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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 March 2021 (18.03.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/050968 A1

(51) International Patent Classification:
C12N 15/74 (2006.01) *C12N 15/52* (2006.01)
C12N 15/66 (2006.01)

(21) International Application Number:
PCT/US2020/050528

(22) International Filing Date:
11 September 2020 (11.09.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/898,891 11 September 2019 (11.09.2019) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, California 94607-5200 (US).

(72) Inventors: **MELIS, Anastasios**; 2745 Del Monte Avenue, El Cerrito, California 94530 (US). **BETTERLE, Nico**; 3950 Stoneridge Drive #2, Pleasanton, California 94588 (US). **HIDALGO MARTINEZ, Diego**; 7347 Hotchkiss Avenue, El Cerrito, California 94530 (US). **SAPHIRE, Andrew C.**; #21 675 South Sierra Avenue, Solana Beach, California 92075 (US).

(74) Agent: **LOCKYER, Jean M.** et al.; Kilpatrick Townsend & Stockton LLP, Mailstop: IP Docketing - 22 1100 Peachtree Street, Suite 2800, Atlanta, GA 30309 (US).

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

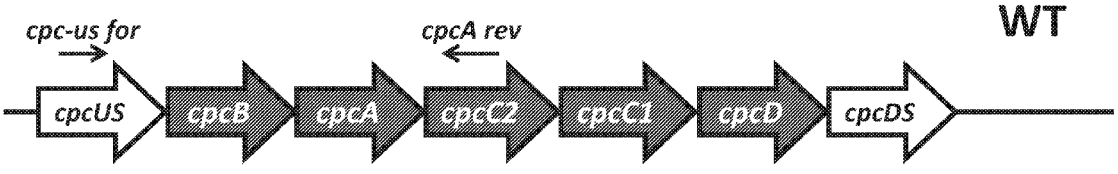
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR,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

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*
— *of inventorship (Rule 4.17(iv))*

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

(54) Title: FUSION CONSTRUCTS TO EXPRESS BIOPHARMACEUTICAL POLYPEPTIDES IN CYANOBACTERIA

Fig. 1 A



(57) Abstract: This invention provides compositions and methods for providing high product yield of transgenes encoding biopharmaceutical polypeptides in cyanobacteria and microalgae.

WO 2021/050968 A1

FUSION CONSTRUCTS TO EXPRESS BIOPHARMACEUTICAL POLYPEPTIDES IN CYANOBACTERIA

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority benefit of U.S. Provisional Application No. 62/898,891, filed September 11, 2019, which is incorporated by reference in its entirety for all purposes.

BACKGROUND OF THE INVENTION

[0002] Efforts to express human therapeutic proteins in photosynthetic microorganisms abound in the literature. In their preponderance, these entail heterologous transformation of a microalgal chloroplast as a synthetic biology platform for the production of biopharmaceutical and therapeutic proteins (Dyo and Purton 2018, and references therein). The vast majority of such efforts have employed transformation of the chloroplast in the model green microalga *Chlamydomonas reinhardtii* via double homologous recombination of exogenous constructs encoding heterologous proteins (Surzycki et al. 2009; Tran et al. 2009; Coragliotti et al. 2011; Gregory et al. 2013; Jones and Mayfield 2013; Rasala and Mayfield 2015; Baier et al. 2018). A common feature of these efforts is the low yield of the transgenic biopharmaceutical proteins, not exceeding 1% of the total *Chlamydomonas reinhardtii* protein (Dyo and Purton 2018). In general, there is a need to develop methods that will systematically and reliably over-express eukaryotic, including human therapeutic, proteins in photosynthetic microorganisms, at levels that exceed 1% of the total cell protein.

[0003] Cyanobacteria such as *Synechocystis* and other microalgae can be used as photosynthetic platforms for the heterologous generation of products of interest. For example, bacterial proteins can be heterologously over-expressed in cyanobacteria, reportedly up to 20% of total soluble protein, by using the strong *cpc* operon promoter and possibly other endogenous or exogenous promoters (Zhou et al. 2014, Kirst et al. 2014; Formighieri and Melis 2017). By way of illustration, Zhou et al. (2014), described the function of a modified (partial) endogenous cyanobacterial promoter (*Pcpc560*), derived from the native cyanobacterial *cpc* operon promoter. They examined the efficacy of this promoter to express (i) the trans-enoyl-CoA reductase (Ter) protein from T*reponema denticola*, a Gram-negative,

obligate anaerobic bacterium, and (ii) the D-lactate dehydrogenase (DldhE) protein from *Escherichia coli*. Both of these bacterial-origin genes and proteins were readily overexpressed in cyanobacteria under the control of the Pcpc. Kirst et al. (2014) showed that *Synechocystis* readily overexpressed, at the protein level and under the native *Pcpc*, the *nptI* gene from *Escherichia coli*, encoding the neomycin phosphotransferase, a kanamycin resistance conferring protein. Similarly, Xiong et al. (2015) showed overexpression of the *Pseudomonas syringae efe* gene, encoding an ethylene forming enzyme, in *Synechocystis sp.* PCC 6803 and enhanced EFE protein accumulation upon transformation of *Synechocystis* with multiple copies of the *P. syringae efe* gene (Xiong et al. 2015). Likewise, Chaves and co-workers provided evidence that cyanobacteria will over-express, at the protein level, the *cmR* gene from *Escherichia coli*, encoding a chloramphenicol resistance protein (Chaves et al. 2016), and the isopentenyl diphosphate isomerase (*fni*) gene from S*treptococcus pneumoniae,* either under the native *Pcpc* (Chaves et al. 2016) or heterologous *Ptrc* promoter (Chaves et al. 2018).

[0004] In separate work, Desplancq et al. (2005) showed that transgenic *Anabaena* sp. PCC 7120, a filamentous cyanobacterium, was able to express the *Escherichia coli*, e.g. bacterial origin, maltose-binding protein (MBP), yielding up to 250 mg MBP per L of culture. In further work, Desplancq et al. (2008) showed that *Anabaena* was also able to express 100 mg per L of gyrase B (GyrB), a 23 kD *Escherichia coli* protein. This is consistent with the notion that cyanobacteria easily express other "bacterial" origin proteins.

[0005] However, recent experience has also shown that heterologous expression of eukaryotic plant and yeast genes occurs at low protein levels, regardless of the promoter used and mRNA levels achieved in the cyanobacterial cytosol (Formighieri and Melis 2016). For example, plant terpene synthases could not be expressed well in cyanobacteria under the control of different strong endogenous and heterologous promoters (Formighieri and Melis 2014; Englund et al. 2018). Heterologous expression in cyanobacteria of the isoprene synthase (Lindberg et al. 2010; Bentley and Melis 2012), β-phellandrene synthase (Bentley et al. 2013), geranyl diphosphate (GPP) synthase from a higher plant origin (Bentley et al 2014; Formighieri et al 2017; Betterle and Melis 2018), and the alcohol dehydrogenase (*ADH1*) gene from yeast (Chen et al. 2013), all showed low levels of recombinant protein expression, even under the control of strong endogenous (e.g. *psbA2, rbcL, cpc*) or strong heterologous promoters (e.g. *Ptrc*), and even after following a careful codon-use optimization of the target transgene (Lindberg et al. 2010; Bentley and Melis 2012; Ungerer et al. 2012; Bentley et al.

2013; Chen and Melis 2013; Formighieri and Melis 2014a; Englund et al. 2018). Similarly, only low levels of expression were reported for a chimeric complex of plant enzymes, including the ethylene synthase *efe* gene from *Solanum lycopersicum* (tomato) (Jindou et al. 2014; Xue et al. 2014), limonene synthase from *Mentha spicata* (spearmint) (Davies et al. 2014) and *Picea sitchensis* (Sitka spruce) (Halfmann et al. 2014a), the sesquiterpene farnesene and bisabolene synthases from *Picea abies* (Norway spruce) (Halfmann et al. 2014b) and *Abies grandis* (grand fir) (Davies et al 2014). In these and other studies, transgenic protein levels were not evident on an SDS-PAGE Coomassie stain of protein extracts and, frequently, shown by sensitive Western blot analysis only.

[0006] Animal-origin eukaryotic transgenes are difficult to express in cyanobacteria. Desplancq et al. (2008) showed that the eukaryotic (human) oncogene E6 protein, when expressed in cyanobacteria, is toxic to the cells. To manage the toxicity, they separated in time cell growth from recombinant protein expression. They resorted to using the inducible nitrate assimilation nir promoter of the filamentous cyanobacterium *Anabaena*, as the promoter for the expression of their transgenes. The latter is repressed in the presence of ammonium (NH4+) salts but induced in the absence of ammonium and presence of nitrate (NO3-). They grew Anabaena to high cell density in the presence of ammonium (NH4+), thereby blocking the expression of the transgenes. By the time cells reached a high density in the culture, the pre-calculated amount of ammonium was either consumed, or experimentally replaced with nitrate salts. Cells then activated the nitrate reductase *nir* promoter, as they were forced to rely on nitrate nutrients for further growth. This induction process resulted in the accumulation of small amounts of the transgenic eukaryotic (human) oncogene E6 protein, although this product again proved to be lethal to the cells under these conditions. Since efforts to express the oncogene E6 by itself failed due to toxicity of the product, Desplancq et al. (2008) undertook to express it as a fusion-protein with the highly-expressed maltose-binding protein as the leader sequence in a MBP*E6 fusion. This effort resulted in a yield of 1 mg protein per L after 5 days of *nir* induction, i.e., 0.4% of the amount measured with MBP as the solo recombinant protein. They suggested that the MBP*E6 fusion protein has an inhibitory effect on its own expression and further that this oncoprotein is toxic to *Anabaena* cells, evidenced from the about 50% inhibition in cell growth observed in the MBP*E6 expressing transformants.

[0007] Interferons (IFNs) are a group of signaling proteins made and released by host cells in response to the presence of viruses. Interferon a-2a (IFNA2) is a member of the Type I

interferon cytokine family, known for its antiviral and anti-proliferative functions. Recombinant *Escherichia coli* expression of IFNA2 resulted in inclusion body formation, or required numerous purification steps that decreased the protein yield. Bis et al. (2014) described an expression and purification scheme for IFNA2 using a pET-SUMO bacterial expression system and a single purification step. Using the SUMO protein as the fusion tag increased the soluble protein expression and minimized the amount of inclusion bodies in *E. coli*. Following protein expression, the SUMO tag was cleaved with the Ulp1 protease leaving no additional amino acids on the fusion terminus following cleavage (Bis et al. 2014). The purified protein had antiviral and anti-proliferative activities comparable to the WHO International Standard, NIBSC 95/650, and the IFNA2 standard available from PBL Assay Science.

[0008] Tissue-type plasminogen activator (tPA) is a protein involved in the breakdown of blood clots. Human tPA has a molecular weight of ~70 kD in the single-chain form. tPA has five peptide domains: The N-terminal finger, epidermal growth factor, serine protease, Kringle 1, and Kringle 2 domain (Youchun et al. 2003). The active part of tPA, the thrombolytic Kringle 2 domain, serine protease domain, two functional regions of protease (176-527 amino acid residues), plus the 1 to 3 amino acids of the N-terminal is known as the "truncated human tissue plasminogen activator" (K2S, reteplase), which has a longer plasma half-life and higher fibrinolytic activity than tPA (Nordt and Bode 2003; Hidalgo et al. 2017). tPA can be manufactured using recombinant DNA technologies based on transgenic microorganism cultures such as *Escherichia coli* and *Saccharomyces cerevisiae* in fermentative bioreactors (Demain and Vaishnav 2009). The biotechnological production of recombinant tissue plasminogen activator protein (K2S, reteplase) from transplastomic tobacco cell cultures was also reported (Hidalgo et al. 2017).

[0009] Recombinant insulin protein is used as a treatment of diabetic patients. The recombinant protein is produced predominantly in *Escherichia coli* and *Saccharomyces cerevisiae*.

[0010] There is a need to develop additional recombinant DNA technologies for the generation of low-cost biopharmaceutical proteins, without relying on animal systems, and without causing depletion of natural resources, pollution, or other environmental degradation. In this respect, a direct photosynthetic production of such compounds is promising. Recently, fusion constructs were designed as protein overexpression vectors that could be used in

cyanobacteria for the over-expression of recalcitrant genes, *e.g.*, plant terpene synthases (WO2016210154). In this approach, highly-expressed endogenous cyanobacteria genes, such as the *cpcB* gene, encoding the β-subunit of phycocyanin, or highly-expressed heterologous genes, such as the *nptI* gene, encoding the kanamycin resistance protein, were employed as leader sequences in such fusion constructs, resulting in the accumulation of eukaryotic proteins up to ~20% of the total cyanobacterial protein.

BRIEF SUMMARY OF SOME ASPECTS OF THE INVENTION

[0011] The present invention is based, in part on the discovery of fusion protein constructs that can be used in cyanobacteria as transgenic protein over-expression vectors to provide high levels of transgenic animal protein accumulation and thus provide high rates of production of biopharmaceutical products such as insulin, interferons, or tissue plasminogen activator (tPA), or tPA derivatives, *e.g.*, an active truncated form of tPA.

[0012] In one aspect, provided herewith is an expression construct comprising a nucleic acid sequence comprising a transgene that encodes a biopharmaceutical protein, wherein the transgene is fused to the 3' end of a nucleic acid sequence that encodes a cyanobacteria protein that is expressed in cyanobacteria at a level of at least 1% of the total cellular protein or encodes an exogenous protein that is over-expressed in cyanobacteria at a level of at least 1% of the total cellular protein. In some embodiments, the transgene is fused to the 3' end of a nucleic acid sequence that encodes a cyanobacteria protein that is expressed in cyanobacteria at a level of at least 1% of the total cellular protein. In some embodiments, the cyanobacteria protein is a β-subunit of phycocyanin (cpcB), an α-subunit of phycocyanin (cpcA), a phycoerythrin subunit (cpeA or cpeB), an allophycocyanin subunit (apcA or apcB), a large subunit of Rubisco (rbcL), a small subunit of Rubisco (rbcS), a D1/32 kD reaction center protein (psbA) of photosystem-II, a D2/34 kD reaction center protein (psbD) of photosystem-II, a CP47 (psbB) or CP43 (psbC) reaction center protein of photosystem-II, a psaA or psaB reaction center protein of photosystem-I, a psaC or psaD reaction center protein of photosystem-I, an rpl ribosomal RNA protein, or an rps ribosomal RNA protein. In some emboidments, the transgene encode insulin, *e.g.*, human insulin. In some embodiments the transgene encode an interferon, *e.g.*, *a* human interferon alpha, such as IFNA2. In some embodiments, the transgene encodes a human tissue plasminogen activator, for example, a truncated human tissue plasminogen activator (K2S, reteplase), which includes the Kringle 2 domain and the serine protease domain. In some embodiments, the transgene encodes a

SARS-CoV2 receptor binding domain. In other embodiments, the transgene encodes a a Tetanus Toxin Fragment C polypeptide. In some embodiments, the transgene is fused to the 3' end of a nucleic acid sequence that encodes an exogenous protein that is over-expressed in cyanobacteria at a level of at least 1% of the total cellular protein. For example, the exogenous protein may be an antibiotic resistance protein such as kanamycin, chloramphenicol, streptomycin, erythromycin, zeocin, or spectinomycin. n some emboidments, the transgene encode insulin, *e.g.*, human insulin. In some embodiments the transgene encode an interferon, *e.g.*, *a* human interferon alpha, such as IFNA2. In some embodiments, the transgene encodes a human tissue plasminogen activator, for example, a truncated human tissue plasminogen activator (K2S, reteplase), which includes the Kringle 2 domain and the serine protease domain. In some embodiments, the transgene encodes a SARS-CoV2 receptor binding domain. In other embodiments, the transgene encodes a a Tetanus Toxin Fragment C polypeptide.

