methods of the present invention are directed to the proper expression, translocation, and processing of chloroplast-targeted sequences in plants and plant cells under the control of the CTP sequences disclosed herein. For the purposes of the present invention, a "processed" chloroplast targeted protein is one in which the CTP

has been removed. At the time of translocation of a chloroplast targeted protein into the chloroplast of a plant cell, the CTP is removed from the targeted protein by cleavage at a particular "cleavage site" between the CTP and the mature protein. The cleavage site can be determined experimentally, or may be predicted based on

5 sequence structure (e.g., by alignment of the unprocessed protein with chloroplast targeted proteins in which the cleavage site is known, by analyzing the sequence for the presence of characteristic CTP domains, and the like) or by using one or more algorithms for cleavage site prediction (e.g., SignalP or PSORT).

Depending on the polypeptide of interest targeted to the chloroplast, the

10 transgenic plants may have a change in phenotype, including, but not limited to, an altered pathogen or insect defense mechanism, an increased resistance to one or more herbicides, an increased ability to withstand stressful environmental conditions, a modified ability to produce starch, a modified level of starch production, a modified oil content and/or composition, a modified ability to utilize, partition and/or store

15 nitrogen, and the like. These results can be achieved through the expression and targeting of a polypeptide of interest to chloroplasts in plants, wherein the polypeptide of interest functions in the chloroplast. The CTP sequences provided herein are useful for targeting native sequences as well as heterologous (non-native) sequences in plants.

20

Non-limiting examples of methods and compositions disclosed herein are as follows:

1. A recombinant polynucleotide encoding a chloroplast transit peptide (CTP) operably linked to a heterologous polynucleotide encoding a polypeptide of interest,

25 wherein the CTP comprises

a) an amino acid sequence comprising the amino acids of SEQ ID NOS: 6 or 7;

b) an amino acid sequence having at least 85% sequence identity to SEQ ID NOS: 6 or 7, wherein said amino acid sequence has CTP activity; or,

30 c) an amino acid sequence having at least 17 consecutive amino acids of SEQ ID NOS: 6 or 7, wherein said amino acid sequence has CTP activity.

2. A recombinant polynucleotide encoding a chimeric chloroplast transit peptide (CTP) operably linked to a heterologous polynucleotide encoding a polypeptide of interest, wherein said chimeric CTP comprises an N-terminal domain,

a central domain, and a C-terminal domain, or variant thereof, wherein at least one of said N-terminal domain, said central domain, said C-terminal domain or variant thereof is heterologous to at least one of said domains.

3. The recombinant polynucleotide of embodiment 2, wherein said N-terminal domain, said central domain or said C-terminal domain is from a CTP from *Oryza sativa* 1-deoxy-D xyulose-5-Phosphate Synthase, *Oryza sativa* -Superoxide dismutase, *Oryza sativa* -soluble starch synthase, *Oryza sativa* -NADP-dependent Malic acid enzyme, *Oryza sativa* -Phospho-2-dehydro-3-deoxyheptonate Aldolase 2, *Oryza sativa* -L-Ascorbate peroxidase 5, *Oryza sativa* -Phosphoglucan water dikinase, *Zea Mays* ssRUBISCO, *Zea Mays* -beta-glucosidase, *Zea Mays* -Malate dehydrogenase, *Zea Mays* Thioredoxin M-type or active variants thereof.

4. The recombinant polynucleotide of embodiment 2 or 3, wherein said N-terminal domain is from a CTP from *Oryza sativa* 1-deoxy-D xyulose-5-Phosphate Synthase, *Oryza sativa* -NADP-dependent Malic acid enzyme, *Zea Mays* -Malate dehydrogenase or active variants thereof.

5. The recombinant polynucleotide of embodiment 2 or 3, wherein said central domain is from a CTP from *Oryza sativa* -Superoxide dismutase, *Oryza sativa* -Phospho-2-dehydro-3-deoxyheptonate Aldolase 2, *Oryza sativa* -L-Ascorbate peroxidase 5, *Zea Mays* ssRUBISCO or active variants thereof.

6. The recombinant polynucleotide of embodiment 2 or 3, wherein said C-terminal domain is from a CTP from *Oryza sativa* -soluble starch synthase, *Oryza sativa* -Superoxide dismutase, *Oryza sativa* -Phosphoglucan water dikinase, *Zea Mays* Thioredoxin M-type, *Zea Mays* -beta-glucosidase or active variants thereof.

7. The recombinant polynucleotide of embodiment 2 or 3, wherein said N-terminal domain is from the *Oryza sativa* 1-deoxy-D xyulose-5-Phosphate Synthase CTP or an active variant thereof, said central domain is from the *Zea Mays* ssRUBISCO CTP or an active variant thereof and said C-terminal domain is from the *Zea Mays* -beta-glucosidase CTP or an active variant thereof.

8. The recombinant polynucleotide of embodiment 2 or 3, wherein said N-terminal domain is from the *Zea Mays* -Malate dehydrogenase CTP or an active variant thereof, said central domain is from the *Oryza sativa* -Superoxide dismutase CTP or an active variant thereof and said C-terminal domain is from the *Oryza sativa* -soluble starch synthase CTP or an active variant thereof.

9. The recombinant polynucleotide of embodiment 2 or 3, wherein said N-terminal domain is from the *Oryza sativa* -NADP-dependent Malic acid enzyme CTP or an active variant thereof, said central domain is from the *Oryza sativa* -Phospho-2-dehydro-3-deoxyheptonate Aldolase 2 CTP or an active variant thereof
 5 and said C-terminal domain is from the *Zea Mays* Thioredoxin M-type CTP or an active variant thereof.

10. The recombinant polynucleotide of embodiment 2, wherein at least one of said N-terminal domain, said central domain, or said C-terminal domain comprises a chimeric domain.

10 11. The recombinant polynucleotide of embodiment 10, wherein at least one portion of said chimeric N-terminal domain is from the N-terminal domain of the *Oryza sativa* -NADP-dependent Malic CTP, *Zea Mays* -Malate dehydrogenase CTP or active variants thereof.

12. The recombinant polynucleotide of embodiment 10, wherein at least
 15 one portion of said chimeric central domain is from the central domain of the *Oryza sativa* -L-Ascorbate peroxidase 5 CTP, *Zea Mays* ssRUBISCO CTP or active variants thereof.

13. The recombinant polynucleotide of embodiment 10, wherein at least one portion of said chimeric C-terminal domain is from the C-terminal domain of the
 20 *Oryza sativa* -soluble starch synthase CTP, *Zea Mays* Thioredoxin M-type CTP, *Oryza sativa* -Superoxide dismutase CTP, *Oryza sativa* -Phosphoglucan water dikinase CTP or active variants thereof.