[0013] In another aspect, the disclosure provide a host cell comprising an expression construct as described herein, *e.g.*, in the preceiding paragraph. In some embodiments, the host cell is a cyanobacteria host cell, such as a single celled cyanobacteria, *e.g.*, *a Synechococcus* sp., *a Thermosynechococcus elongatus, a Synechocystis* sp., or a *Cyanothece* sp. In some embodiments, the cyanobacteria are micro-colonial cyanobacteria, *e.g.*, a *Gloeocapsa magma, Gloeocapsa phylum, Gloeocapsa alpicola, Gloeocpasa atrata, Chroococcus* spp., or *Aphanothece* sp. In some emboidments, the cyanobacteria is a filamentous cyanobacteria, such as an *Oscillatoria* spp., a *Nostoc* sp., an *Anabaena* sp., or an *Arthrospira* sp.

[0014] In further aspects, provided a cyanobacterial cell culture comprising cyanobacteria genetically modified as described herein to produce a biopharmaceutical protein, *e.g.*, as described in the preceding paragraph. In some embodiments, the disclosure provide a photobioreactor containing such a cyanobacterial cell culture.

[0015] In an addition as expect, the disclosure provides a method of expressing a transgene at high levels, the method comprising culturing a cyanobacterial cell culture as described herein, *e.g.*, in the preceding paragraph under conditions in which the transgene is expressed.

[0016] In a further aspect provided herein is a method of modifying a cyanobacterial cell to express a transgene at high levels, the method comprising introducing an expression construct as described herein, *e.g.*, in the preceding paragraphs, into the cell.

[0017] In other aspective provided herein is an isolated fusion protein comprising a biopharmaceutical protein to be expressed in cyanobacteria fused to the 3' end of a heterologous protein that is expressed in cyanobacteria at a level of at least 1% of the total cellular protein. In some embodiments, the heterologous protein is a native cyanobacteria protein.

[0018] In a further aspect, provided herein is a cyanobacterial host cell comprising an expression unit comprising: (i) a nucleic acid sequence comprising a transgene that encodes a biopharmaceutical protein, wherein the transgene is fused to the 3' end of a nucleic acid sequence that encodes a cyanobacteria β-subunit of phycocyanin (cpcB) polypeptide to produce a fusion polypeptide comprises cpcB and the biopharmaceutical protein; (ii) a nucleic acid sequence encoding a cyanobacteria α-subunit of phycocyanin (cpcA) polypeptide; and (iii) a nucleic acid sequence encoding a cyanobacterial cpcC1, cpcC2 and cpcD polypeptide. In some embodiments, the recombinant expression unit is operably linked to an endogenous cyanobacteria cpc promoter. In some embodiments, the transgene encodes a native human interferon polypeptide. In some embodiments, the transgene encodes an interferon polypeptide having at least 95% identity to SEQ ID NO:1. In some embodiments, the fusion protein comprises a protease cleavage site, *e.g.*, a Factor Xa cleavage site, between cpcB and the interferon polypeptide. In some embodiments, the transgene encodes a native human tissue plasminogen activator (tPA) polypeptide or truncated native human tPA polypeptide. In some embodiments, the transgene encodes a tPA polypeptide having at least 95% identity to the region of SEQ ID NO:2 that lacks the signal peptide or having at least 95% identity to SEQ ID NO:3. In some embodiments, the fusion protein comprises a protease cleavage site, *e.g.*, a Factor Xa cleavage site, between cpcB and the tPA polypeptide. In some embodiments, the transgene encodes a native Tetanus Toxin Fragment C (TTFC) polypeptide or a TTFC polypeptide having at least 95% identity to SEQ ID NO:15. In some embodiments, the fusion protein comprises a protease cleavage site, *e.g.*, a Tobacco Etch virus (TEV) cleavage site, between cpcB and the TTFC polypeptide. In some embodiments, the transgene encodes a native Cholera Toxin Fragment B polypeptide or a Cholera Toxin Fragment B polypeptide having at least 95% identity to SEQ ID NO:18. In some embodiments, the fusion protein comprises a protease cleavage site, *e.g.*, a TEV cleavae site, between cpcB and the Cholera Toxin Fragment B polypeptide. In some embodiments, the transgene encodes a native human insulin polypeptide. In some embodiments, the transgene encodes an insulin polypeptide having at least 95% identity to SEQ ID NO:4. In some

emobdiments, the fusion protein comprises a protease cleavage site, *e.g.*, a Factor Xa cleavage site, between the cpcB and insulin polypeptide. In some embodiments,the transgene encodes a SARS-CoV2 polypeptide having at least 95% identity to SEQ ID NO:16 or 17. In some embodiments,the fusion protein comprises a protease cleavage site between cpcB and the SARS-CoV2 polypeptide. In some embodiments, an expression unit as provided herein comprises an antibiotic resistance gene, *e.g.*, a chloramphenicol or streptomycin antibiotic resistance gene, between the transgene and cpcA.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] FIG. 1. Schematic overview of DNA constructs designed for the transformation of the *Synechocystis* PCC 6803 (*Synechocystis*) genome. (A) The native *cpc* operon, as it occurs in wild type S*ynechocystis*. This DNA sequence and associated strain are referred to as the wild type (WT). (B) Insertion in the *cpc* operon of the codon-optimized human interferon (*IFN*) gene followed by the chloramphenicol (*cmR)* resistance cassette in an operon configuration, replacing the phycocyanin-encoding β-subunit *cpcB* gene of *Synechocystis*. This DNA construct is referred to as *IFN*; (C) Insertion in the *cpc* operon of the codon-optimized *IFN* gene immediately downstream of the phycocyanin-encoding β-subunit *cpcB* gene of *Synechocystis*, followed by the *cmR* resistance cassette, in an operon configuration. This DNA construct is referred to as *cpcB-IFN*; (D) Insertion in the *cpc* operon of the codon-optimized *IFN* gene as a fusion construct with the phycocyanin-encoding β-subunit *cpcB* gene, with the latter in the leader sequence position. The fusion construct *cpcB*IFN* was followed by the *cmR* resistance cassette in an operon configuration. *cpcB* and *IFN* genes were linked by the DNA sequence encoding the Factor Xa cleavage site. The latter comprises the Ile-Glu/Asp-Gly-Arg amino acid sequence. This DNA construct is referred to as the *cpcB*IFN*.

[0020] FIG. 2. Genomic DNA PCR analysis testing for transgenic DNA copy homoplasmy in *Synechocystis* transformants. Wild type and transformant strains were probed in genomic DNA PCR reactions for product generation and transgenic DNA segregation. Primers <*cpc-us for*> and <*cpcA rev*> showed substantially different and unique products in the wild type and the different transformants comprising the constructs of Fig. 1. Wild type PCR products had a 1,289 bp size, whereas the IFN, cpcB-IFN, and the cpcB*IFN transformants generated 2,094, 2,723, and 2,619 bp size products, respectively. Absence of wild type products from the latter was evidence of DNA copy homoplasmy for the transformants. (The cpcB-IFN

construct generated a product size slightly larger than that of the cpcB*IFN because it contained the *Synechocystis* native cpcB-cpcA intergenic DNA sequence. Please see gene nucleotide sequences in the Supplementary Materials.)

[0021] FIG. 3. Coloration of cells from photoautotrophically-grown liquid cultures showing a blue-green wild type (WT) phenotype, and greenish phenotype for the IFN, CpcB-IFN, and CpcB*IFN-containing transformants. The latter did not assemble phycocyanin rods, hence the absence of the distinct blue cyanobacterial coloration from the cells.

[0022] FIG. 4. Protein expression analysis of *Synechocystis* wild type and transformants. (A) Total cellular protein extracts were resolved by SDS-PAGE and visualized by Coomassie-stain. Two independent replicates of total protein extracts from wild type (WT), and IFN, CpcB-IFN and CpcB*IFN transformant cells were loaded onto the SDS-PAGE. Individual native and heterologous proteins of interest are indicated to the right of the gel. Sample loading corresponds to 0.25 µg of chlorophyll. Note the clear presence of a heterologous protein migrating to ~36 kD in the CpcB*IFN fusion extracts. (B) Total protein extracts of (A) were subjected to Western-blot analysis with loading of the lanes as per Fig. 4A. Specific polyclonal antibodies against the human IFN protein were used to probe target proteins. Sample loading corresponds to 0.25 µg of chlorophyll. Note the specific antibody cross-reaction with proteins migrating to ~36 and ~108 kD in the cpcB*IFN fusion and the absence of a cross reaction with any protein from the IFN and cpcB-IFN transformant cells. The latter do not seem to make / accumulate IFN.

[0023] FIG. 5. Protein expression analysis of *Synechocystis* wild type (WT) and transformants harboring the *cpcB*IFN* fusion construct. Total cellular protein extracts were resolved by SDS-PAGE and visualized by Coomassie-stain. Two different versions of the IFN gene were used: the human native IFN' and the *Synechocystis* codon-optimized IFN gene. Note the presence of heterologous proteins migrating to ~36 kD (CpcB*IFN) and ~23 kD (CmR) in the transformants but not in the wild type. Also note the presence of the ~19 kD CpcB β-subunit and the ~17 kD CpcA α-subunit of phycocyanin in the wild type but not in the transformants. Sample loading corresponds to 0.5 µg of chlorophyll. Quantification of the CpcB*IFN protein accumulation relative to that of the Rubisco large subunit (RbcL) is given in the results of Table 1.

[0024] FIG. 6. Protein expression analysis of *Synechocystis* wild type (WT) and transformants harboring the *cpcB*IFN* fusion construct. Total cellular protein extracts were

resolved by SDS-PAGE and visualized by Coomassie-stain. Two different versions of the fusion construct were used comprising the *CpcB*IFN* fusion and the more extensive *cpcB*His*Xa*IFN* fusion configuration, followed by the *cmR* resistance cassette. Equivalent amount of the CpcB*IFN and the CpcB*His*Xa*IFN fusion proteins were expressed in *Synechocystis*. Individual native and heterologous proteins of interest are indicated to the right of the gel. Sample loading corresponds to 0.25 μg of chlorophyll.

[0025] FIG. 7. Batch-scale purification of the recombinant CpcB*His*Xa*IFN protein through cobalt affinity chromatography. Protein purification was conducted employing a small amount of resin as solid phase. The latter was mixed and incubated with the cell extracts. The resin was pelleted and washed repeatedly with buffers containing imidazole at different concentrations.

Lane 1 shows the cell extracts (upper panel) and the resin pellet (lower panel) of the wild type, CpcB*IFN, and CpcB*His*Xa*IFN fusion construct cells prior to incubation with the resin. Note the natural pink coloration of the latter.

Lane 2 shows the cell extracts (upper panel) and the resin pellet (lower panel) of the wild type, CpcB*IFN, and CpcB*His*Xa*IFN fusion construct cells following a 5-min incubation with the resin in the presence of 10 mM imidazole. Note the blue coloration of the resin and the green coloration of the supernatant.

Lanes 3-5 show the remaining cell extracts (upper panel) and the resin pellet (lower panel) of the wild type, CpcB*IFN, and CpcB*His*Xa*IFN fusion construct cells following a consecutive wash of the resin three times with a buffer containing 10 mM of imidazole. Note the resulting clear supernatant and the pink coloration of the resin after the third wash (lane 5) for the wild type and CpcB*IFN, suggesting absence of His-tagged proteins. Also note the blue coloration of the resin in the CpcB*His*Xa*IFN sample, which was retained in this pellet (lanes 3-5) in spite of the repeated wash, suggesting the presence of resin-bound blue-colored His-tagged proteins.

Lanes 6-8 show the subsequent extracts (upper panel) and the resin pellet (lower panel) of the wild type, CpcB*IFN, and CpcB*His*Xa*IFN fusion construct cells following a wash three times with a buffer containing 250 mM of imidazole, designed to dissociate His-tagged proteins from the resin. Note the bluish supernatant in lanes 6 and 7 and the corresponding loss of the blue color from the resin pellet, suggesting the specific removal of His-tagged proteins from the resin.

[0026] **FIG. 8.** Coomassie-stained SDS-PAGE gel analysis of fractions eluted with different imidazole concentrations. Fractions were obtained upon affinity chromatography purification as shown in Fig. 7. Samples were loaded on a per volume basis. Note the ~108, ~38, and ~17 kD proteins eluted from the CpcB*His*Xa*IFN extract (marked with arrows).

[0027] **FIG. 9.** Absorbance spectra of purified *Synechocystis* complexes. (A) Absorbance spectra of eluent E1 fractions from wild type, CpcB*IFN, and CpcB*His*Xa*IFN samples, as shown in Fig. 8. (B) Absorbance spectra of cellular protein extracts from wild type, Δcpc deletion mutant (Kirst et al., 2014) and CpcB*His*Xa*IFN transformant cells.

[0028] **FIG. 10.** Column-based purification of the CpcB*His*Xa*IFN fusion protein through cobalt affinity chromatography.

Lane 1, upper panel, shows the CpcB*His*Xa*IFN cell extracts in the presence of 5 mM imidazole prior to resin application. Lane 1, lower panel, shows the SDS-PAGE protein profile of these extracts, indicating presence of all *Synechocystis* proteins.

Lane 2, upper panel, shows the CpcB*His*Xa*IFN cell extracts after incubation with the resin but prior to washing with additional imidazole applications. Lane 2, lower panel, shows the SDS-PAGE protein profile of these extracts, obtained upon a prior removal of the resin from the mix, indicating presence of all *Synechocystis* proteins.

Lanes 3-6, upper panel, show the CpcB*His*Xa*IFN cell extracts that passed through the resin upon four consecutive washes with 5 mM imidazole and, lower panel, the SDS-PAGE protein profile of these extracts, showing a steep depletion (from lane 3 to lane 6) of total protein.

Lanes 7-9, upper panel, show the further removal of resin-bound proteins from the CpcB*His*Xa*IFN cell extracts that eluted upon three consecutive washes with 250 mM imidazole and, lower panel, the SDS-PAGE protein profile of these extracts, showing substantial enrichment in mainly four proteins with apparent molecular weights of 108, 36, 27, and 17 kD. The majority of these proteins were eluted with the first application of the 250 mM imidazole solution.

[0029] **FIG. 11.** (A) SDS-PAGE and Coomassie-staining analysis of *Synechocystis* wild type, CpcB*IFN, and CpcB*His*Xa*IFN total cell extract, and of proteins eluted from the resin column upon application of 250 mM imidazole. (B) Western blot analysis with specific IFN polyclonal antibodies of the proteins resolved in (A). Note the heterologous ~36 kD

CpcB*His*Xa*IFN and the ~108 kD putative CpcB*His*Xa*IFN trimer (marked by arrowheads).

[0030] FIG. 12. (A) SDS-PAGE and Coomassie-stain analysis of *Synechocystis* wild type, CpcB*His*Xa*IFN, and resin-eluted proteins. (B) SDS-PAGE and Zinc-stain analysis of *Synechocystis* wild type, CpcB*His*Xa*IFN, and resin-eluted proteins. Zn-staining is designed to highlight the presence of bilin tetrapyrrole pigments. Individual native and heterologous proteins of interest are indicated to the right of the gels.

[0031] FIG. 13. (A) Map of the *nptI*IFN* fusion construct in the *cpc* operon locus. Note the presence of the His-tag and the Xa protease cleavage site in-between the two genes in the fusion. (B) SDS-PAGE and Coomassie staining of the protein extracts from wild type (WT), the *cpcB*His*Xa*IFN*, and two independent lines of the *nptI*His*Xa*IFN* transformants. (C) Western blot analysis of a duplicate gel as the one shown in (B). Specific anti-IFN polyclonal antibodies were used in this analysis. Note the specific antibody cross reactions with protein bands migrating to ~36 kD (CpcB*His*Xa*IFN) and ~46 kD (NptI*His*Xa*IFN). Also note the antibody cross reactions with protein bands of higher molecular mass.

[0032] FIG. 14. Efficacy of interferon in preventing encephalomyocarditis virus (EMC) infection of human lung cells (A549), as performed by a PBL Assay Science, Piscataway, NJ USA test. (Diamonds) IFN titration curve using a standard recombinant interferon. (Squares) IFN titration curve using the cyanobacterial CpcB*His*Xa*IFN fusion interferon. The analysis showed that 0.002 ng/mL of a standard recombinant interferon was needed to cause 50% inhibition in EMC infection, whereas 0.0875 ng/mL of cyanobacterial CpcB*His*Xa*IFN fusion interferon was required to cause 50% inhibition in EMC infection.

[0033] FIG. 15. (A) Map of the *cpcB*His*Xa*K2S* fusion construct in the *cpc* operon locus. Note the presence of the His-tag and the Xa protease cleavage site in-between the two genes in the fusion. (B) SDS-PAGE and Coomassie stain of the protein extracts from wild type (WT), and three independent lines of the *cpcB*His*Xa*K2S* transformant. (C) Western blot analysis of a duplicate gel as the one shown in (B). tissue-Plasminogen Activase recognizing polyclonal antibodies were used in this assay. Note the specific antibody cross reactions with protein bands migrating to ~58.9 kD protein band in the K2S transformants.

[0034] FIG. 16. (A) Map of the *cpcB*INS* fusion construct in the *cpc* operon locus. (B) SDS-PAGE and Coomassie stain of the protein extracts from wild type (WT), a CpcB*INS

(insulin) containing transformant and, for comparison purposes, a CpcB*PHLS (β-phellandrene synthase) transformant. Note the 19 kD β-subunit and 17 kD α-subunit of phycocyanin in the wild type, the ~27 kD CpcB*INS (insulin) in the *cpcB*INS* transformant, and the ~84 kD CpcB*PHLS protein in the *cpcB*PHLS* transformant.

[0035] FIG. 17. (A) Map of the *cpcB*L7*His*TEV*TTFC* fusion construct in the *cpc* operon locus, including a linker of seven aminoacids (L7) and a Hisx6-tag (His). (B, left panel) SDS-PAGE and Coomassie stain analysis of the protein extracts from wild type (WT), the LTV recipient strain, and three *Synechocystis* transformant lines of the cpcB*L7*His*TEV*TTFC (Tetanus Toxin Fragment C). Note the presence of the 19 kD CpcB β-subunit and 17 kD CpcA α-subunit of phycocyanin in the wild type only, the ~72 kD cpcB*L7*His*TEV*TTFC protein (denoted as cpcB*TTFC) in the TTFC transformants, and the ~55 kD RBCL (large sububit of Rubisco) protein in all strains. Hashtag (#) denotes the electrophoretic mobility position of the cpcB*L7*TEV*ISPS fusion protein from the respective isoprene synthase (*ISPS*)-containing strain that was used as the recipient strain of the *cpcB*L7*His*TEV*TTFC* construct. Densitometric analysis of the SDS-PAGE Coommassie stain showed that the cpcB*L7*His*TEV*TTFC fusion protein accounted for about 28% of the total cell protein. (B, right panel) Western blot analysis of the protein profile shown in B (left panel), probed with specific polyclonal antibodies agains the TTFC polypeptide. Note the antibody cross reaction with the 72 kD CpcB*L7*His*TEV*TTFC fusion protein, the ~290 kD putative trimeric [CpcB*L7*His*TEV*TTFC]x3 undissolved fusion protein complex, plus some lower molecular size putative proteolysis fragments.

[0036] FIG. 18. (A) Map of the *cpcB*L7*His*TEV*RBD* fusion construct in the *cpc* operon locus, including a linker of seven amino acids (L7), a Hisx6-tag (His) and the TEV cleavage site (TEV), followed by the Receptor Binding Domain (RBD) of the S1 protein from the SARS-CoV-2. (B, left panel) SDS-PAGE and Coomassie stain of the protein extracts from wild type (WT), the LTV recipient strain, and a *Synechocystis* transformant line harboring the *cpcB*L7*His*TEV*RBD* fusion protein (RBD). The arrow points to the electrophoretic mobility of the 45 kD RBD fusion protein. (B, center panel). Western blot analysis of the protein profile shown in B (left panel), probed with specific polyclonal antibodies agains the leader CpcB protein in the fusion construct. Note the antibody cross reaction with the 45 kD cpcB*L7*His*TEV*RBD fusion protein. (B, right panel) SDS-PAGE and Zinc-stain analysis of *Synechocystis* expressing the LTV and RBD fusion construct phenotypes. Zn-staining is designed to highlight the presence of bilin tetrapyrrole

pigments. Note the Zn-staining of a band at 45 kD in the RBD expressing transformant, and the staining of a band migrating to ~85 kD in the LTV (*cpcB*L7*REV*ISPS*) transformant.

[0037] **FIG. 19.** Panels A-D provide schematics of illustrative expression constructs.

DETAILED DESCRIPTION OF THE INVENTION

[0038] The term "naturally-occurring" or "native" as used herein as applied to a nucleic acid, a protein, a cell, or an organism, refers to a nucleic acid, protein, cell, or organism that is found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism that can be isolated from a source in nature and which has not been intentionally modified by a human in the laboratory is naturally occurring.

[0039] The term "heterologous nucleic acid," as used herein, refers to a nucleic acid wherein at least one of the following is true: (a) the nucleic acid is foreign ("exogenous") to (i.e., not naturally found in) a given host microorganism or host cell; (b) the nucleic acid comprises a nucleotide sequence that is naturally found in (e.g., is "endogenous to") a given host microorganism or host cell (e.g., the nucleic acid comprises a nucleotide sequence endogenous to the host microorganism or host cell. In some embodiments, a "heterologous" nucleic acid may comprise a nucleotide sequence that differs in sequence from the endogenous nucleotide sequence but encodes the same protein (having the same amino acid sequence) as found endogenously; or two or more nucleotide sequences that are not found in the same relationship to each other in nature, e.g., the nucleic acid is recombinant. An example of a heterologous nucleic acid is a nucleotide sequence encoding a fusion protein comprising two proteins that are not joined to one another in nature.

[0040] The term "recombinant" polynucleotide or nucleic acid refers to one that is not naturally occurring, e.g., is made by the artificial combination of two otherwise separated segments of sequence through human intervention. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. A "recombinant" protein is encoded by a recombinant polynucleotide. In the context of a genetically modified host cell, a "recombinant" host cell refers to both the original cell and its progeny.

[0041] As used herein, the term "genetically modified" refers to any change in the endogenous genome of a cyanobacteria cell compared to a wild-type cell. Thus, changes that are introduced through recombinant DNA technology and/or classical mutagenesis

techniques are both encompassed by this term. The changes may involve protein coding sequences or non-protein coding sequences such as regulatory sequences as promoters or enhancers.

[0042] An "expression construct" or "expression cassette" as used herein refers to a recombinant nucleic acid construct, which, when introduced into a cyanobacterial host cell in accordance with the present invention, results in increased expression of a fusion protein encoded by the nucleic acid construct. The expression construct may comprise a promoter sequence operably linked to a nucleic acid sequence encoding the fusion protein or the expression cassette may comprise the nucleic acid sequence encoding the fusion protein where the construct is configured to be inserted into a location in a cyanobacterial genome such that a promoter endogenous to the cyanobacterial host cell is employed to drive expression of the fusion protein. An "expression unit" as used herein refers to a minimal region of a polynucleotide that is expressed that provided for high level protein expression, which comprises the polynucleotide that encodes the fusion protein, as well as other genes, *e.g.*, *cpcA* and *cpc* operon genes encoding cpc linker polypeptides CpcC2, CpcC1, and CpcD. In some embodiments, the expression unit additionally include a gene encoding an antibiotic resistance polypeptide, such as a chloramphenicol resistance gene or streptomycin resistance gene. The expression unit may also comprise additiona sequences, such as nucleic acid sequences encoding a protease cleavage sites, a linker polypeptide, or a polypeptide tagging sequence, such as a His tag.

[0043] By "construct" is meant a recombinant nucleic acid, generally recombinant DNA, which has been generated for the purpose of the expression of a specific nucleotide sequence(s), or is to be used in the construction of other recombinant nucleotide sequences.

[0044] As used herein, the term "exogenous protein" refers to a protein that is not normally or naturally found in and/or produced by a given cyanobacterium, organism, or cell in nature. As used herein, the term "endogenous protein" refers to a protein that is normally found in and/or produced by a given cyanobacterium, organism, or cell in nature.

[0045] An "endogenous" protein or "endogenous" nucleic acid is also referred to as a "native" protein or nucleic acid that is found in a cell or organism in nature.

[0046] The terms "nucleic acid" and "polynucleotide" are used synonymously and refer to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. A nucleic acid of the present invention will generally contain

phosphodiester bonds, although in some cases, nucleic acid analogs may be used that may have alternate backbones, comprising, *e.g.*, phosphoramidate, phosphorothioate, phosphorodithioate, or O-methylphophoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press); and peptide nucleic acid backbones and linkages. Other analog nucleic acids include those with positive backbones; non-ionic backbones, and non-ribose backbones. Thus, nucleic acids or polynucleotides may also include modified nucleotides, that permit correct read through by a polymerase. "Polynucleotide sequence" or "nucleic acid sequence" may include both the sense and antisense strands of a nucleic acid as either individual single strands or in a duplex. As will be appreciated by those in the art, the depiction of a single strand also defines the sequence of the complementary strand; thus the sequences described herein also provide the complement of the sequence. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. The nucleic acid may be DNA, both genomic and cDNA, RNA or a hybrid, where the nucleic acid may contain combinations of deoxyribo- and ribo-nucleotides, and combinations of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine, isoguanine, etc.

[0047] The term "promoter" or "regulatory element" refers to a region or sequence determinants located upstream or downstream from the start of transcription that are involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A "cyanobacteria promoter" is a promoter capable of initiating transcription in cyanobacteria cells. Such promoters need not be of cyanobacterial origin, for example, promoters derived from other bacteria or plant viruses, can be used in the present invention.

[0048] A polynucleotide sequence is "heterologous to" a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified by human action from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants.

[0049] Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the

same when aligned for maximum correspondence as described below. The term "complementary to" is used herein to mean that the sequence is complementary to all or a portion of a reference polynucleotide sequence.

[0050] Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman Add. *APL. Math.* 2:482 (1981), by the homology alignment algorithm of Needle man and Wunsch *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman *Proc. Natl. Acad. Sci.* (U.S.A.) 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

[0051] "Percentage of sequence identity" is 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 percentage of sequence identity.

[0052] The term "substantial identity" in the context of polynucleotide or polypeptide sequences means that a polynucleotide or polypeptide comprises a sequence that has at least 50% sequence identity to a reference nucleic acid or polypeptide sequence. Alternatively, percent identity can be any integer from 40% to 100%. Exemplary embodiments include at least: 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% compared to a reference sequence using the programs described herein; preferably BLAST using standard parameters, as described below.

[0053] Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other, or a third nucleic acid, under stringent conditions. Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of the target sequence

hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is about 0.02 molar at pH 7 and the temperature is at least about 60°C.

[0054] The term "isolated", when applied to a nucleic acid or protein, denotes that the nucleic acid or protein is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state and may be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high-performance liquid chromatography. A protein which is the predominant species present in a preparation is substantially purified. In particular, an isolated gene is separated from open reading frames which flank the gene and encode a protein other than the gene of interest.

[0055] The term "reactor" as used herein refers to the vessel in which cyanobacteria are grown.

Introduction

[0056] The present invention is based, in part, on the discovery of fusion protein constructs that can be used in cyanobacteria as transgenic protein over-expression vectors to provide high levels of transgenic animal proteins, *e.g.*, interferons, insulin, or tPA polypeptides. Expression of transgenes in cyanobacteria using such vectors results in high levels of accumulation of a protein encoded by the transgene.

[0057] A fusion protein of the present invention comprises a protein that is to be expressed in cyanobacteria, typically a non-native protein that is not expressed in cyanobacteria, *e.g.*, a plant protein fused to a protein that is expressed at high levels in cyanobacteria. In the context of the present invention, a protein that is "expressed at high levels in cyanobacteria" refers to a protein that accumulates to at least 1%. Such proteins, when fused at the N-terminus of a protein of interest to be expressed in cyanobacteria, are also referred to herein as "leader proteins", "leader peptides", or "leader sequences". A nucleic acid encoding a leader protein is typically referred to herein as a "leader polynucleotide" or "leader nucleic acid sequence" or "leader nucleotide sequence".

[0058] In some embodiments, a protein that is expressed at high levels is a naturally occurring protein that is expressed at high levels in wild-type cyanobacteria, and is used as endogenous "leader polypeptide sequence" in the cyanobacterial strain of origin. Such

proteins include, *e.g.*, a phycocyanin β-subunit (cpcB), a phycocyanin α-subunit (cpcA), a phycoerythrin α-subunit (cpeA), a phycoerythrin β-subunit (cpeB), an allophycocyanin α-subunit (apcA), an allophycocyanin β-subunit (apcB), a large subunit of Rubisco (rbcL), a small subunit of Rubisco (rbcS), a photosystem II reaction center protein, a photosystem I reaction center protein, or a rpl or rps cyanobacterial ribosomal RNA protein. In some embodiments, a protein that is expressed at high levels is a naturally occurring protein that is expressed at high levels in wild-type cyanobacteria, and it is used as heterologous leader sequence in a different cyanobacterial strain.

[0059] In some embodiments, a protein that is expressed at high levels is an exogenous protein that the cyanobacteria have been genetically modified to express at high levels. For example, proteins that provide for antibiotic resistance that are expressed to high levels in cyanobacteria, *e.g.*, a bacterial kanamycin resistance protein, NPT, or a bacterial chloramphenicol resistance protein, CmR, may be used as a leader sequence.