14. The recombinant polynucleotide of embodiment 10, wherein said chimeric CTP comprises

25 a) a chimeric N-terminal domain, wherein said chimeric N-terminal domain comprises a portion of the N-terminal domain from the *Zea Mays* -Malate dehydrogenase CTP fused in frame to a portion of the N-terminal domain of the *Oryza sativa* -NADP-dependent Malic acid enzyme CTP;

b) a central domain, wherein said central domain is from the *Zea Mays*
 30 ssRUBISCO CTP; and,

c) a chimeric C-terminal domain, wherein said chimeric C-terminal domain comprises a portion of the C-terminal domain from the *Oryza sativa* -soluble starch synthase CTP fused in frame to a portion of the C-terminal domain from the *Zea Mays* Thioredoxin M-type CTP;

wherein said chimeric CTP has CTP activity.

15. The recombinant polynucleotide of embodiment 10, wherein said chimeric CTP comprises

5 a) a chimeric N-terminal domain, wherein said chimeric N-terminal domain comprises a portion of the N-terminal domain from the *Zea Mays* -Malate dehydrogenase CTP fused in frame to a portion of the N-terminal domain of the *Oryza sativa* -NADP-dependent Malic acid enzyme CTP;

10 b) a chimeric central domain, wherein said chimeric central domain comprises a portion of the central domain from the *Oryza sativa* -L-Ascorbate peroxidase 5 CTP fused in frame to a portion of the central domain of the *Zea Mays* ssRUBISCO CTP; and,

15 c) a chimeric C-terminal domain, wherein said chimeric C-terminal domain comprises a portion of the C-terminal domain from the *Oryza sativa* -Superoxide dismutase CTP fused in frame to a portion of the C-terminal domain of the *Oryza sativa* -Phosphoglucan water dikinase CTP;

wherein said chimeric CTP has CTP activity.

16. The recombinant polynucleotide of embodiment 3, wherein the chimeric CTP comprises

20 a) an amino acid sequence comprising the amino acids of SEQ ID NOS: 1, 2 or 3;

b) an amino acid sequence having at least 85% sequence identity to SEQ ID NOS: 1, 2 or 3, wherein said amino acid sequence has CTP activity; or

c) an amino acid sequence having at least 17 consecutive amino acids of SEQ ID NOS: 1, 2 or 3, wherein said amino acid sequence has CTP activity.

25 17. The recombinant polynucleotide of embodiment 14, wherein the chimeric CTP comprises

a) an amino acid sequence comprising the amino acids of SEQ ID NO: 4;

30 b) an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 4, wherein said amino acid sequence has CTP activity; or

c) an amino acid sequence having at least 17 consecutive amino acids of SEQ ID NO: 4, wherein said amino acid sequence has CTP activity.

18. The recombinant polynucleotide of embodiment 15, wherein the chimeric CTP comprises

a) an amino acid sequence comprising the amino acids of SEQ ID NO: 5;

b) an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 5, wherein said amino acid sequence has CTP activity; or

5 c) an amino acid sequence having at least 17 consecutive amino acids of SEQ ID NO: 5, wherein said amino acid sequence has CTP activity.

19. The recombinant polynucleotide of any one of embodiments 1-18, wherein said polypeptide of interest comprises a *Bacillus thuringiensis* polypeptide having insecticidal activity.

10 20. The recombinant polynucleotide of embodiment 19, wherein said *Bacillus thuringiensis* polypeptide having insecticidal activity comprises an IP2-127 polypeptide.

21. A nucleic acid construct comprising the recombinant polynucleotide of any one of embodiments 1-20.

15 22. The nucleic acid construct of embodiment 21, further comprising a promoter operably linked to said recombinant polynucleotide.

23. A cell comprising at least one recombinant polynucleotide of any of embodiments 1-20 or the nucleic acid construct of any one of embodiments 21 or 22.

24. The cell of embodiment 23, wherein said cell is a plant cell.

20 25. The cell of embodiment 24, wherein said polynucleotide or nucleic acid construct is stably incorporated into the genome of said plant cell.

26. The cell of any one of embodiments 24 or 25, wherein said plant cell is from a monocot.

25 27. The cell of embodiment 26, wherein said monocot is maize, wheat, rice, barley, sorghum, sugarcane or rye.

28. The cell of any one of embodiments 24 or 25, wherein said plant cell is from a dicot.

29. The cell of embodiment 28, wherein the dicot is soybean, Brassica, sunflower, cotton or alfalfa.

30 30. A plant comprising at least one plant cell of any one of embodiments 24-29.

31. A plant explant comprising at least one plant cell of any one of embodiments 24-29.

32. A transgenic seed produced by the plant of embodiment 30, wherein said seed comprises said recombinant polynucleotide.

33. A recombinant polypeptide encoded by the polynucleotide of any one of embodiments 1-20.

5

34. A method of targeting a polypeptide of interest to a chloroplast comprising expressing the recombinant polynucleotide of any one of embodiments 1-20 or the nucleic acid construct of embodiment 21 or 22 in a plant cell.

35. A method of targeting a polypeptide of interest to a chloroplast
10 comprising introducing the recombinant polynucleotide of any one of embodiments 1-20 or the nucleic acid construct of embodiment 21 or 22 in a plant cell and expressing said recombinant polynucleotide in the plant cell.

36. The method of embodiment 34 or 35, wherein said method further comprises regenerating a transgenic plant from said plant cell.

15 37. The method of any one of embodiments 34-36, wherein said plant cell is from a monocot.

38. The method of embodiment 37, wherein said monocot is selected from the group consisting of maize, wheat, rice, barley, sorghum, sugarcane or rye.

39. The method of any one of embodiments 34-36, wherein said plant cell
20 is from a dicot.

40. The method of embodiment 39, wherein said dicot is selected from the group consisting of soybean, Brassica, sunflower, cotton or alfalfa.

41. The method of any one of embodiments 35-40, wherein said
25 polypeptide of interest comprises an insecticidal protein and expression of said polypeptide controls a pest.

42. The method of embodiment 41, wherein said polypeptide of interest comprises a *Bacillus thuringiensis* polypeptide having insecticidal activity.

43. The method of embodiment 42, wherein said *Bacillus thuringiensis* polypeptide having insecticidal activity comprises an IP2-127 polypeptide.

30

EXPERIMENTAL

The following examples are offered to illustrate, but not to limit, the claimed invention. It is understood that the examples and embodiments described herein are for illustrative purposes only, and persons skilled in the art will recognize various

reagents or parameters that can be altered without departing from the spirit of the invention or the scope of the appended claims.

Example 1. Development of novel chloroplast targeting peptides (CTPs) for maize.

Nuclear encoded plant proteins that are translated in the cytosol are targeted to the chloroplast using an N-terminal transit peptide. CTPs are both necessary and sufficient for correct chloroplast targeting and these signal peptides are of variable length and sequence. Although there is no consensus peptide sequence CTPs do share a similar structural framework consisting of an uncharged N-terminus, a central region lacking acidic residues but enriched in hydroxylated amino acids, and a basic arginine-rich amphipathic C-terminus.