[0060] The invention additionally provides nucleic acids encoding a fusion protein as described herein, as well as expression constructs comprising the nucleic acids and host cells that have been genetically modified to express such fusion proteins. In further aspects, the invention provides methods of modifying a cyanobacterial cell to overexpress a protein of interest using an expression construct of the invention and methods of producing the protein of interests and products generated by the proteins using such genetically modified cyanobacterial cells.

[0061] The invention employs various routine recombinant nucleic acid techniques. Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those commonly employed in the art. Many manuals that provide direction for performing recombinant DNA manipulations are available, *e.g.*, Sambrook, Molecular Cloning, A Laboratory Manual (4th Ed, 2012); and Current Protocols in Molecular Biology (Ausubel et al., eds., 1994-2015).

Proteins expressed at high levels in cyanobacteria

[0062] In the present invention, nucleic acid constructs are created in which a polynucleotide sequence encoding a protein of interest is fused to the C-terminal end of a polynucleotide that encodes a leader protein, i.e., a protein that is expressed at high levels in cyanobacteria as described herein. The protein of interest is then also expressed at high

levels in conjunction with the leader sequence. In the context of the invention, a protein that is "expressed at high levels" in cyanobacteria refers to a protein that is at least 1%, typically at least 2%, at least 3%, at least 4%, at least 5%, or at least 10%, or greater, of the total protein expressed in the cyanobacteria. Expression levels in cyanobacteria may be evaluated in cells that are logarithmically growing, but may be alternatively determined in cells in a stationary phase of growth. The level of protein expression can be assessed using various techniques. In the present invention, high level expression is typically determined using SDS PAGE analysis. Following electrophoresis, the gel is stained and the level of proteins assessed by scanning the gel and quantifying the amount of protein using an image analyzer.

[0063] In some embodiments, a leader sequence in accordance with the invention encodes a naturally occurring cyanobacteria protein that is expressed at high levels in native cyanobacteria. Thus, in some embodiments, the protein is endogenous to cyanobacteria. Examples of such proteins include cpcB, cpcA, cpeA, cpeB, apcA, apcB, rbcL, rbcS, psbA, rpl, or rps. In some embodiments, the leader sequence encodes less than the full-length of the protein, but typically comprises a region that encodes at least 25%, typically at least 50%, or at least 75%, or at least 90%, or at least 95%, or greater, of the length of the protein. As appreciated by one of skill in the art, use of an endogenous cyanobacterial polynucleotide sequence for constructing an expression construct in accordance with the invention provides a sequence that need not be codon-optimized, as the sequence is already expressed at high levels in cyanobacteria. Examples of cyanobacterial polynucleotides that encode cpcB, cpcA, cpeA, cpeB, apcA, apcB, rbcL, rbcS, psbA, rpl, or rps are available at the website www.genome.microbedb.jp/cyanobase under accession numbers, as follows:

- cpcA: *Synechocystis* sp. PCC6803 sll1578, *Anabaena* sp. PCC7120 arl0529, *Thermosynechococcus elongatus* BP-1 tlr1958, *Synechococcus elongatus* PCC6301 syc0495_c, syc0500_c
- cpcB: *Synechocystis* sp. PCC6803 sll1577, *Anabaena* sp. PCC7120 arl0528, *Thermosynechococcus elongatus* BP-1 tlr1957, *Synechococcus elongatus* PCC6301 syc0496_c, syc0501_c
- cpeA: *Prochlorococcus marinus* SS120 Pro0337, *Synechococcus* sp. WH8102 SYNW2009, SYNW2016
- cpeB: *Prochlorococcus marinus* SS120 Pro0338, *Synechococcus* sp. WH8102 SYNW2008, SYNW2017

- apcA: *Synechocystis* sp. PCC 6803, slr2067; *Anabaena* sp. PCC 7120, all0450, alr0021; *Synechococcus elongatus* PCC 6301, syc1186_d
- apcB: *Synechocystis* sp. PCC 6803, slr1986, *Anabaena* sp. PCC 7120, alr0022, *Synechococcus elongatus* PCC 6301, syc1187_d
- rbcL RubisCO large subunit: *Synechocystis* sp. PCC 6803 slr0009
- rbcS RubisCO small subunit: *Synechocystis* sp. PCC 6803 slr0012
- rpl: 50S ribosomal protein of *Synechocystis*, *e.g.* sll1803; sll1810; ssr1398 and
- rps: 30S ribosomal protein of *Synechocystis*, *e.g.* sll1804; slr1984.

[0064] The polynucleotide sequence that encodes the leader protein need not be 100% identical to a native cyanobacteria polynucleotide sequence. A polynucleotide variant having at least 50% identity or at least 60% identity, or greater, to a native cyanobacterial polynucleotide sequence, *e.g.*, a native cpcB, cpcA, cpeA, cpeB, rbcL, rbcS, psbA, rpl, or rps cyanobacteria polynucleotide sequence, may also be used, so long as the codons that vary relative to the native cyanobacterial polynucleotide are codon optimized for expression in cyanobacteria and the codons that vary relative to the wild type sequence do not substantially disrupt the structure of the protein. In some embodiments, a polynucleotide variant that has at least 70% identity, at least 75% identity, at least 80% identity, or at least 85% identity, or greater to a native cyanobacterial polynucleotide sequence, *e.g.*, a native *cpcB*, *cpcA*, *cpeA*, *cpeB*, *rbcL*, *rbcS*, *psbA*, *rpl*, or *rps* cyanobacteria polynucleotide sequence, is used, again maintaining codon optimization for cyanobacteria. In some embodiments, a polynucleotide variant that has least 90% identity, or at least 95% identity, or greater, to a native cyanobacterial polynucleotide sequence, *e.g.*, a native *cpcB*, *cpcA*, *cpeA*, *cpeB*, *rbcL*, *rbcS*, *psbA*, *rpl*, or *rps* cyanobacteria polynucleotide sequence, is used. The percent identity is typically determined with reference the length of the polynucleotide that is employed in the construct, i.e., the percent identity may be over the full length of a polynucleotide that encodes the leader polypeptide sequence, or may be over a smaller length, *e.g.*, in embodiments where the polynucleotide encodes at least 25%, typically at least 50%, or at least 75%, or at least 90%, or at least 95%, or greater, of the length of the protein. The protein encoded by a variant polynucleotide sequence as described need not retain a biological function, however, a codon that varies from the wild-type polynucleotide is typically selected such that the protein structure of the native cyanobacterial sequence is not

substantially altered by the changed codon, *e.g.*, a codon that encodes an amino acid that has the same charge, polarity, and/or is similar in size to the native amino acid is selected.

[0065] In some embodiments, a polynucleotide variant of a naturally over-expressed (more than 1% of the total cellular protein) cyanobacterial gene is employed, that encodes for a polypeptide sequence that has at least 70%, or 80%, or at least 85% or greater identity to the protein encoded by the wild-type gene. In some embodiments, the polynucleotide encodes a protein that has 90% identity, or at least 95% identity, or greater, to the protein encoded by the wild-type gene. Variant polynucleotides may also be codon optimized for expression in cyanobacteria.

[0066] In some embodiments, a protein that is expressed at high levels in cyanobacteria is not native to cyanobacteria in which a fusion construct in accordance with the invention is expressed. For example, polynucleotides from bacteria or other organisms that are expressed at high levels in cyanobacteria may be used as leader sequences. In some embodiments, the polynucleotides from other organisms may be codon-optimized for expression in cyanobacteria. In some embodiments, codon optimization is performed such that codons used with an average frequency of less than 12% by *Synechocystis* are replaced by more frequently used codons. Rare codons can be defined, e.g., by using a codon usage table derived from the sequenced genome of the host cyanobacterial cell. See, *e.g.*, the codon usage table obtained from Kazusa DNA Research Institute, Japan (website www.kazusa.or.jp/codon/) used in conjunction with software, *e.g.*, "Gene Designer 2.0" software, from DNA 2.0 (website www.dna20.com/) at a cut-off thread of 15%.

[0067] In some embodiments, a leader sequence in accordance with the present invention encodes a protein that confers antibiotic resistance. For example, in some embodiments, the leader sequence encodes neomycin phosphotransferase *e.g.*, NPT1, which confers neomycin and kanamycin resistance. Other polynucleotides that may be employed include a chloramphenicol acetyltransferase polynucleotide, which confers chloramphenicol resistance; or a polynucleotide encoding a protein that confers streptomycin, ampicillin, erythromycin, zeocin, or tetracycline resistance, or resistance to another antibiotic. In some embodiments, the leader sequence encodes less than the full-length of the protein, but typically comprises a region that encodes at least 25%, typically at least 50%, or at least 75%, or at least 90%, or at least 95%, or greater, of the length of the protein. In some embodiments, a polynucleotide variant of a naturally occurring antibiotic resistance gene is employed. As noted above, a

variant polynucleotide need not encode a protein that retains the native biological function. A variant polynucleotide typically encodes a protein that has at least 80% identity, or at least 85% or greater, identity to the protein encoded by the wild-type antibiotic resistance gene. In some embodiments, the polynucleotide encodes a protein that has 90% identity, or at least 95% identity, or greater, to the wild-type antibiotic resistance protein. Such variant polynucleotides employed as leader sequence may also be codon-optimized for expression in cyanobacteria. The percent identity is typically determined with reference to the length of the polynucleotide that is employed in the construct, i.e., the percent identity may be over the full length of a polynucleotide that encodes the leader polypeptide sequence, or may be over a smaller length, *e.g.*, in embodiments where the polynucleotide encodes at least 25%, typically at least 50%, or at least 75%, or at least 90%, or at least 95%, or greater, of the length of the protein. A protein encoded by a variant polynucleotide sequence need not retain a biological function, however, codons that are present in a variant polynucleotide are typically selected such that the protein structure relative to the wild-type protein structure is not substantially altered by the changed codon, *e.g.*, a codon that encodes an amino acid that has the same charge, polarity, and/or is similar in size to the native amino acid is selected.

[0068] Other leader proteins can be identified by evaluating the level of expression of a candidate leader protein in cyanobacteria. For example, a leader polypeptide that does not occur in wild type cyanobacteria may be identified by measuring the level of protein expressed from a polynucleotide codon optimized for expression in cyanobacteria that encodes the candidate leader polypeptide. A protein may be selected for use as a leader polypeptide if the protein accumulates to a level of at least 1%, typically at least 2%, at least 3%, at least 4%, at least 5%, or at least 10%, or greater, of the total protein expressed in the cyanobacteria when the polynucleotide encoding the leader polypeptide is introduced into cyanobacteria and the cyanobacteria cultured under conditions in which the transgene is expressed. The level of protein expression is typically determined using SDS PAGE analysis. Following electrophoresis, the gel is scanned and the amount of protein determined by image analysis.

Transgenes

[0069] A fusion construct of the invention may be employed to provide high level expression in cyanobacteria for any desired biopharmaceutical protein. Thus, for example, cyanobacteria can be engineered to express an animal biopharmaceutical polypeptide such as

an antibody, hormone, cytokine, therapeutic enzyme and the like, as a fusion polypeptide with a protein expressed at a high level in cyanobacteria, *e.g.* a cpcB or other protein encoded by the Cpc operon. In some embodiments the biopharmaceutical polypeptide is expressed at a level of at least 1%, or at least 5%, or at least 10%, or at least 15%, or at least 20%, of total cellular protein as described herein.

[0070] In some embodiments, the nucleic acid sequence encoding the animal, *e.g.*, mammalian, biopharmaceutical polypeptide is codon-optimized for expression in cyanobacteria. Alternatively, the nucleic acid sequence need not be codon-optimized, as high-level expression of the fusion polypeptide does not require codon optimization.

[0071] In some embodiments, the mature form of the biopharmaceutical polypeptide lacking the native signal sequence is expressed.

[0072] In some embodiments, the transgene that is expressed encodes an interferon, *e.g.*, an interferon alpha, such as IFNA2. In some embodiments, the interferon is interferon-alpha, such as human interferon α-2. An illustrative polypeptide sequence is available under uniprot number P01563. The amino acid sequence of a mature form of human interferon alpha-2, which lacks the signal polypeptide, is provided in SEQ ID NO:1. In some embodiments, the IFNA2 protein is expressed as a fusion construct with cpcB, *e.g.*, by replacing the *cpcB* gene in the *cpc* operon with a transgene encoding a *cpcB**interferon fusion construct. In some embodiments, the transgene encodes an interferon polypeptide fused to an antibiotic resistance polypeptide, such as Npt1. In some embodiments, such a fusion polypeptide is introduced into the *cpc* operon for expression. In some embodiments, the gene encoding the Npt1*interferon fusion polypeptides is inserted to replace the *cpcb* gene in the *cpc* operon. In some embodiments, the fusion polypeptide comprises a protease cleavage site such as a Factor Xa cleavage site or alternative cleavage site, *e.g.*, a Tobacco Etch Virus (TEV) cysteine protease cleavage site. Alternatively, the fusion polypeptide may comprise an Enteropeptidase, Thrombin, Protease 3C, Sortase A, Genase I, Intein, or a Snac-tag cleavage site (*e.g.*, Kosobokova et al. 2016; Dang et al. 2019). In some embodiments, the fusion polypeptide may comprise a protein purification tag, such as a 6XHis tag.

[0073] In some embodiments, the transgene that is expressed encodes a tPA, *e.g.*, a human tPA lacking a native signal sequence. Human tPA has a molecular weight of about 70kDa in the single-chain form. The tPA polypeptide had five domains: an N-terminal finger domain, an epidermal growth factor domain, a serine protease domain, and Kringle 1 and Kringle 2

domains. In some embodiments, the tPA polypeptide that is expressed is a truncated human tissue plasminogen activator (K2S, reteplase), which includes the Kringle 2 domain and the serine protease domain. Illustrative examples of tPA polypeptide sequences that can be expressed in accordance with the invention are shown in SEQ ID NOS:2 and 3. In some embodiments, the tPA that is expressed lacks the signal polypeptide. In some embodiments, the tPA incorporated into the fusion polypeptide has the amino acid sequence of SEQ ID NO:3. In some embodiments, the IFNA2 protein is expressed as a fusion construct with cpcB, *e.g.*, by replacing the *cpcB* gene in the *cpc* operon with a transgene encoding a *cpcB**tPA fusion construct. In some embodiments, the transgene encodes a tPA polypeptide fused to an antibiotic resistance polypeptide, such as Npt1. In some embodiments, such a fusion polypeptide is introduced into the *cpc* operon for expression. In some embodiments, the gene encoding the Npt1*tPA fusion polypeptides is inserted to replace the *cpcb* gene in the *cpc* operon. In some embodiments, the fusion polypeptide comprises a protease cleavage site such as a Factor Xa cleavage site or alternative cleavage site, *e.g.*, a TEV cysteine protease cleavage site. Alternatively, the fusion polypeptide may comprise an Enteropeptidase, Thrombin, Protease 3C, Sortase A, Genase I, Intein, or a Snac-tag cleavage site (*e.g.*, Kosobokova et al. 2016; Dang et al. 2019). In some embodiments, the fusion polypeptide may comprise a protein purification tag, such as a 6XHis tag.

[0074] In some embodiments, the transgene that is expressed encodes an insulin *e.g.*, a human insulin. An illustrative polypeptide sequence is available under uniprot number P01308. The amino acid sequence of a mature form of human insulin, which lacks the signal polypeptide, is provided in SEQ ID NO:4. In some embodiments, the insulin protein is expressed as a fusion construct with cpcB, *e.g.*, by replacing the *cpcB* gene in the *cpc* operon with a transgene encoding a *cpcB**insulin fusion construct. In some embodiments, the transgene encodes an insulin polypeptide fused to an antibiotic resistance polypeptide, such as Npt1. In some embodiments, such a fusion polypeptide is introduced into the *cpc* operon for expression. In some embodiments, the gene encoding the Npt1*insulin fusion polypeptides is inserted to replace the *cpcb* gene in the *cpc* operon. In some embodiments, the fusion polypeptide comprises a protease cleavage site such as a Factor Xa cleavage site or alternative cleavage site, *e.g.*, a TEV cysteine protease cleavage site. Alternatively, the fusion polypeptide may comprise an Enteropeptidase, Thrombin, Protease 3C, Sortase A, Genase I, Intein, or a Snac-tag cleavage site (*e.g.*, Kosobokova et al. 2016; Dang et al. 2019).

In some embodiments, the fusion polypeptide may comprise a protein purification tag, such as a 6XHis tag.

[0075] As noted above, in some embodiments, the transgene portion of a fusion construct in accordance with the invention may be codon optimized for expression in cyanobacteria. For example, in some embodiments, codon optimization is performed such that codons used with an average frequency of less than 12% by *Synechocystis* are replaced by more frequently used codons. Rare codons can be defined, e.g., by using a codon usage table derived from the sequenced genome of the host cyanobacterial cell. See, *e.g.*, the codon usage table obtained from Kazusa DNA Research Institute, Japan (website www.kazusa.or.jp/codon/) used in conjunction with software, *e.g.*, "Gene Designer 2.0" software, from DNA 2.0 (website www.dna20.com/) at a cut-off thread of 15%; or the software available at the website, idtdna.com/CodonOpt.

Preparation of recombinant expression constructs

[0076] Recombinant DNA vectors suitable for transformation of cyanobacteria cells are employed in the methods of the invention. Preparation of suitable vectors and transformation methods can be prepared using any number of techniques, including those described, *e.g.*, in Sambrook, Molecular Cloning, A Laboratory Manual (4th Ed, 2012); and Current Protocols in Molecular Biology (Ausubel et al., eds., 1994-2015). For example, a DNA sequence encoding a fusion protein of the present invention will be combined with transcriptional and other regulatory sequences to direct expression in cyanobacteria.

[0077] In some embodiments, the vector includes sequences for homologous recombination to insert the fusion construct at a desired site in a cyanobacterial genome, *e.g.*, such that expression of the polynucleotide encoding the fusion construct will be driven by a promoter that is endogenous to the organism. A vector to perform homologous recombination will include sequences required for homologous recombination, such as flanking sequences that share homology with the target site for promoting homologous recombination.

[0078] Regulatory sequences incorporated into vectors that comprise sequences that are to be expressed in the modified cyanobacterial cell include promoters, which may be either constitutive or inducible. In some embodiments, a promoter for a nucleic acid construct is a constitutive promoter. Examples of constitutive strong promoters for use in cyanobacteria include, for example, the *psbD1* gene or the basal promoter of the *psbD2* gene, or the rbcLS promoter, which is constitutive under standard growth conditions. Various other promoters

that are active in cyanobacteria are also known. These include the strong *cpc* operon promoter, the *cpe* operon and *apc* operon promoters, which control expression of phycobilisome constituents. The light inducible promoters of the *psbA1, psbA2,* and *psbA3* genes in cyanobacteria may also be used, as noted below. Other promoters that are operative in plants, *e.g.*, promoters derived from plant viruses, such as the CaMV35S promoters, or bacterial viruses, such as the T7, or bacterial promoters, such as the PTrc, can also be employed in cyanobacteria. For a description of strong and regulated promoters, *e.g.*, active in the cyanobacterium *Anabaena* sp. strain PCC 7120 and *Synechocystis* 6803, see *e.g.*, Elhai, *FEMS Microbiol Lett* 114:179-184, (1993) and Formighieri, *Planta* 240:309–324 (2014).

[0079] In some embodiments, a promoter can be used to direct expression of the inserted nucleic acids under the influence of changing environmental conditions. Examples of environmental conditions that may affect transcription by inducible promoters include anaerobic conditions, elevated temperature, or the presence of light. Promoters that are inducible upon exposure to chemicals reagents are also used to express the inserted nucleic acids. Other useful inducible regulatory elements include copper-inducible regulatory elements (Mett *et al., Proc. Natl. Acad. Sci. USA* 90:4567-4571 (1993); Furst *et al., Cell* 55:705-717 (1988)); copper-repressed petJ promoter in *Synechocystis* (Kuchmina *et al.* 2012, *J Biotechn* 162:75-80); riboswitches, e.g. theophylline-dependent (Nakahira *et al.* 2013, *Plant Cell Physiol* 54:1724-1735; tetracycline and chlor-tetracycline-inducible regulatory elements (Gatz *et al., Plant J.* 2:397-404 (1992); Röder *et al., Mol. Gen. Genet.* 243:32-38 (1994); Gatz, *Meth. Cell Biol.* 50:411-424 (1995)); ecdysone inducible regulatory elements (Christopherson *et al., Proc. Natl. Acad. Sci. USA* 89:6314-6318 (1992); Kreutzweiser *et al., Ecotoxicol. Environ. Safety* 28:14-24 (1994)); heat shock inducible promoters, such as those of the hsp70/dnaK genes (Takahashi *et al., Plant Physiol.* 99:383-390 (1992); Yabe *et al., Plant Cell Physiol.* 35:1207-1219 (1994); Ueda *et al., Mol. Gen. Genet.* 250:533-539 (1996)); and lac operon elements, which are used in combination with a constitutively expressed lac repressor to confer, for example, IPTG-inducible expression (Wilde *et al., EMBO J.* 11:1251-1259 (1992)). An inducible regulatory element also can be, for example, a nitrate-inducible promoter, *e.g.*, derived from the spinach nitrite reductase gene (Back *et al., Plant Mol. Biol.* 17:9 (1991)), or a light-inducible promoter, such as that associated with the small subunit of RuBP carboxylase or the LHCP gene families (Feinbaum *et al., Mol. Gen. Genet.* 226:449 (1991); Lam and Chua, Science 248:471 (1990)).

[0080] In some embodiments, the promoter may be from a gene associated with photosynthesis in the species to be transformed or another species. For example, such a promoter from one species may be used to direct expression of a protein in transformed cyanobacteria cells. Suitable promoters may be isolated from or synthesized based on known sequences from other photosynthetic organisms. Preferred promoters are those for genes from other photosynthetic species, or other photosynthetic organism where the promoter is active in cyanobacteria.

[0081] A vector will also typically comprise a marker gene that confers a selectable phenotype on cyanobacteria transformed with the vector. Such marker genes, include, but are not limited to those that confer antibiotic resistance, such as resistance to chloramphenicol, kanamycin, spectinomycin, G418, bleomycin, hygromycin, and the like.

[0082] Cell transformation methods and selectable markers for cyanobacteria are well known in the art (Wirth, *Mol. Gen. Genet.*, 216(1):175-7 (1989); Koksharova, *Appl. Microbiol. Biotechnol.*, 58(2): 123-37 (2002); Thelwell *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 95:10728-10733 (1998)).

[0083] Any suitable cyanobacteria may be employed to express a fusion protein in accordance with the invention. These include unicellular cyanobacteria, micro-colonial cyanobacteria that form small colonies, and filamentous cyanobacteria. Examples of unicellular cyanobacteria for use in the invention include, but are not limited to, *Synechococcus* and *Thermosynechococcus* sp., *e.g.*, *Synechococcus sp.* PCC 7002, *Synechococcus sp.* PCC 6301, and *Thermosynechococcus elongatus*; as well as *Synechocystis* sp., such as *Synechocystis sp.* PCC 6803; and *Cyanothece sp.*, such as PCC 8801. Examples of micro-colonial cyanobacteria for use in the invention, include, but are not limited to, *Gloeocapsa magma, Gloeocapsa phylum, Gloeocapsa alpicola, Gloeocpasa atrata, Chroococcus* spp., and *Aphanothece* sp. Examples of filamentous cyanobacteria that can be used include, but are not limited to, *Oscillatoria* spp., *Nostoc sp.*, *e.g.*, *Nostoc* sp. PCC 7120, and *Nostoc sphaeroides*; *Anabaena* sp., *e.g.*, *Anabaena variabilis* and *Arthrospira* sp. ("Spirulina"), such as *Arthrospira platensis* and *Arthrospira maxima, and Mastigocladus laminosus*. Cyanobacteria that are genetically modified in accordance with the invention may also contain other genetic modifications, *e.g.*, modifications to the terpenoid pathway, to enhance production of a desired compound.

[0084] Cyanobacteria can be cultured to high density, *e.g.*, in a photobioreactor (*see, e.g.*, Lee *et al.*, *Biotech. Bioengineering* 44:1161-1167, 1994; Chaumont, *J Appl. Phycology* 5:593-604, 1990) to produce the protein encoded by the transgene. In some embodiments, the protein product of the transgene is purified. In many embodiments, the cyanobacteria culture is used to produce a desired, non-protein product, *e.g.*, isoprene, a hemiterpene; β-phellandrene, a monoterpene; farnesene, a sesquiterpene; or other products. The product produced from the cyanobacteria may then be isolated or collected from the cyanobacterial cell culture.

EXAMPLES

[0085] The following examples illustrate the over-expression of illustrative biopharmaceutical polypeptides in cyanobacteria.

Example 1. Expression of an interferon in cyanobacteria.

*cpcB*IFN fusion constructs*

[0086] This example demonstrates the expression of the mature human interferon α-2 protein (Uniprot No. P01563), referred to in this example as IFN, in the cyanobacteria *Synechocystis* sp. PCC 6803 (*Synechocystis*). To validate the fusion constructs approach, three different DNA constructs were designed for the transformation of wild type (WT) *Synechocystis* through double homologous DNA recombination in the *cpc* operon locus (**Fig. 1A**). The nucleic acid construct *IFN* (**Fig. 1B**) was codon optimized for expression in *Synechocystis*, and designed to replace the *cpcB* gene in the *cpc* operon. *IFN* was followed by the chloramphenicol resistance cassette (*cmR*) in an operon configuration. Construct *cpcB-IFN* (**Fig. 1C**) was designed to insert both the *IFN* and the *cmR* genes after the *cpcB* gene in an operon configuration. Finally, construct *cpcB*IFN* (**Fig. 1D**) was designed to replace the *cpcB* gene in the *cpc* operon with the fusion construct *cpcB*IFN*, followed by the *cmR* gene in an operon configuration. A Factor Xa cleavage-encoding sequence was inserted between the *cpcB* and *IFN* genes in the construct of Fig. 1D.

[0087] PCR analysis to determine whether transgenic DNA copy homoplasmy was achieved. Primers *cpc-us for* and *cpcA rev* were designed on the flanking regions of the transgenic DNA insertion sites (**Fig. 1**). PCR amplification using WT genomic DNA as a template generated a product of 1,289 bp (**Fig. 2**). PCR amplification using DNA from the transformant IFN, CpcB-IFN, and CpcB*IFN strains generated the expected product sizes of

2,094 bp, 2,723 bp and 2,619 bp, respectively. DNA copy homoplasmy was evidenced by the absence of WT PCR products in the PCR amplification reactions of the IFN transformants.

[0088] After DNA copy homoplasmy was achieved, WT and transformant strains were grown photo-autotrophically in liquid BG-11 cultures. The visual phenotype (**Fig. 3**) was noticeably different between the WT and transformant strains. The WT cells had a blue-green coloration, consistent with the presence of blue phycocyanin and green chlorophyll pigments in their functional light-harvesting antennae. All transformant strains showed a yellow-green pigmentation, suggesting lack of phycocyanin, which is responsible for the blue pigmentation of the cells. This is consistent with previously reported results (Kirst et al. 2014; Formighieri and Melis 2015; Chaves and Melis 2016) and underscores the absence of assembled phycocyanin rods in the transformants.

[0089] Protein analysis of total cell extracts from WT and transformant *Synechocystis* was performed using SDS-PAGE followed by Coomassie blue staining and Western blot analysis (**Fig. 4**). Two replicate samples of WT protein extracts showed the presence of CpcB β-subunit and CpcA α-subunit of phycocyanin as the dominant protein bands, migrating to ~19 and ~17 kD, respectively. Another dominant band in the SDS-PAGE profile was the large subunit of Rubisco (RbcL), migrating to about ~56 kD (**Fig. 4A**). The latter was used as a normalization factor in protein quantification and as a loading control of the gels.

[0090] CpcB and CpcA subunits were not evident in the protein extracts of the transformants because of inability of these transformants to assemble the phycobilisome-peripheral phycocyanin rods. The *IFN* and *cpcB-IFN* transformants failed to show accumulation of recombinant IFN protein in the expected ~19 kD region, both in the SDS-PAGE and the associated Western blot (**Fig. 4B, IFN** and **CpcB-IFN**), suggesting either very-low levels or absence of the recombinant IFN protein from these samples. These results show that the powerful cpc promoter was not sufficient to support IFN (~19 kD) protein expression / accumulation in *Synechocystis*. In contrast, protein extracts from the *cpcB*IFN* fusion transformants showed a clear presence of an abundant protein with electrophoretic mobility to ~36 kD. This band was attributed to accumulation of the CpcB*IFN fusion protein (**Fig. 4A, CpcB*IFN**). Identification of the ~36 kD protein was tested by Western blot analysis with specific polyclonal antibodies raised against the human IFN protein (**Fig. 4B, CpcB*IFN**). A strong reaction between the polyclonal antibodies and a protein band

migrating to ~36 kD suggested that this band is the recombinant CpcB*IFN protein. Moreover, binding was also detected with protein bands at a higher MW, suggesting the formation / presence of complexes (~108 kD) containing the CpcB*IFN fusion protein.

[0091] To evaluate the effect of DNA codon-use optimization on the IFN protein expression level, CpcB*IFN fusion DNA constructs were designed using the *Synechocystis* codon optimized IFN as well as the native unoptimized human DNA sequence (termed IFN') for comparative expression measurements in *Synechocystis*. The latter construct harbored the same elements of the CpcB*IFN fusion, with the exception of the *IFN* gene that was replaced by the human native *IFN'* sequence (no codon-use optimization). Wild type (WT), *cpcB*IFN'*, and *cpcB*IFN* transformant strains were grown in parallel, and total cell proteins were extracted and subjected to SDS-PAGE analysis. Upon Coomassie staining of the SDS-PAGE (**Fig. 5**), the WT protein extract showed as main subunits the 56 kD RbcL, 19 kD CpcB, and 17 kD CpcA. The latter two subunits were missing from the extract of the transformant cells, shown in three independent replicates per transformant in Fig. 5. Densitometric analysis of Coomassie stained SDS-PAGE (**Fig. 5**) showed the presence of RbcL to ~12.5% of total cellular protein. Fusion constructs accumulated to ~10.2% in the *cpcB*IFN'* and ~11.8% in *cpcB*IFN* codon-optimized transformant strains. Validation of the Coomassie stained SDS-PAGE protein assignments was obtained through Western blot analysis with specific polyclonal antibodies (not shown).