Two approaches were employed to develop these targeting peptides based on the alignment of a set of known or predicted chloroplast transit peptides from monocotyledonous species (see Fig. 1). The first approach utilized was to generate chimeric CTPs based on the predicted boundaries of the different CTP domains described above. The new CTPs can be derived from 3 or more different plant CTP sequences that when combined together collectively reconstitute a CTP. A set of 5 different chimeric CTPs were generated based on the CTP alignment found in Fig 1.

msCTP1 (SEQ ID NO: 1:

MALTTFSISRGGFVGALQGLKSTASLPNNESFSRHHLPPSSSPQSSKRRCNLSFT
TR) was generated from the combination of domains in sequential order from *Oryza sativa* (Os) 1-deoxy-D xylulose-5-Phosphate Synthase CTP (aa 1-17), *Zea mays* (Zm) ssRUBISCO CTP (aa 18-27) and Zm-beta-glucosidase CTP (aa 28-56). msCTP2

(SEQ ID NO: 2:

MGLSTVYSPAGPRLVPAPASLFQSPSSGCHSCWGP GPGGGRRRLPS
PRRRPITGTRS) was generated from the combination of domains in sequential order from Zm-Malate dehydrogenase (NADP) CTP (aa 1-17), Os-Superoxide dismutase (SOD) CTP (aa 18- 27) and Os-Soluble starch synthase CTP (aa 28-52). msCTP3

(SEQ ID NO: 3:

MLSARAAATAAAAAAASPQPRLAATFLVLPSKRALAPLLSVGRVA
TRRPRHVCQ) was generated from the combination of the following domains in sequential order from Os-NADP-dependent Malic acid enzyme CTP (aa 1-17), Os-

Phospho-2-dehydro-3-deoxyheptonate (PHD) Aldolase 2 CTP (aa 18-27), and Zm Thioredoxin M-type (TRX) CTP (aa 28-54). msCTP4 (SEQ ID NO: 4:

MGLSTVYSPAAAAAASPPQPRSTASLPGCHSCWGPGLLSVGRVATRRPR

HVCQ) was generated with the combination of domains in sequential order from Zm-

5 Malate dehydrogenase (NADP) CTP (aa 1-9), Os-NADP-dependent Malic acid enzyme CTP CTP (aa 10-21), Zm-ssRUBSICO CTP(aa 22-27), Os-Soluble starch synthase CTP (aa 28-37), and Zm-Thioredoxin (TRX) CTP (aa 38-54). The design of msCTP4 incorporated sequences derived from 2 separate CTPs for the first and third domains. msCTP5 (SEQ ID NO: 5:

10 MGLSTVYSPAAAAAASPPSLRSTASLPARPFHSLRLAAG

RRGFACRGRSAAS) was generated with the combination of domains in sequential order from Zm-Malate dehydrogenase (NADP) CTP (aa 1-9), Os-NADP-dependent Malic acid enzyme CTP CTP (aa 10-17), Os- L-Ascorbate peroxidase 5(OsAPx05) (aa 18-21); Zm-ssRUBSICO CTP (aa 22-27), Os-Superoxide dismutase (OsSOD) 15 CTP (aa 28-39), and Os-Phosphoglucan water dikinase (OsPGDK) CTP (aa 40-52).

The second approach used the most frequent amino acid at each position based on the alignment of the different CTPs that were of a similar size (50-60 aa). In some cases where no dominant amino acid residue was apparent one of the more frequent amino acid residues was chosen to be incorporated into the sequence. Two CTPs were 20 developed using this strategy. msCTP6 (SEQ ID NO: 6:

MALASVMAAAAASVVSFPAGRGSGG

SSVLRSLSLAGSRRSAAAVRRLAL) and msCTP7 (SEQ ID NO: 7:

MAVATVLAAAALAAVSPPGLRSSLGFPVRRSLPSAARGGSPAATRRRCRAA).

A comparison of the amino acid identity levels for the different CTPs 25 developed using this strategy is found in Table 1. The homology between all the CTPs ranged from 16 – 64%.

Table 1: msCTP identity table.

	msCTP1	msCTP3	msCTP4	msCTP5	msCTP2	msCTP6	msCTP7
msCTP1		16	26	28	16	20	22
msCTP3			64	38	21	38	36
msCTP4				59	46	31	36
msCTP5					33	34	44
msCTP2						35	26
msCTP6							47
msCTP7							

Example 2. Construction of vectors for testing the ability of the novel CTPs to target an insecticidal toxin to the chloroplast.

5 A transient expression vector was generated to evaluate the ability of the novel CTPs to target an insecticidal toxin, IP2-127, to the maize chloroplast. This vector contained a fusion gene with IP2-127 at the N-terminus and AcGFP at the C terminus separated by a short linker sequence (SEQ ID NO: 8:

10 ATGGGCAACAGCGTGCTCAACAGCGGACGCACCACCATCTGCGACGCCTA
 CAACGTGGCCGCGCACGACCCGTTTCAGCTTCCAGCACAAAGAGCCTCGACA
 CCGTGCAGCGCGAGTGGACCGAGTGGAAGAAGAACAACCACAGCCTCTA
 CCTCGACCCGATCGTGGGCACCGTGGCCAGCTTCCTCCTCAAGAAGGTGG
 GCAGCCTCGTGGGCAAGCGCATCCTCAGCGAGCTGCGCAACCTCATCTTC
 CCGAGCGGCAGCACCAACCTCATGCAGGACATCCTCCGCGAGACCGAGCA
 15 GTTCCTCAACCAGCGCCTCGACACCGACACCCTCGCCAGGGTGAACGCCG
 AGCTGACCGGCCTCCAGGCCAACGTGGAGGAGTTCAACCGCCAGGTGGAC
 AACTTCCTCAACCCGAACCGCAACGCCGTGCCGCTCAGCATCACCAGCAG
 CGTGAACACCATGCAGCAGCTCTTCCTCAACCGCCTCCCGCAGTTCCAGAT
 GCAGGGCTACCAGCTCCTGCTCCTGCCGCTCTTCGCCCAGGCCGCCAACCT
 20 CCACCTCAGCTTCATCCGCGACGTGATCCTCAACGCCGACGAGTGGGGCA
 TCAGCGCCGCCACCCTCCGCACCTACCGCGACTACCTCAAGAACTACACC
 CGCGACTACAGCAACTACTGCATCAACACCTACCAGAGCGCCTTCAAGGG