[0092] The above results showed that IFN successfully accumulated in *Synechocystis* only when expressed in a fusion construct configuration with the native highly-expressed CpcB subunit of phycocyanin, regardless of whether the *IFN* gene was codon-optimized or not. In order to isolate the recombinant fusion protein, we designed a new DNA construct referred to as the *cpcB*His*Xa*IFN*, based on the previous CpcB*IFN construct (**Fig. 6**). A DNA fragment encoding the domain of six histidines and the Factor Xa cleavage-site was inserted between the *cpcB* and the *IFN* genes in the fusion construct. Protein analysis was then conducted on the transformant lines. Coomassie staining of the SDS-PAGE profile (**Fig. 6**) showed the abundant RbcL, CpcB and CpcA subunits in the wild type extracts (**Fig. 6, WT**). The *cpcB*IFN* transformants lacked the CpcB and CpcA proteins but accumulated the CpcB*IFN as a ~36 kD protein (**Fig. 6, CpcB*IFN**). The *cpcB*His*Xa*IFN* transformants also lacked the CpcB and CpcA proteins but accumulated an abundant protein band with a slightly higher apparent molecular mass than that of the CpcB*IFN (**Fig. 6, CpcB*His*Xa*IFN**). This band was attributed to the CpcB*His*Xa*IFN protein. The fact

that CpcB*His*Xa*IFN protein band showed a similar abundance as that of the CpcB*IFN construct suggested that the His*Xa addition to the CpcB*IFN fusion did not adversely affect the expression level of this recombinant protein.

*Batch-based purification of the cpcB*His*Xa*IFN recombinant protein*

[0093] We initially applied a "batch" purification procedure to the recombinant CpcB*His*Xa*IFN protein using a His-Select resin (Sigma) and by following the manufacturer's instructions. The procedure was conducted in Eppendorf tubes, thereby minimizing the amount of resin and cell extract used. Total cell extracts from WT, *cpcB*IFN*, and *cpcB*His*Xa*IFN* fusion construct transgenic cells were employed in a side-by-side comparative resin treatment and purification analysis. Prior to incubation with the resin, cellular extracts were incubated on ice for 20 min in the presence of 1% Triton X-100 to disperse cellular aggregates that appeared to interfere with the precipitation of the resin upon centrifugation. Un-solubilized cell debris were pelleted and discarded following a brief centrifugation. The supernatant, containing the cellular protein extracts, was incubated with the resin for 5 min, followed by centrifugation to pellet the resin and any His-tagged proteins bound to it.

[0094] Lane 1 in **Fig. 7** shows the cell extracts (upper panel) and the resin (lower panel) of the wild type, *cpcB*IFN*, and *cpcB*His*Xa*IFN* fusion construct transgenic cells prior to incubation with the resin. The resin had a natural pink coloration.

[0095] Lane 2 in **Fig. 7** shows the cell extracts (upper panel) and the resin pellet (lower panel) of the wild type, *cpcB*IFN*, and *cpcB*His*Xa*IFN* cell lines following a 5-min incubation with the resin and a subsequent centrifugation. There was a blue coloration of the resin pellet and green coloration of the supernatant.

[0096] Lanes 3-5 in **Fig. 7** show the remaining extracts (upper panels) and the resin pellet (lower panels) of the wild type, *cpcB*IFN*, and *cpcB*His*Xa*IFN* cell lines following a consecutive wash of the resin with a buffer containing 10 mM imidazole to remove non-target proteins. The supernatant was clear and there was a pink coloration of the resin after the third wash (lane 5) for the wild type and *cpcB*IFN* transformants, suggesting absence of His-tagged proteins. There was a blue coloration of the resin in the *cpcB*His*Xa*IFN* sample, which was retained in this pellet (lanes 3-5) in spite of the repeated 10 mM imidazole wash, suggesting the presence and binding to the resin of blue-colored His-tagged proteins.

[0097] Lanes 6-8 in **Fig. 7** show the subsequent extracts (upper panel) and the resin pellet (lower panel) of the wild type, *cpcB*IFN*, and *cpcB*His*Xa*IFN* cell lines following a wash of the resin three times with a buffer containing 250 mM of imidazole, designed to dissociate His-tagged proteins from the resin. There was a bluish color to the supernatant in lanes 6 and 7 and a corresponding loss of the blue color from the resin pellet, suggesting the specific removal of His-tagged proteins from the resin under these conditions.

[0098] Fractions eluted from the resin upon application of 250 mM imidazole were analyzed by SDS-PAGE (**Fig. 8**). Elution fractions from both WT and the *cpcB*IFN* transgenic extracts showed no protein bands in the Coomassie stained gels (**Fig. 8**, left and middle panels), whereas eluent 1 (E1) from the *cpcB*His*Xa*IFN* extracts clearly showed the presence of protein bands, with the most abundant migrating to ~36 kD, attributed to the CpcB*His*Xa*IFN fusion protein. Secondary bands migrating to ~17 kD, ~27 kD, and ~108 kD were also noted (**Fig. 8**, right panel). The ~17 kD protein was attributed to the CpcA α-subunit of phycocyanin. The ~27 kD protein could be the CpcG1 subunit of the phycobilisome, a phycocyanin rod-core linker polypeptide (Kondo et al., 2005), and the ~108 kD band is tentatively attributed to a CpcB*His*Xa*IFN trimer, as it was shown to contain the CpcB*His*Xa*IFN fusion protein (see below).

[0099] The nature of the pigmentation of proteins from eluent 1 of the cell extracts was investigated through spectrophotometric analysis (**Fig. 9A**). The spectra of E1 from the WT and CpcB*IFN extracts did not show any absorbance features, consistent with absence of coloration in lanes 6-8 (Fig. 7) of these samples. Eluent 1 from the CpcB*His*Xa*IFN sample showed a distinct absorbance band with a peak at ~625 nm and a secondary broad band peaking in the UV-A region of the spectrum. This closely resembled the absorbance spectrum of phycocyanin from *Synechocystis* (Kirst et al. 2014), suggesting the presence of bilin pigment covalently-bound to the CpcB*His*Xa*IFN fusion protein. To further investigate this observation, absorbance spectra of total protein extracts from WT and *cpcB*His*Xa*IFN* transformant cells were also measured. These were compared with the absorbance spectrum of cells lacking phycocyanin due to a Δcpc operon deletion (Kirst et al. 2014). The spectrum of WT cells showed typical absorbance bands of chlorophyll at 680 nm and phycocyanin at 625 nm (**Fig. 9B**). The extract from the Δcpc transformants showed the specific Chl absorbance peak at 680 nm, whereas the phycocyanin absorbance peak at around 625 nm was missing (**Fig. 9B**). The absorbance spectrum from the *cpcB*His*Xa*IFN* transformant cells showed a substantially lower absorbance at about 625 nm due to depletion

of phycocyanin, but this lowering was not as extensive as that observed with the Δcpc cells (**Fig. 9B**). The difference, and apparent low-level absorbance of the *cpcB*His*Xa*IFN* cells at 625 nm, suggests that the CpcB protein, albeit in a fusion construct configuration with the IFN, and/or the CpcA protein that apparently accompanies this recombinant protein, covalently bind at least some of the phycobilin pigment that is naturally associated with it, and which is manifested in the blue coloration of the E1 eluent.

*Column-based purification of the cpcB*His*Xa*IFN recombinant proteins*

[0100] Based on the initial encouraging results obtained with the "batch" purification approach, we proceeded to conduct a "column-based" purification of the His-tagged proteins (**Fig. 10**). This experimental work was conducted as an alternative method in an attempt to elute a greater amount of the CpcB*His*Xa*IFN protein. Total protein extract from the *cpcB*His*Xa*IFN* transformant cells, mixed with 5 mM imidazole, was loaded onto the resin. Four subsequent washing steps were conducted with 5 mM imidazole to remove non-target proteins from the resin. After these washing steps, elution of the target protein with 250 mM imidazole was undertaken. The pigmentation pattern of the resulting fractions was in accordance with the results obtained with the "batch-based" purification (please see below).

[0101] Lane 1 in **Fig. 10,** upper panel, shows the *cpcB*His*Xa*IFN* cell extracts that were incubated in the presence of 5 mM imidazole prior to loading on the resin. Lane 1 in **Fig. 10,** lower panel, shows the SDS-PAGE protein profile of these extracts, indicating presence of all expected *Synechocystis* proteins.

[0102] Lane 2 in **Fig. 10,** upper panel, shows the *cpcB*His*Xa*IFN* cell extracts after incubation with the resin but prior to washing with additional imidazole. Lane 2 in **Fig. 10,** lower panel, shows the SDS-PAGE protein profile of these extracts, obtained upon removal of the resin from the mix, again indicating presence of all expected *Synechocystis* proteins

[0103] Lanes 3-6 in **Fig. 10** (upper panel) show the *cpcB*His*Xa*IFN* cell extracts that were removed from the resin upon four consecutive washes with 5 mM imidazole and (**Fig. 10,** lower panel) the SDS-PAGE protein profile of these extracts, showing removal of the majority of cellular proteins in the first wash (Fig. 10, lane 3) and the virtual absence of cell proteins (lane 4 to lane 6) in three additional wash steps with 5 mM imidazole.

[0104] Lanes 7-9 in **Fig. 10** (upper panel) show the further removal of bound His-tagged proteins from the *cpcB*His*Xa*IFN* cell extracts. These eluted from the resin upon three consecutive washes with 250 mM imidazole. **Fig. 10** (lower panel) is the SDS-PAGE protein profile of these extracts, showing substantial enrichment in mainly four proteins with apparent molecular weights of ~108, 36, 27, and 17 kD. The majority of these proteins were eluted upon the first application of the 250 mM imidazole (**Fig. 10,** lane 7), as subsequent elution treatments (**Fig. 10,** lanes 8 and 9) produced much lower levels of protein eluent. Western blot analysis with specific anti-IFN antibodies showed strong cross reactions with the 36 and 108 kD protein bands only (**Fig. 11**). The ~17 kD protein was attributed to the CpcA α-subunit of phycocyanin, as it reacted with CpcA-specific antibodies (not shown, but see also below), whereas the 27 kD protein was attributed to the CpcG1 linker polypeptide (Kondo et al. 2005) that helped to bind the CpcA α-subunit to the CpcB*His*Xa*IFN fusion complex, thereby explaining the simultaneous elution of all three proteins from the resin.

Blue coloration of the target proteins

[0105] The blue coloration of the target proteins (Figs. 7 and 10) and the absorbance spectral evidence of Fig. 9A, suggested the presence of bilin in association with the recombinant CpcB*His*Xa*IFN protein. This finding was surprising as *CpcB*fusion* constructs are known to abolish the assembly of the phycocyanin peripheral rods of the phycobilisome (Formighieri and Melis 2015; 2016; Chaves et al. 2017; Betterle and Melis 2018; 2019), leading to the assumption of a CpcB inability to bind bilin. To further test the spectrophotometric suggestion of bilin presence (Fig. 9A), SDS-PAGE analysis of protein extracts from wild type, the *cpcB*His*Xa*IFN* transformant, and the resin column-based 1st eluent proteins of the latter (**Fig. 12A**) were subjected to "zinc-staining" (please see Materials and methods). Zinc-staining is designed to specifically label the open tetrapyrroles that are covalently bound to *Synechocystis* proteins. **Fig. 12B** shows the result of the Zn-staining of proteins in a duplicate gel, as the one shown in Fig. 12A. In the WT, Zn-staining occurred for proteins migrating to ~19 and ~17 kD, attributed to the native CpcB and CpcA phycocyanin subunits. Zn-staining of the total CpcB*His*Xa*IFN transformant cell extract occurred for protein bands migrating to ~36 and ~17 kD, attributed to the CpcB*His*Xa*IFN and the CpcA proteins, respectively. Zn-staining of the first resin eluent (E1) fraction occurred for protein bands migrating to ~108, ~36 and ~17 kD, putatively attributed to a CpcB*His*Xa*IFN trimer, the CpcB*His*Xa*IFN monomer and the CpcA proteins, respectively. These results corroborate the evidence based on spectrophotometry and

Western blot analysis, clearly showing the presence of bilin in association with the CpcB*His*Xa*IFN fusion and residual CpcA proteins.

*nptI*IFN fusion constructs*

[0106] To further evaluated fusion constructs in the expression and accumulation of biopharmaceutical proteins, two different fusion constructs were designed for the transformation of wild type (WT) *Synechocystis*, based on the *nptI* gene serving as the leader sequence in a *nptI*IFN* configuration and through homologous DNA recombination in the *cpc* operon or *glgA1* locus sites (**Fig. 13A**). In such constructs, the NptI protein served as the antibiotic selection marker, in addition to being the leader protein sequence in the fusion construct (Betterle and Melis 2018; 2019). SDS-PAGE profile of *Synechocystis* protein extracts showed absence of IFN from the wild type, as expected (**Fig. 13B**, WT). The *cpcB*His*Xa*IFN* transformant showed the expected accumulation of a protein band migrating to about 36 kD (**Fig. 13B**, cpcB*His*Xa*IFN), whereas two different lines of a transformant expressing the *nptI*His*Xa*IFN* construct in the *cpc* operon locus showed the presence of a 46 kD protein attributed to this fusion. Positive identification of these assignments was offered by the Western blot analysis of duplicate gels as the one shown in **Fig. 13C**, further confirming the relative abundance of the fusion constructs expressed in the different *Synechocystis* genome loci.

*Antiviral activity of the native and CpcB*IFN fusion protein*

[0107] Activity the cyanobacterial recombinant CpcB*His*Xa*IFN protein was compared with that of commercially-available native interferon provided by the PBL Assay Science, Piscataway, NJ, USA (**Fig. 14**). The results showed that 0.0875 ng/mL of CpcB*His*Xa*IFN fusion interferon was needed to cause a 50% inhibition in encephalomyocarditis (EMC) virus infection, whereas the commercial control required 0.002 ng/mL to cause a 50% inhibition in EMC infection. Part of the difference in sensitivity is probably due to the presence of the CpcB leader sequence in the CpcB*His*Xa*IFN fusion protein, which may have slowed the activity of the fusion IFN. This assumption was validated upon measurements with the cyanobacterial recombinant IFN protein, from which the CpcB leader sequence was removed (Xa excision function).

Example 2. Expression of tissue plasminogen activator derivative K2S protein in cyanobacteria.