CCTCAACACCCGCCTCCACGGCACCCCTCGAGTTCCGCACCTACATGTTTCCT
CAACGTCTTCGAGTACGTGAGCATCTGGAGCCTCTTCAAGTACCAGAGCC
TCCTCGTGAGCAGCGGGCGCCAACCTCTACGCCAGCGGGCAGCGGGCCCGCAG
CAGACCCAGAGCTTCACCAGCCAGGACTGGCCGTTCTCTACAGCCTCTTC
5 CAGGTGAACAGCAACTACGTGCTCAACGGCTTCAGCGGGCGCCAGGCTCAG
CAACACCTTCCCGAACATCGGGCGGCCTCCCGGGCAGCACCACCACCCACG
CCCTCCTCGCGGGCCAGGGTGAACCTACAGCGGGCGGCATCAGCAGCGGGCGAC
ATCGGGCGCCAGCCCGTTCAACCAGAACTTCAACTGCAGCACCTTCCTCCC
GCCGCTCCTCACCCCGTTCTGTGCGCAGCTGGCTCGATAGCGGCAGCGACC
10 GCGAGGGCGTGGCCACCGTGACCAACTGGCAGACCGAGAGCTTCGAGAC
CACACTCGGGCTCAGGAGCGGGCGCCTTCACCGCCCGCGGCAACAGCAACT
ACTTCCCGGACTACTTCATCCGGAACATCTCCGGCGTTCCGTTGGTGGTCC
GTAACGAGGATCTCAGGAGGCCGCTGCACTACAACGAGATCCGCAACATC
GCTTCGCCCAGCGGGACCCCAAGGTGGAGCACGGGCCTACATGGTGTCCGT
15 GCACAACCGGAAGAACAACATCCACGCGGTCCATGAGAACGGCAGCATG
ATCCACCTGGCTCCTAACGACTACACGGGGTTTACAATCTCTCCGATCCAT
GCTACTCAAGTCAACAACCAGACCAGGACGTTTCATCTCGGAGAAGTTCGG
CAACCAGGGAGACTCCTTGAGGTTCGAGCAGAACACAACAACTGCCCCGCT
ACACCCTTCGGGGCAACGGGAACAGCTACAACCTCTACCTGCGCGTCAGC
20 TCCATCGGCAACTCGACGATCAGGGTCACGATCAACGGAAGGGTCTACAC
TGCGACCAACGTGAACACGACAATAACAACGACGGCGTCAACGACAAC
GGCGCTAGGTTCTCCGACATCAACATCGGGAACGTTGTGGCAAGCTCCAA
CTCGGATGTCCCTCTTGACATCAACGTCACCTTCAACTCTGGAACGCAGTT
CGATCTGATGAACACAATGCTGGTGCCAACTAACATCAGCCCTCTGTACG
25 GTGGAGGCGGCAGCGGTGGCGGAGGCTCCGGAGGCGGTGGCTCCATGGT
GAGCAAGGGCGCCGAGCTGTTACCGGCATCGTGCCCATCCTGATCGAGC
TGAATGGCGATGTGAATGGCCACAAGTTCAGCGTGAGCGGGCGAGGGCGA
GGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCG
GCAAGCTGCCTGTGCCCTGGCCCACCCTGGTGACCACCCTGAGCTACGGC
30 GTGCAGTGCTTCTCACGCTACCCCGATCACATGAAGCAGCACGACTTCTTC
AAGAGCGCCATGCCTGAGGGCTACATCCAGGAGCGCACCATCTTCTTCGA
GGATGACGGCAACTACAAGTCGCGCGCCGAGGTGAAGTTCGAGGGCGAT
ACCCTGGTGAATCGCATCGAGCTGACCGGCACCGATTTCAAGGAGGATGG
CAACATCCTGGGCAATAAGATGGAGTACAACGCCCACAATGTGT

ACATCATGACCGACAAGGCCAAGAATGGCATCAAGGTGAACTTCAAGATC
CGCCACAACATCGAGGATGGCAGCGTGCAGCTGGCCGACCACTACCAGCA
GAATACCCCATCGGGCGATGGCCCTGTGCTGCTGCCCCGATAACCACTACC
TGTCACCCAGAGCGCCCTGTCCAAGGACCCCAACGAGAAGCGCGATCAC
5 ATGATCTACTTCGGCTTCGTGACCGCCGCCCATCACCCACGGCATGGAT
GAGCTGTACAAGTGA) which encoded a IP2-127-AcGFP fusion protein (SEQ ID
NO: 9:

MGNSVLNSGRITICDAYNVAAHDPFSFQHKSLDTVQREWTEWKKNNHSL
YLDPIVGTVASFLLKKVGSLVGKRILSELRLNIFPSGSTNLMQDILRETEQFLN
10 QRLDTDTLARVNAELTGLQANVEEFNRQVDNFLNPNRNAVPLSITSSVNTMQ
QLFLNRLPQFQMGGYQLLLLPLFAQAANLHLSFIRDVILNADEWGISAATLRT
YRDYLKNYTRDYSNYCINTYQSAFKGLNTRLHGTLEFRTYMFLNVFEYVSIW
SLFKYQSLLVSSGANLYASGSGPQQTQSFTSQDWPLYSLFQVNSNYVLNGFS
GARLSNTFPNIGGLPGSTTTHALLAARVNYSGGISSGDIGASPFNQNFNCSTFL
15 PPLLTPFVRSWLDSDREGVATVTNWQTESFETTLGLRSGAFTARGNSNYFP
DYFIRNISGVPLVVRNEDLRRPLHYNEIRNIASPSGTPGGARAYMVSVHNRKN
NIHAVHENGSMIHLAPNDYTGFTISPIHATQVNNQTRTFISEKFGNQGDSLRF
QNNTTARYTLRGNGNSYNLYLRVSSIGNSTIRVTINGRVYTATNVNTTTNND
GVNDNGARFSDINIGNVASSNSDVPLDINVTFNSGTQFDLMNTMLVPTNISP
20 LYGGGGSGGGGSGGGGSMVSKGAELFTGIVPILIELNGDVNGHKFSVSGEGE
GDATYGKLTCLKFICTTGKLPVPWPTLVTTLSYGVQCFSRYPDHMKQHDFFKS
AMPEGYIQERTIFFEDDGNYKSRAEVKFEGDTLVNRIELTGTDFKEDGNILGN
KMEYNYNAHNVYIMTDKAKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGD
GPVLLPDNHYLSTQSALSKDPNEKRDHMIYFGFVTAATHGMDELYK)

25 consisting of IP2-127 from amino acid 1-634, a short 15 aa linker from amino acid
635-649, and AcGFP from amino acid 650-888. The IP2-127::AcGFP fusion gene is
under control of the strong constitutive maize Ubiquitin 1 promoter-5'UTR- intron1
regulatory element in vector pSK-UBI-IP2-127::AcGFP with a pinII transcriptional
terminator sequence. The vector contains unique BamHI and KpnI restriction enzyme
30 sites immediately upstream of the IP2-127 translational start codon to facilitate an in
frame insertion of different CTP sequences at the N-terminus of the fusion.

The different novel monocot CTPs were synthesized by DNA2.0 (Menlo park,
CA). Each CTP was subcloned into pSK-UBI-IP2-127::AcGFP using the unique

BamHI and KpnI restriction sites. The base vector, pSK-UBI-IP2-127::AcGFP, was used as a control for non-CTP targeted IP2-127::AcGFP. A vector containing IP2-127::AcGFP fused to a previously characterized CTP derived by gene shuffling (6H1-CTP) (SEQ ID NO: 10:

5 MAATTLTSALPGAFSSSQRPSPFNLQRSPRVLRRFNRKTGRQ
PRGLVRAAKAQ) was used as a positive control for chloroplast targeting in transient expression assays.

Example 3: Transient expression assays to identify novel CTPs effective at targeting IP2-127::AcGFP to the maize chloroplast.

10 Maize seedlings were generated in soilless artificial condition by embedding kernels between two sheets of seed germination paper in a roll and its bottom portion was submerged in 0.1 mg/ml sucrose solution. Leaf segments were detached from seedlings at 15 days post-planting immediately before ballistic co-bombardment with
15 colloidal gold particles transformation. The lower epidermis of the leaf segments were excised and overlaid on top of filter papers in 100 mm Petri dishes.

The samples were co-bombarded with DNAs from both a DS-RED plasmid vector and individual CTP testing vectors using the PDS-1000 He biolistic particle delivery system (Bio-Rad, Hercules CA). Gold particles (1.0 μ m in diameter; Bio-
20 Rad) were coated with plasmid DNAs following the procedure described by Sanford et al. (1993) with modifications. Briefly, 50 μ l of freshly prepared gold particles in water (20 mg/ml), and 20 μ l of DNA mixture, which contain 10 μ g of equimolar quantities of the DS Red helper plasmid and CTP testing plasmids, were combined and 50 μ l of a 2.5 M CaCl_2 solution and 20 μ l of freshly prepared 0.1 M spermidine
25 (Sigma-Aldrich, St Louis MO) were slowly added with gentle vortexing. The mixture was incubated at room temperature for 5 min and pelleted at 13,000g in a microcentrifuge for 5 sec. The supernatant was carefully removed and the pellet was resuspended in 85 μ l of 100% ethanol. While gently vortexing, a 6 μ l aliquot of suspension was drawn and dispensed onto the center of a macrocarrier membrane.
30 The membrane was allowed to air dry completely for 2–5 min and used immediately. Leaf segments were bombarded at a distance of 9 cm from an 1100-psi rupture disk. Two replicate shots were performed from each coating preparation. After

bombardment, the leaf samples were incubated in a moist chamber at 28 degree Celsius.

Initial examination was conducted at approximately 24 h post-bombardment with a Lumar fluorescence stereomicroscope (Carl Zeiss Inc., Thornwood NY) equipped with both a green-emitting (Zeiss Set 10) and red-emitting (Zeiss Set 43 HE) filter set to image the AcGFP and the DsRed2, respectively. The leaf segments containing AcGFP-positive cells identified in the stereomicroscope were placed in a 0.01% Tween 20 solution and a vacuum was applied for about 10 min to remove internal air and to wet the leaf surface. The leaves were placed into coverglass chambers (Nalge Nunc International, Rochester NY) in the same solution, sealed with an additional coverglass and examined in the LSM510 (Carl Zeiss). AcGFP fluorescence was captured using a 488 nm argon laser for excitation and a 500-550 nm band pass emission filter. DsRed fluorescence was imaged using a 561 nm diode laser for excitation and a 575-615 nm band pass emission filter. Chlorophyll fluorescence was captured by combining 561 nm excitation and a 650-710 nm band pass emission filter.

DsRed expression was used to assess the overall transformation rate and was very useful for identifying transformed cells in the confocal microscope. Although epidermal cells were transformed with the highest frequency by the bombardment procedure, mesophyll cells were used to assess plastid targeting. Plastid targeting was confirmed by co-localizing the AcGFP signal with chlorophyll fluorescence. Plastid targeting was quantified with the confocal microscope by counting the number of mesophyll cells showing plastid-targeted AcGFP as a percentage of the total number of transformed cells (i.e., those exhibiting DsRed fluorescence).

The results of this analysis are outlined in Table 2. No colocalization of IP2-127::AcGFP with the chloroplast was observed in the non-targeted control where AcGFP fluorescence was limited exclusively to the cytosolic compartment. The majority of AcGFP derived fluorescence from the positive control, 6H1-CTP-IP2-127::AcGFP, was found to colocalize to chloroplasts and was scored at the highest level of +++. CTPs, msCTP1 and msCTP4, showed equivalent levels of chloroplast colocalization of IP2-127::AcGFP as observed with 6H1CTP. msCTP2 and msCTP6 directed IP2-127::AcGFP to the chloroplast although there was equal signal between chloroplast targeted and cytosolic localized fluorescence observed. This suggested that these two CTPs were not as efficient as msCTP1 or msCTP4 in chloroplast

targeting. msCTP5 directed more cytosolic localization of IP2-127::AcGFP than chloroplast colocalization but detectable levels of chloroplast colocalization was observed. msCTP7 failed to direct any IP2-127::AcGFP to the chloroplast and was similar to the non-targeted control where IP2-127::AcGFP was cytosolic.

5

Table 2: Effectiveness of chloroplast targeting of novel CTPs based on colocalization of AcGFP fluorescence with maize chloroplasts in transient expression assays.

Construct	Colocalization with chloroplasts
IP2-127::AcGFP (non-targeted control)	—
6H1CTP-IP2-127::AcGFP (targeted control)	+++
msCTP1-IP2-127::AcGFP	+++
msCTP2-IP2-127::AcGFP	++
msCTP3-IP2-127::AcGFP	—
msCTP4-IP2-127::AcGFP	+++
msCTP5-IP2-127::AcGFP	+
msCTP6-IP2-127::AcGFP	++
msCTP7-IP2-127::AcGFP	—

+++ IP2-127::AcGFP mostly chloroplast localized

++ equal IP2-127::GFP detected in chloroplast and cytosol.

10 + some IP2-127::AcGFP in chloroplast but mostly in cytosol

- IP2-127::GFP entirely in cytosol

Example 4. Transgenic plant evaluation of novel CTPs.

The effect of chloroplast targeting was extended from transient expression assays to stable transgenic events expressing IP2-127 with different msCTPs.

15 Chloroplast targeting of IP2-127 generally results in higher accumulation of IP2-127 in plants than would be observed when non-targeted. This difference in accumulation may be related to improved stability of IP2-127 in the chloroplast and/or phytotoxicity issues associated with high levels of accumulation of IP2-127 in the cytosol during the transformation process. Transformation vectors were generated using IP2-127 and
20 a subset of the msCTPs – mCTP1, mCTP2, msCTP4, msCTP5, msCTP6 – that were selected to represent the different qualitative results from the transient assays. Emphasis was put on those msCTPs that demonstrated some level of chloroplast colocalization in the transient experiments. The transformation vectors were generated

by using a base vector containing the IP2-127 gene with unique BamHI and KpnI restriction enzyme sites directly upstream of the translation start codon (ATG) of IP2-127. Subcloning each of the msCTPs into the BamHI and KpnI sites created an N-terminal fusion with the msCTP protein sequence and the IP2-127 protein sequence.