K2S fusion constructs

[0108] The fusion constructs approach was also implemented with the tissue plasminogen activator derivative K2S protein. The modified ***cpc*** operon with the *cpcB*His*Xa*K2S* construct was coupled with the chloramphenicol (*cmR*) resistance cassette and expressed under the control of the ***cpc*** promoter (**Fig. 15A**). A similar construct was made in which the Factor Xa protease cleavage domain was replaced by the Tobacco Etch Virus (TEV) cysteine protease cleavage site. SDS-PAGE analysis of the total protein content of wild type, *cpcB*His*Xa*K2S*, and *cpcB*His*TEV*K2S* are shown in **Fig. 15B**. A single WT and three independent lines of each the *cpcB*His*Xa*K2S*, and *cpcB*His*TEV*K2S* transformants are shown in this figure. Western blot analysis of the same protein profile was conducted with polyclonal antibodies raised against the CpcA α-subunit of phycocyanin, which also recognize the CpcB β-subunit (**Fig. 15C**). The results clearly show that dominant in the wild type ~19 kD CpcB β-subunits and CpcA α-subunits of phycocyanin are absent in the *cpcB*His*Xa*K2S*, and *cpcB*His*TEV*K2S* transformants. This is consistent with previous results on the protein phenotype of cpcB*fusion transformants, and it serves as evidence that the *cpcB*His*Xa*K2S*, and *cpcB*His*TEV*K2S* transformants have reached a state of transgenic DNA homoplasmy, underlined by the absence of wild type products in the CpcB and CpcA electrophoretic mobility region. The results also show expression of the CpcB*His*Xa*K2S, and CpcB*His*TEV*K2S transgenic proteins, evidenced by the presence of 58.9 kD protein bands in the gels and the corresponding Western blots (**Fig. 15C**).

Example 3. Expression of insulin in cyanobacteria

Insulin fusion construct

[0109] The fusion constructs approach was further implemented with the human pro-insulin protein expression. The modified *cpc* operon with the *cpcB*INS* construct was coupled with the kanamycin (*nptI*) resistance cassette and expressed under the control of the *cpc* promoter (**Fig. 16A**). SDS-PAGE profile analysis of the total protein content of wild type, *cpcB*INS*, and an earlier transformant carrying the β-phellandrene synthase gene (*PHLS*) from lavender were compared (**Fig. 16B**). The results clearly showed that dominant in the wild type ~19 kD CpcB β-subunit and ~17 kD CpcA α-subunit of phycocyanin are absent in the *cpcB*INS*, as they are also absent from the *cpcB*PHLS* transformants. This is consistent with previous results on the protein phenotype of "*cpcB*fusion*" transformants,

and serves as evidence that the *cpcB*INS* transformants have reached a state of transgenic DNA homoplasmy, underscored by the absence of wild type products in the CpcB and CpcA electrophoretic mobility region. The results also showed expression of the CpcB*INS transgenic protein, evidenced by the presence of ~28 kD protein band specifically in the respective gel lanes (**Fig. 16B**, CpcB*INS).

Example 4. Expression of the Tetanus Toxin Fragment C (TTFC) in cyanobacteria

TTFC fusion construt

[0110] The fusion construct approach was also reduced to practice with the over-expression of the Tetanus Toxin Fragment C (TTFC) protein in cyanobacteria. The modified *cpc* operon, in this case with the *cpcB*L7*His*TEV*TTFC* construct, was coupled with the streptomycin (*smR*) resistance cassette and expressed under the control of the *cpc* promoter (**Fig. 17A**). The work compared the SDS-PAGE profile of the total protein content of wild type, the recipient LTV strain (a transformant carrying the isoprene synthase gene from lavender), and the *cpcB*L7*His*TEV*TTFC* fusion construct (**Fig. 17B, left panel**). In this configuration, presence of the His-tag allowed for a subsequent isolation and purification of the fusion protein. The SDS-PAGE Coomassie stain results clearly showed that the dominant in the wild type ~19 kD CpcB β- and ~17 kD CpcA α-subunits of phycocyanin are absent from the TTFC (*cpcB*L7*His*TEV*TTFC)* transformant, as they are also absent from the LTV (*cpcB*L7*TEV*ISPS)* transformant. This is consistent with previous results on the protein phenotype of *"cpcB*fusion"* transformants, and serves as evidence that the *cpcB*L7*His*TEV*TTFC* transformants have reached a state of transgenic DNA homoplasmy, underscored by the absence of wild type products in the CpcB and CpcA electrophoretic mobility region. Importantly, densitometric analysis of the SDS-PAGE Coommassie stain showed that the 72 kD cpcB*L7*His*TEV*TTFC fusion protein accounted for about 28% of the total cell protein. These results were validated by Western blot analysis, probed with specific polyclonal antibodies agains the TTFC polypeptide (**Fig. 17B, right panel**). Noted was the antibody cross reaction with the 72 kD cpcB*L7*His*TEV*TTFC fusion protein, but also with a ~290 kD putative trimeric [cpcB*L7*His*TEV*TTFC]x3 undissolved fusion protein complex, plus some lower molecular size putative proteolysis fragments of the cpcB*L7*His*TEV*TTFC fusion protein.

Example 5. Expression of the Receptor Binding Domain (RBD) of the SARS-CoV-2 virus in cyanobacteria

RBD fusion construt

[0111] The fusion construct approach was also reduced to practice with the over-expression of a viral protein, the Receptor Binding Domain (RBD) of the spike (S) protein from the SARS-CoV-2, whicht causes the coronavirus disease 2019 (COVID-19). Map of the modified ***cpc*** operon expressing the *cpcB*L7*His*TEV*RBD* fusion construct, including a linker of seven amino acids (L7), a Hisx6-tag (His) and the TEV cleavage factor, followed by the Receptor Binding Domain (RBD) of the spike (S1) protein from the SARS-CoV-2 virus is shown in **FIG. 18** (A). SDS-PAGE and Coomassie stain of the protein extracts from the LTV recipient strain (LTV), and a transformant line harboring the cpcB*L7*His*TEV*RBD fusion protein (RBD) are shown in **FIG. 18** (B, left panel). The arrow points to the electrophoretic mobility of the 45 kD RBD fusion protein, which partially overlaps a native *Synechocystis* 44 kD protein. Western blot analysis of the electropheretically-resolved protein profile for the LTV and RBD *Synechocystis* strains, probed with specific polyclonal antibodies agains the leader CpcB protein, showed an antibody cross reaction with the 45 kD cpcB*L7*His*TEV*RBD fusion protein (**FIG. 18 B,** middle panel). Further identification of the 45 kD protein in the RBD sample was achieved by Zinc-stain analysis of the electophretically-separated proteins from *Synechocystis* expressing the LTV and RBD fusion construct phenotypes (**FIG. 18 B,** right panel). Zn-staining is designed to highlight the presence of bilin tetrapyrrole pigments. Note the specific Zn-staining of a band at 45 kD in the RBD expressing transformant, attributed to the presence of the bilin-binding CpcB protein in the cpcB*L7*His*TEV*RBD fusion protein. (A protein band migrating to about 85 kD is also stained with Zn, and is attributed to the bilin-binding CpcB protein in the *cpcB*L7*His*TEV*ISPS* expressing construct, which is larger than the RBD-containing one.)

Summary of Examples

[0112] Eukaryotic transgenes of plant and animal origin are not always expressed to significant levels in cyanobacteria (Desplancq et al. 2005; 2008; Jindou et al. 2014; Formighieri and Melis 2015). Based on these results, the choice of a strong promoter, such as *cpc*, was necessary but not sufficient to provide high levels of terpene synthase expression in cyanobacteria. Previous investigations pointed to the importance of efficient translation

for protein accumulation. This also appears to be the case in the illustrative examples provided above.

[0113] The *cpc* operon promoter controls expression of the abundant phycocyanin subunits and their associated linker polypeptides of the phycobilisome light-harvesting antenna (Fig. 1A). This endogenous strong promoter was employed in an effort to drive heterologous expression of the codon-optimized IFN gene. However, of the three IFN construct configurations (Fig. 1b, 1c, and 1d), only the fusion construct cpcB*Xa*IFN produced substantial amounts of the transgenic IFN protein (Fig. 1d). Earlier real time RT-qPCR analysis compared transcript levels of plant-origin transgenes, under the same different configurations as those depicted in Fig. 1. The analysis revealed that such transgene constructs resulted in about equal rates of transcription and showed comparable steady-state levels of eukaryotic transgene mRNA (Formighieri and Melis 2016). Hence, the rate of transcription does not appear to be the determinant of recombinant protein abundance in this case.

[0114] Protein synthesis was later investigated by analyzing the polyribosomes distribution profile associated with the various transcripts (Formighieri and Melis 2016). A high density of polyribosomes in prokaryotes, such as cyanobacteria, was attributed to a ribosome pileup, when a slower ribosome migration rate on the mRNA causes multiple ribosomes to associate with the same mRNA molecule (Qin and Fredrick 2013). This was observed to be the case for the Fig. 1b- and 1c-type constructs resulting in low transgenic protein accumulation (Formighieri and Melis 2016). Conversely, a low density of polyribosomes is attributed to efficient ribosome migration on the mRNA, resulting in efficient translation and high levels of protein accumulation (Qin and Fredrick 2013). This was observed to be the case for the Fig. 1d-type constructs of high transgenic protein accumulation (Formighieri and Melis 2016).

[0115] It is of interest that elution of the CpcB*His*Xa*IFN protein from the corresponding cell lysates showed a bluish coloration, which was attributed to the binding of the blue bilin to both the CpcB protein in the CpcB*His*Xa*IFN transformant and to the small amounts of the phycocyanin α-subunit present. Both of these apparently carry the tetrapyrrole chromophore, as evidenced by the typical phycocyanin absorbance spectra of these extracts (Fig. 9a) and by the Zn-staining of the proteins (Fig. 12). However, unlike the *in vivo* situation when about equal amounts of CpcB and CpcA are noted (Fig. 5, WT), there

appeared to be no stoichiometry of CpcB*His*Xa*IFN and CpcA in the transformants (Fig. 5, IFN). The role of small amounts of CpcA in stabilizing the CpcB*His*Xa*IFN recombinant protein is not known at present.

MATERIALS AND METHODS

Synechocystis strains, recombinant constructs, and culture conditions.

[0116] The cyanobacterium *Synechocystis* sp. PCC 6803 (*Synechocystis*) was used as the experimental strain in this work and referred to as the wild type (WT). Gene sequences encoding the human interferon α-2 protein (referred to in the Examples as IFN) and human pro-insulin protein, both without the corresponding N-terminal signal peptides, were codon optimized for protein expression in *Synechocystis* using an open software system available on website, idtdna.com/CodonOpt. Gene sequence encoding the tissue plasminogen activator derivative K2S protein (sequence vaialble at www site drugbank.ca/drugs/DB00015) was codon optimized using the same above-cited open software. DNA constructs for *Synechocystis* transformation were synthesized by Biomatik USA (Wilmington, Delaware). Sequences of the DNA constructs are shown in the Supplemental Materials.

[0117] *Synechocystis* transformations were carried out according to established protocols (Eaton-Rye, 2011; Williams, 1988; Lindberg et al., 2010). Wild type and transformants were maintained on BG11 media supplemented with 1% agar, 10 mM TES-NaOH (pH 8.2) and 0.3% sodium thiosulfate. Liquid cultures of BG11 were buffered with 25 mM sodium bicarbonate, pH 8.2, and 25 mM dipotassium hydrogen phosphate, pH 9, and incubated in the light upon slow continuous bubbling with air at 26°C. Transgenic DNA copy homoplasmy in the cells was achieved upon transformant incubation on agar in the presence of increasing concentrations of chloramphenicol (3-25 μg/mL). Growth of the cells was promoted by using a balanced combination of white LED bulbs supplemented with incandescent light to yield a final visible light (PAR) intensity of ~100 μmol photons m^{-2} s^{-1}.

Genomic DNA PCR analysis of Synechocystis transformants.

[0118] Genomic DNA templates were prepared, as previously described (Formighieri and Melis, 2014a). A 20 μL culture aliquot was provided with an equal volume of 100% ethanol followed by brief vortexing. A 200 μL aliquot of a 10 % (w/v) Chelex®100 Resin (BioRad) suspension in water was added to the sample prior to mixing and heating at 98 °C for 10 min to lyse the cells. Following centrifugation at 16,000 *g* for 10 min to pellet cell debris, 5 μL of

the supernatant was used as a genomic DNA template in a 25 µL PCR reaction mixture. Q5® DNA polymerase (New England Biolabs) was used to perform the genomic DNA PCR analyses. Transgenic DNA copy homoplasmy in *Synechocystis* was tested using suitable primers listed in the Supplemental Materials. The genomic DNA location of these primers is indicated in Fig. 1 for the appropriate DNA constructs.

Protein analysis

[0119] Cells in the mid exponential growth phase (OD_{730} ~1) were harvested by centrifugation at 4,000 *g* for 10 min. The pellet was resuspended in a solution buffered with 25 mM Tris-HCl, pH 8.2, also containing a cOmplete™ mini protease inhibitor cocktail (Roche; one 50 mg tablet was added per 50 mL suspension). Cells were broken by passing the suspension through a French press cell at 1,500 psi. A slow speed centrifugation (350 *g* for 3 min) was applied to remove unbroken cells. For protein electrophoretic analysis, sample extracts were solubilized upon incubation for 1 h at room temperature in the presence of 125 mM Tris-HCl, pH 6.8, 3.5% SDS, 10% glycerol, 2 M urea, and 5% β-mercaptoethanol. SDS-PAGE was performed using Mini-PROTEAN TGX precast gels (BIORAD). Densitometric quantification of target proteins was performed using the BIORAD (Hercules, CA) Image Lab software. A subsequent Western blot analysis entailed transfer of the SDS-resolved proteins to a 0.1 µm pore size PVDF membrane (Life Technologies, Carlsbad, CA). Protein transfer to PVDF was followed by protein probing with rabbit-raised CpcA specific polyclonal antibodies (Abbiotec, San Diego, CA), as previously described (Formighieri and Melis, 2015;), or IFN-specific polyclonal antibodies (Abcam, Cambridge, MA).

Recombinant protein purification

[0120] Total cellular extracts (concentration 100 µg dcw mL^{-1}) from wild-type and transformant strains of *Synechocystis* were gently solubilized upon incubation with 1% Triton X-100 at 0 °C for 20 min. Solubilization of the extracts was conducted in an ice-water bath, upon gentle shaking. Following this solubilization treatment, samples were centrifuged at 10,000 *g* for 10 min to remove cell debris and insoluble material. His-Select resin (Sigma, Saint Louis, MO) was employed as a solid phase for protein binding and purification through cobalt affinity chromatography. Manufacturer's instructions were followed for both batch-type and column-based binding and purification. The washing solution was buffered with 20 mM Hepes, pH 7.5, and contained 150 mM NaCl and 10 mM imidazole to help remove non-

target proteins. The elution solution was buffered with 20 mM Hepes, pH 7.5, and contained 150 mM NaCl and 250 mM imidazole to elute target protein from the resin.

Zn-staining

[0121] SDS-PAGE was incubated in 5 mM zinc sulfate for 30 min X (Li et al. 2016). To detect covalent chromophore-binding polypeptides, zinc induced fluorescence was monitored by Chemidoc imaging system (BIORAD), employing UV light as a light source. Loading of total protein extracts was the same as for the Coomassie-stained SDS-PAGE.