5 Non-targeted or msCTP targeted versions of the genes were placed under control of the maize Ubiquitin 1 promoter – 5'UTR- Ubiquitin intron1 and terminated with the pinII terminator sequence from Potato creating the msCTP test cassette. The msCTP test cassettes were introduced into a binary transformation vector using standard GatewayTM LR Clonase reactions that were facilitated by the presence of *attL3* and
10 *attL4* recombination sites flanking the test cassette and *attR3* and *attL4* sites in the destination transformation vector. The final product of the LR reaction was a transformation binary vector containing the msCTP-IP2-127 test cassette upstream of a cassette containing the maize Ubiquitin1 promoter-5'UTR-Ubiquitin intron1 controlling expression of a PAT selectable marker gene with the 35S terminator
15 sequence.

Transgenic events derived from this set of msCTP testing vectors were evaluated for expression of IP2-127 by ELISA. The results of the ELISA analysis are shown in Table 3. Accumulation of IP2-127 in the cytosol was 511 ppm. The addition of the 6H1-CTP to the N-terminus of IP2-127 improved accumulation ~2.5-fold to
20 1294 ppm demonstrating the effect of an effective CTP on IP21-27 accumulation in plants. Two CTPs, msCTP1 and msCTP4, improved IP2-127 accumulation ~7.7-fold and ~5-fold over the non-targeted version and 2-3-fold over the level of accumulation directed by 6H1-CTP. msCTP2 and msCTP6 showed comparable levels of accumulation as 6H1-CTP with both versions improving accumulation ~ 2.5-fold over
25 non-targeted IP2-127. No significant improvement in IP2-127 accumulation was observed with msCTP5 as the levels of accumulation were only about ~1.5-fold improved over the negative control. Overall, accumulation of IP2-127 in transgenic plants correlated well with the results of colocalization of IP2-127::AcGFP observed in transient expression assays. The results demonstrated that this strategy of
30 developing new synthetic CTPs was effective at providing novel chloroplast targeting peptides and that many of the msCTPs developed enhanced accumulation of IP2-127 in transgenic maize plants.

Table 3. Accumulation of IP2-127 in transgenic maize events.

Construct ID	No of Events Tested	IP2-127 Expression (PPM)
UBI-IP2-127	25	511
UBI-6H1-CTP-IP2-127	21	1294
UBI-msCTP1-IP2-127	23	3931
UBI-msCTP2-IP2-127	24	1408
UBI-msCTP4-IP2-127	25	2548
UBI-msCTP5-IP2-127	23	802
UBI-msCTP6-IP2-127	25	1406

Table 4. Summary of CTP domains.

CTP	N-terminal Domain	Central Domain	C-terminal Domain
msCTP1	SEQ ID NO: 24	SEQ ID NO: 25	SEQ ID NO: 26
msCTP2	SEQ ID NO: 27	SEQ ID NO: 28	SEQ ID NO: 29
msCTP3	SEQ ID NO: 30	SEQ ID NO: 31	SEQ ID NO: 32
msCTP4	SEQ ID NO: 33/ SEQ ID NO: 34	SEQ ID NO: 35	SEQ ID NO: 36/ SEQ ID NO: 37
msCTP5	SEQ ID NO: 38/ SEQ ID NO: 39	SEQ ID NO: 40/ SEQ ID NO: 41	SEQ ID NO: 42/ SEQ ID NO: 43

Table 5. Summary of SEQ ID NOS

SEQ ID NO	NA/AA	Description
1	AA	msCTP1
2	AA	msCTP2
3	AA	msCTP3
4	AA	msCTP4
5	AA	msCTP5
6	AA	msCTP6
7	AA	msCTP7
8	NA	Nucleotide sequence of IP2-127-AcGFP fusion protein
9	AA	Amino acid sequence of IP2-127-AcGFP fusion protein
10	AA	6H1-CTP (positive control CTP)
11	AA	CTP Consensus Sequence
12	AA	IP2-127 Amino Acid sequence
13	AA	OS- 1-deoxy-D-xyulose-5-Phosphate Synthase CTP
14	AA	OS- Superoxide dismutase CTP
15	AA	OS- soluble starch synthase CTP
16	AA	OS- NADP dependent Malic acid enzyme CTP
17	AA	OS-Phospho-2-dehydro-3-deoxyheptonate Aldolase 2 CTP
18	AA	OS- L-Ascorbate Peroxidase 5 CTP
19	AA	OS- Phosphoglucan water dikinase
20	AA	ZM- ssRUBISCO CTP
21	AA	ZM- beta-glucosidase CTP
22	AA	ZM- Malate dehydrogenase CTP
23	AA	ZM- Thioredoxin M-type
24	AA	Amino acids 1-17 of OS- 1-deoxy-D-xyulose-5-Phosphate Synthase CTP, N-terminal domain of CTP1
25	AA	Amino acids 18-27 of ZM- ssRUBISCO CTP, Central domain of CTP1
26	AA	Amino acids 28-56 of ZM- beta-glucosidase CTP, C-terminal domain of CTP1
27	AA	Amino acids 1-17 of ZM- Malate dehydrogenase CTP, N-terminal domain of CTP2
28	AA	Amino acids 18-27 of OS- Superoxide dismutase CTP, Central domain of CTP2
29	AA	Amino acids 28-52 of OS- soluble starch synthase CTP, C-terminal domain of CTP2
30	AA	Amino acids 1-17 of OS- NADP dependent Malic acid enzyme CTP, N-terminal domain of CTP3
31	AA	Amino acids 18-27 of OS-Phospho-2-dehydro-3-deoxyheptonate Aldolase 2 CTP, Central domain of CTP3

32	AA	Amino acids 28-54 of ZM- Thioredoxin M-type, C-terminal domain of CTP3
33	AA	Amino acids 1-9 of ZM- Malate dehydrogenase CTP, portion of N-terminal domain of CTP4
34	AA	Amino acids 10-21 of OS- NADP dependent Malic acid enzyme CTP, portion of N-terminal domain of CTP4
35	AA	Amino acids 22-27 of ZM- ssRUBISCO CTP, Central domain of CTP4
36	AA	Amino acids 28-37 of OS- soluble starch synthase CTP, portion of C-terminal domain of CTP4
37	AA	Amino acids 38-54 of ZM- Thioredoxin M-type, portion of C-terminal domain of CTP4
38	AA	Amino acids 1-9 of ZM- Malate dehydrogenase CTP, portion of N-terminal domain of CTP5
39	AA	Amino acids 10-17 of OS- NADP dependent Malic acid enzyme CTP, portion of N-terminal domain of CTP5
40	AA	Amino acids 18-21 of OS- L-Ascorbate Peroxidase 5 CTP, portion of central domain of CTP5
41	AA	Amino acids 22-27 of ZM- ssRUBISCO CTP, portion of central domain of CTP5
42	AA	Amino acids 28-39 of OS- Superoxide dismutase CTP, portion of C-terminal domain of CTP5
43	AA	Amino acids 40-52 of OS- Phosphoglucan water dikinase, portion of C-terminal domain of CTP5

The article “a” and “an” are used herein to refer to one or more than one (i.e., to at least one) of the grammatical object of the article. By way of example, “an element” means one or more element.

All publications and patent applications mentioned in the specification are indicative of the level of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

THAT WHICH IS CLAIMED:

1. A recombinant polynucleotide encoding a chloroplast transit peptide (CTP) operably linked to a heterologous polynucleotide encoding a polypeptide of interest, wherein the CTP comprises
 - a) an amino acid sequence comprising the amino acids of SEQ ID NOS: 6 or 7;
 - b) an amino acid sequence having at least 85% sequence identity to SEQ ID NOS: 6 or 7, wherein said amino acid sequence has CTP activity; or,
 - c) an amino acid sequence having at least 17 consecutive amino acids of SEQ ID NOS: 6 or 7, wherein said amino acid sequence has CTP activity.
2. A recombinant polynucleotide encoding a chimeric chloroplast transit peptide (CTP) operably linked to a heterologous polynucleotide encoding a polypeptide of interest, wherein said chimeric CTP comprises an N-terminal domain, a central domain, and a C-terminal domain, or variant thereof, wherein at least one of said N-terminal domain, said central domain, said C-terminal domain or variant thereof is heterologous to at least one of said domains.
3. The recombinant polynucleotide of claim 2, wherein said N-terminal domain, said central domain or said C-terminal domain is from a CTP from *Oryza sativa* 1-deoxy-D xyulose-5-Phosphate Synthase, *Oryza sativa* -Superoxide dismutase, *Oryza sativa* -soluble starch synthase, *Oryza sativa* -NADP-dependent Malic acid enzyme, *Oryza sativa* -Phospho-2-dehydro-3-deoxyheptonate Aldolase 2, *Oryza sativa* -L-Ascorbate peroxidase 5, *Oryza sativa* -Phosphoglucan water dikinase, *Zea Mays* ssRUBISCO, *Zea Mays* -beta-glucosidase, *Zea Mays* -Malate dehydrogenase, *Zea Mays* Thioredoxin M-type or active variants thereof.
4. The recombinant polynucleotide of claim 2 or 3, wherein said N-terminal domain is from a CTP from *Oryza sativa* 1-deoxy-D xyulose-5-Phosphate Synthase, *Oryza sativa* -NADP-dependent Malic acid enzyme, *Zea Mays* -Malate dehydrogenase or active variants thereof.

5. The recombinant polynucleotide of claim 2 or 3, wherein said central domain is from a CTP from *Oryza sativa* -Superoxide dismutase, *Oryza sativa* -Phospho-2-dehydro-3-deoxyheptonate Aldolase 2, *Oryza sativa* -L-Ascorbate peroxidase 5, *Zea Mays* ssRUBISCO or active variants thereof.

5

6. The recombinant polynucleotide of claim 2 or 3, wherein said C-terminal domain is from a CTP from *Oryza sativa* -soluble starch synthase, *Oryza sativa* -Superoxide dismutase, *Oryza sativa* -Phosphoglucan water dikinase, *Zea Mays* Thioredoxin M-type, *Zea Mays* -beta-glucosidase or active variants thereof.

10

7. The recombinant polynucleotide of claim 2 or 3, wherein said N-terminal domain is from the *Oryza sativa* 1-deoxy-D xyulose-5-Phosphate Synthase CTP or an active variant thereof, said central domain is from the *Zea Mays* ssRUBISCO CTP or an active variant thereof and said C-terminal domain is from the *Zea Mays* -beta-glucosidase CTP or an active variant thereof.

15

8. The recombinant polynucleotide of claim 2 or 3, wherein said N-terminal domain is from the *Zea Mays* -Malate dehydrogenase CTP or an active variant thereof, said central domain is from the *Oryza sativa* -Superoxide dismutase CTP or an active variant thereof and said C-terminal domain is from the *Oryza sativa* -soluble starch synthase CTP or an active variant thereof.

20

9. The recombinant polynucleotide of claim 2 or 3, wherein said N-terminal domain is from the *Oryza sativa* -NADP-dependent Malic acid enzyme CTP or an active variant thereof, said central domain is from the *Oryza sativa* -Phospho-2-dehydro-3-deoxyheptonate Aldolase 2 CTP or an active variant thereof and said C-terminal domain is from the *Zea Mays* Thioredoxin M-type CTP or an active variant thereof.

25

10. The recombinant polynucleotide of claim 2, wherein at least one of said N-terminal domain, said central domain, or said C-terminal domain comprises a chimeric domain.

30

11. The recombinant polynucleotide of claim 10, wherein at least one portion of said chimeric N-terminal domain is from the N-terminal domain of the *Oryza sativa* -NADP-dependent Malic CTP, *Zea Mays* -Malate dehydrogenase CTP or active variants thereof.

5

12. The recombinant polynucleotide of claim 10, wherein at least one portion of said chimeric central domain is from the central domain of the *Oryza sativa* -L-Ascorbate peroxidase 5 CTP, *Zea Mays* ssRUBISCO CTP or active variants thereof.

10

13. The recombinant polynucleotide of claim 10, wherein at least one portion of said chimeric C-terminal domain is from the C-terminal domain of the *Oryza sativa* -soluble starch synthase CTP, *Zea Mays* Thioredoxin M-type CTP, *Oryza sativa* -Superoxide dismutase CTP, *Oryza sativa* -Phosphoglucan water dikinase CTP or active variants thereof.

15

14. The recombinant polynucleotide of claim 10, wherein said chimeric CTP comprises

a) a chimeric N-terminal domain, wherein said chimeric N-terminal domain comprises a portion of the N-terminal domain from the *Zea Mays* -Malate dehydrogenase CTP fused in frame to a portion of the N-terminal domain of the *Oryza sativa* -NADP-dependent Malic acid enzyme CTP;

20

b) a central domain, wherein said central domain is from the *Zea Mays* ssRUBISCO CTP; and,

25

c) a chimeric C-terminal domain, wherein said chimeric C-terminal domain comprises a portion of the C-terminal domain from the *Oryza sativa* -soluble starch synthase CTP fused in frame to a portion of the C-terminal domain from the *Zea Mays* Thioredoxin M-type CTP;

wherein said chimeric CTP has CTP activity.

30

15. The recombinant polynucleotide of claim 10, wherein said chimeric CTP comprises

a) a chimeric N-terminal domain, wherein said chimeric N-terminal domain comprises a portion of the N-terminal domain from the *Zea Mays* -Malate

dehydrogenase CTP fused in frame to a portion of the N-terminal domain of the *Oryza sativa* -NADP-dependent Malic acid enzyme CTP;

b) a chimeric central domain, wherein said chimeric central domain comprises a portion of the central domain from the *Oryza sativa* -L-Ascorbate peroxidase 5 CTP fused in frame to a portion of the central domain of the *Zea Mays* ssRUBISCO CTP; and,

c) a chimeric C-terminal domain, wherein said chimeric C-terminal domain comprises a portion of the C-terminal domain from the *Oryza sativa* -Superoxide dismutase CTP fused in frame to a portion of the C-terminal domain of the *Oryza sativa* -Phosphoglucan water dikinase CTP;

wherein said chimeric CTP has CTP activity.

16. The recombinant polynucleotide of claim 3, wherein the chimeric CTP comprises

a) an amino acid sequence comprising the amino acids of SEQ ID NOS: 1, 2 or 3;

b) an amino acid sequence having at least 85% sequence identity to SEQ ID NOS: 1, 2 or 3, wherein said amino acid sequence has CTP activity; or

c) an amino acid sequence having at least 17 consecutive amino acids of SEQ ID NOS: 1, 2 or 3, wherein said amino acid sequence has CTP activity.

17. The recombinant polynucleotide of claim 14, wherein the chimeric CTP comprises

a) an amino acid sequence comprising the amino acids of SEQ ID NO: 4;

b) an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 4, wherein said amino acid sequence has CTP activity; or

c) an amino acid sequence having at least 17 consecutive amino acids of SEQ ID NO: 4, wherein said amino acid sequence has CTP activity.

18. The recombinant polynucleotide of claim 15, wherein the chimeric CTP comprises

a) an amino acid sequence comprising the amino acids of SEQ ID NO: 5;

- b) an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 5, wherein said amino acid sequence has CTP activity; or
- c) an amino acid sequence having at least 17 consecutive amino acids of SEQ ID NO: 5, wherein said amino acid sequence has CTP activity.

5

19. The recombinant polynucleotide of any one of claims 1-18, wherein said polypeptide of interest comprises a *Bacillus thuringiensis* polypeptide having insecticidal activity.

10

20. The recombinant polynucleotide of claim 19, wherein said *Bacillus thuringiensis* polypeptide having insecticidal activity comprises an IP2-127 polypeptide.

15

21. A nucleic acid construct comprising the recombinant polynucleotide of any one of claims 1-20.

22. The nucleic acid construct of claim 21, further comprising a promoter operably linked to said recombinant polynucleotide.

20

23. A cell comprising at least one recombinant polynucleotide of any of claims 1-20 or the nucleic acid construct of any one of claims 21 or 22.

24. The cell of claim 23, wherein said cell is a plant cell.

25

25. The cell of claim 24, wherein said polynucleotide or nucleic acid construct is stably incorporated into the genome of said plant cell.

26. The cell of any one of claims 24 or 25, wherein said plant cell is from a monocot.

30

27. The cell of claim 26, wherein said monocot is maize, wheat, rice, barley, sorghum, sugarcane or rye.

28. The cell of any one of claims 24 or 25, wherein said plant cell is from a dicot.

29. The cell of claim 28, wherein the dicot is soybean, Brassica, sunflower,
5 cotton or alfalfa.

30. A plant comprising at least one plant cell of any one of claims 24-29.

31. A plant explant comprising at least one plant cell of any one of claims
10 24-29.

32. A transgenic seed produced by the plant of claim 30, wherein said seed comprises said recombinant polynucleotide.

15 33. A recombinant polypeptide encoded by the polynucleotide of any one of claims 1-20.

34. A method of targeting a polypeptide of interest to a chloroplast comprising expressing the recombinant polynucleotide of any one of claims 1-20 or
20 the nucleic acid construct of claim 21 or 22 in a plant cell.

35. A method of targeting a polypeptide of interest to a chloroplast comprising introducing the recombinant polynucleotide of any one of claims 1-20 or the nucleic acid construct of claim 21 or 22 in a plant cell and expressing said
25 recombinant polynucleotide in the plant cell.

36. The method of claim 34 or 35, wherein said method further comprises regenerating a transgenic plant from said plant cell.

30 37. The method of any one of claims 34-36, wherein said plant cell is from a monocot.

38. The method of claim 37, wherein said monocot is selected from the group consisting of maize, wheat, rice, barley, sorghum, sugarcane or rye.

39. The method of any one of claims 34-36, wherein said plant cell is from a dicot.

5 40. The method of claim 39, wherein said dicot is selected from the group consisting of soybean, Brassica, sunflower, cotton or alfalfa.

41. The method of any one of claims 35-40, wherein said polypeptide of interest comprises an insecticidal protein and expression of said polypeptide controls
10 a pest.

42. The method of claim 41, wherein said polypeptide of interest comprises a *Bacillus thuringiensis* polypeptide having insecticidal activity.

15 43. The method of claim 42, wherein said *Bacillus thuringiensis* polypeptide having insecticidal activity comprises an IP2-127 polypeptide.

Figure 1
Strategy for CTP Development

CTP framework	Uncharged N-terminus		R rich C terminus Amphipathic helix
msCTP1	OsdXP	ssRUBISCO (Zm)	β Gluc
msCTP2	NADP	OsSOD	Os soluble starch
msCTP3	OsNADP-malic enzyme	OsPHD Aldolase 2	TRX
msCTP4	NADP/OsNADP-malic enzyme	ssRUBISCO (Zm)	Os sol. starch/TRX
msCTP5	NADP/OsNADP	OsAPx05/ ssRUBISCO	OsSOD/OsPGDK
msCTP6	Most freq AA	Most freq AA	Most freq AA
msCTP7	Most freq AA	Most freq AA	Most freq AA

Alignment of Monocot CTPs

CTP framework	
Uncharged N-terminus	S, T, K, R Lack of Acidic residues
R rich C-terminus Amphiphilic B-strand	

Figure 1

Strategy for CTP Development ^{AA}

CTP framework	Uncharged N-terminus		R rich C terminus Amphipathic helix	BB
msCTP1	OsdXP	ssRUBISCO (Zm)	β Gluc	
msCTP2	NADP	OsSOD	Os soluble starch	CC
msCTP3	OsNADP-malic enzyme	OsPHD Aldolase 2	TRX	
msCTP4	NADP/OsNADP- malic enzyme	ssRUBISCO (Zm)	Os sol. starch/TRX	
msCTP5	NADP/OsNADP	OsAPx05/ ssRUBISCO	OsSOD/OsPGDK	
msCTP6	Most freq AA	Most freq AA	Most freq AA	
msCTP7	Most freq AA	Most freq AA	Most freq AA	