Interferon activity

[0122] Viruses replicate by co-opting normal host cell functions, turning cells into viral factories. Interferon protects cells by binding to extracellular receptors activating a cascade of signals that shuts down both de novo protein and DNA synthesis, depriving the invader the means to replicate. This puts the cells into a semi dormant state, preventing the production of new virus. This is most evident in the life cycle of lytic viruses which normally burst or lyse target cells, but fail to do so when cells are in an interferon-induced antiviral state. One can assess interferon activity by visually comparing the number of intact/lysed cells for a particular concentration of interferon added.

[0123] To assess interferon activity, we contracted the services of PBL Assay Science, Piscataway, NJ, USA, a commercial biomedical testing company, to impartially compare a commercially-available interferon against our own cyanobacterially-generated fusion IFN using the cytopathic effect (CPE) assay.

[0124] The PBL test entailed cells that were (1) untreated; (2) incubated with the encephalomyocarditis (EMC) virus alone; (3) pre-incubated with increasing concentrations of commercial interferon (provided by PBL Assay Science, Piscataway, NJ, USA); or (4) pre-incubated with our cyanobacteria-derived interferon at various concentrations of protein ranging from $1x10^{-3}$ to $1x10^{-7}$ µg/mL.

[0125] Samples were titrated in 96 well plates, and protection against the EMC virus was determined in comparison to the virus (no IFN) and cell (no virus) controls. The samples were run in duplicate alongside Human Interferon Alpha (INF-α) in a viral challenge assay using the encephalomyocarditis virus (EMC) on A549 human cells.

[0126] After maturation of the viral cytopathic effect (CPE), the live cells were fixed and stained using a mixture of 2 mL of 4% formaldehyde, 5% glycerol and 0.5% crystal violet stains per well and allowed to sit at for 60 min at room temperature. Plates were then washed 6-times in running water and dried upside down on filter paper. The dye was subsequently solubilized and assayed by absorbance readings at 570 nm.

[0127] All references, including publications, accession numbers, patent applications, and patents, cited herein are hereby incorporated by reference for the purpose for which it is cited to the same extent as if each reference were individually and specifically indicated to be incorporated by reference.

Listing of Refences Cited by in Specificaton by Author, Publication Year

Baier T, Kros D, Feiner RC, Lauersen KJ, Müller KM, Kruse O (2018) Engineered fusion proteins for efficient protein secretion and purification of a human growth factor from the green microalga *Chlamydomonas reinhardtii*. ACS Synth Biol. 7(11):2547-2557. doi: 10.1021/acssynbio.8b00226.

Bentley FK, Melis A (2012) Diffusion-based process for carbon dioxide uptake and isoprene emission in gaseous/aqueous two-phase photobioreactors by photosynthetic microorganisms. Biotech Bioeng 109:100-109 doi:10.1002/bit.23298

Bentley FK, García-Cerdán JG, Chen H-C, Melis A (2013) Paradigm of monoterpene (β-phellandrene) hydrocarbons production via photosynthesis in cyanobacteria. Bioenergy Res 6, 917–929. doi: 10.1007/s12155-013-9325-4

Bentley FK, Zurbriggen A, Melis A (2014) Heterologous expression of the mevalonic acid pathway in cyanobacteria enhances endogenous carbon partitioning to isoprene. Molecular Plant 7:71-86; doi:10.1093/mp/sst134

Betterle N, Melis A (2018) Heterologous leader sequences in fusion constructs enhance expression of geranyl diphosphate synthase and yield of β-phellandrene production in cyanobacteria (*Synechocystis*). ACS Synth Biol 7:912–921

Bis RL, Stauffer TM, Singh SM, Lavoie TB, Krishna M.G. Mallela KMG (2014) High yield soluble bacterial expression and streamlined purification of recombinant human interferon a-2a. Protein Expression and Purification 99, 138–146

Chaves JE, Rueda Romero P, Kirst H, Melis A (2016) Role of isopentenyl-diphosphate isomerase in heterologous cyanobacterial (*Synechocystis*) isoprene production. Photosynth Res 130:517-527. doi:10.1007/s11120-016-0293-3

Chaves JE, Melis A (2018) Biotechnology of cyanobacterial isoprene production. Appl Microbiol Biotechnol 102(15):6451-6458

Chen H-C, Melis A (2013) Marker-free genetic engineering of the chloroplast in the green microalga *Chlamydomonas reinhardtii*. Plant Biotech J. 11, 818–828; DOI: 10.1111/pbi.12073

Clark ED (2001) Protein refolding for industrial processes. Curr Opin Biotechnol 12, 202-207

Coragliotti AT, Beligni MV, Franklin SE, Mayfield SP (2011) Molecular factors affecting the accumulation of recombinant proteins in the Chlamydomonas reinhardtii chloroplast. Mol Biotechnol 48:60–75

Dang B, Mravic M, Hu H, Schmidt N, Mensa B, DeGrado WF (2019) Nat Methods 16(4):319-322. doi: 10.1038/s41592-019-0357-3. Epub 2019 Mar 25.

Demain AL, Vaishna P (2009) Production of recombinant proteins by microbes and higher organisms. Biotechnol. Adv. 27, 297-306

Desplancq D, Rinaldi A-S, Horzer H, Ho Y, Nierengarten H, R. Andrew Atkinson RA, Kieffer B, Weiss E (2005) Combining inducible protein overexpression with NMR-grade triple isotope labeling in the cyanobacterium *Anabaena* sp. PCC 7120. BioTechniques 39, 405-411

Desplancq D, Rinaldi A-S, Horzer H, Ho Y, Nierengarten H, R. Andrew Atkinson RA, Kieffer B, Weiss E (2008) Automated overexpression and isotopic labelling of biologically active oncoproteins in the cyanobacterium *Anabaena* sp. PCC 7120. Biotechnol Appl Biochem 51, 53–61 doi:10.1042/BA20070276

Davies FK, Work VH, Beliaev AS, Posewitz M C (2014) Engineering limonene and bisabolene production in wild type and a glycogen-deficient mutant of *Synechococcus* sp. PCC7002. Front. Bioeng. Biotechnol. 2,21.

Dyo YM, Purton S (2018) The algal chloroplast as a synthetic biology platform for production of therapeutic proteins. Microbiol 164(2):113-121. doi: 10.1099/mic.0.000599.

Englund E, Shabestary K, Hudson EP, Lindberg P (2918) Systematic overexpression study to find target enzymes enhancing production of terpenes in Synechocystis PCC 6803, using isoprene as a model compound. Metab Eng. 49:164-177. doi: 10.1016/j.ymben.2018.07.004. Epub 2018 Jul 17

Formighieri C, Melis A (2014a) Regulation of β-phellandrene synthase gene expression , recombinant protein accumulation, and monoterpene hydrocarbons production in *Synechocystis* transformants. Planta 240, 309–324. doi: 10.1007/s00425-014-2080-8

Gregory JA, Topol AB, Doerner DZ, Mayfield S (2013) Alga-produced cholera toxin-Pfs25 fusion proteins as oral vaccines. Appl Environ Microbiol 79:3917-3925.

Halfmann C, Gu L, Zhou R (2014a) Engineering cyanobacteria for the production of a cyclic hydrocarbon fuel from CO_2 and H_2O. Green Chem. 16, 3175-3185

Halfmann C, Gu L, Gibbons W, Ruanbao Zhou R (2014b) Genetically engineering cyanobacteria to convert CO_2, water, and light into the long-chain hydrocarbon farnesene. Appl Microbiol Biotechnol 98:9869-9877

Hidalgo D, Abdoli -Nasab M, Jalali-Javaran M, Bru-Martinez R, Cusido RM, Corchete P, Palazon J (2017) Biotechnological production of recombinant tissue plasminogen activator protein (reteplase) from transplastomic tobacco cell cultures. Plant Physiol Biochem 118, 130-137. http://dx.doi.org/10.1016/j.plaphy.2017.06.013

Kondo K, Geng X, Katayama M, Ikeuchi M (2005) Distinct roles of CpcG1 and CpcG2 in phycobilisome assembly in the cyanobacterium *Synechocystis* sp. PCC 6803. Photosyn Res 84:269-73

Jindou S, Ito Y, Mito N, Uematsu K, Hosoda A, Tamura H (2014) Engineered platform for bioethylene production by a cyanobacterium expressing a chimeric complex of plant enzymes. *ACS Synth. Biol.* 37, 487-496

Jones CS, Mayfield SP (2013) Steps toward a globally available malaria vaccine: harnessing the potential of algae for future low cost vaccines. Bioengineered 4:164-167

Kosobokova EN, Skrypnik KA, Kosorukov VS (2016) Overview of fusion tags for recombinant proteins. Biochemistry (Mosc). 81(3):187-200. doi: 10.1134/S0006297916030019.

Lin Y, Garvey CJ, Birch D, Corkery RW, Loughlin PC, Scheer H, Willows RD, Chen M (2016) Characterization of red-shifted phycobilisomes isolated from the chlorophyll *f*-containing cyanobacterium *Halomicronema hongdechloris*. Biochim Biophys Acta. 1857, 107-114

Lindberg P, Park S, Melis A (2010) Engineering a platform for photosynthetic isoprene production in cyanobacteria, using *Synechocystis* as the model organism. Metab Eng 12:70–79. doi: 10.1016/j.ymben.2009.10.001

Luo X-G, Tian W-J, Ni M, Jing X-L, Lv L-H, Wang N, Jiang Y, Zhang T-C (2011) Soluble expression of active recombinant human tissue plasminogen activator derivative (K2S) in *Escherichia coli*. Pharmaceutical Biology 49:653-657

Nordt TK, Bode C (200). Thrombolisys: newer trombolytic agents and their role in clinical medicine. Hearth 89, 1358-1362

Parkin J, Cohen B (2001) An overview of the immune system. *Lancet*. 357 (9270): 1777-1789. doi:10.1016/S0140-6736(00)04904-7

Qin D, Fredrick K (2013) Analysis of polysomes from bacteria. Methods Enzymol 530:159–172

Qiu JI, Swartz JR, Georgiou G (1998) Expression of active human tissue-type plasminogen activator in *Escherichia coli*. Appl Environ Microbiol 64:4891–4896

Rasala BA, Mayfield SP (2015) Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. Photosynth Res 123:227–239

Sonksen P, Sonksen J (2000) Insulin: understanding its action in health and disease. British Journal of Anaesthesia 85(1): 69–79. doi:10.1093/bja/85.1.69. PMID 10927996

Stryer L (1995) Biochemistry (Fourth ed.). New York: W.H. Freeman and Company. pp. 773–774. ISBN 0 7167 2009 4

Surzycki R, Greenham K, Kitayama K, Dibal F, Wagner R, Rochaix J-D, Ajam T, Surzycki S (2009) Factors effecting expression of vaccines in microalgae. Biologicals 37:133–138

Tran M, Zhou B, Pettersson PL, Gonzalez MJ, Mayfield SP (2009) Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. Biotechnol Bioeng 104:663-673.

Ungerer J, Tao L, Davis M, Ghirardi M, Maness P-C, Yu J (2012) Sustained photosynthetic conversion of CO_2 to ethylene in recombinant cyanobacterium *Synechocystis* 6803. Energy Environ Sci 5:8998–9006

Vijay D, M Akhtar, MK, Hess WR (2019) Genetic and metabolic advances in the engineering of cyanobacteria. Current Opinion in Biotechnology 59:150-156.

Voet D, Voet JG (2011) *Biochemistry* (4th ed.). New York: Wiley.

Wilson SA, Roberts SC (2012) Recent advances towards development and commercialization of plant cell culture processes for the synthesis of biomolecules. Plant Biotech. J. 10, 249-268.

Xiong W, Morgan JA, Ungerer J, Wang B, Maness P-C, Yu J (2015) The plasticity of cyanobacterial metabolism supports direct CO_2 conversion to ethylene. Nature Plants 1, Article Number 15053.

Youchun Z, Ge W, Kong Y, Zhang C (2003) Cloning, expression and renaturation studies of reteplase. J. Microbiol. Biotechnol. 13 (6), 989-992.

Zhou J, Zhang H, MengH, Zhu Y, Bao G, Zhang Y, Li Y, Ma Y (2014) Discovery of a super-strong promoter enables efficient production of heterologous proteins in cyanobacteria. Scientific Rep 4(1), 4500

Table 1 Quantification of the RbcL and CpcB*IFN fusion proteins as percent of the total *Synechocystis* proteins loaded onto the SDS-PAGE lanes of Fig. 5. RbcL levels were measured to account for ~12.5%±0.5, CpcB*IFN' accounted for 10.2%±0.2, whereas the CpcB*IFN accounted for 11.8%±0.1 of the total cellular proteins.

Protein measured	IFN' 1	IFN' 2	IFN' 3	IFN 1	IFN 2	IFN 3
RbcL	12.1	12.4	13.2	11.9	12.9	12.6
CpcB*IFN	10.4	9.9	10.2	11.8	11.9	11.7

Illustrative sequences.

SEQ ID NO:1 Human interferon alpha-2 (165 amino acids in length)

CDLPQTHSLGSRRTLMLLAQMRKISLFSCLKDRHDFGFPQEEFGNQFQKAETIPVLHE
MIQQIFNLFSTKDSSAAWDETLLDKFYTELYQQLNDLEACVIQGVGVTETPLMKEDSI
LAVRKYFQRITLYLKEKKYSPCAWEVVRAEIMRSFSLSTNLQESLRSKE

SEQ ID NO:2 Human tissue-type plasminogen activator (562 amino acids in length). The signal peptide is underlined.

MDAMKRGLCCVLLLCGAVFVSPSQEIHARFRRGARSYQVICRDEKTQMIYQQHQSW
LRPVLRSNRVEYCWCNSGRAQCHSVPVKSCSEPRCFNGGTCQQALYFSDFVCQCPE
GFAGKCCEIDTRATCYEDQGISYRGTWSTAESGAECTNWNSSALAQKPYSGRRPDAI
RLGLGNHNYCRNPDRDSKPWCYVFKAGKYSSEFCSTPACSEGNSDCYFGNGSAYRG
THSLTESGASCLPWNSMILIGKVYTAQNPSAQALGLGKHNYCRNPDGDAKPWCHVL
KNRRLTWEYCDVPSCSTCGLRQYSQPQFRIKGGLFADIASHPWQAAIFAKHRRSPGE
RFLCGGILISSCWILSAAHCFQERFPPHHLTVILGRTYRVVPGEEEQKFEVEKYIVHKE
FDDDTYDNDIALLQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGKHEA
LSPFYSERLKEAHVRLYPSSRCTSQHLLNRTVTDNMLCAGDTRSGGPQANLHDACQ
GDSGGPLVCLNDGRMTLVGIISWGLGCGQKDVPGVYTKVTNYLDWIRDNMRP

SEQ ID NO:3 Truncated human tissue plasminogen activator (K2S reteplase) amino acid sequence (355 amino acids in length)

SYQGNSDCYFGNGSAYRGTHSLTESGASCLPWNSMILIGKVYTAQNPSAQALGLGK
HNYCRNPDGDAKPWCHVLKNRRLTWEYCDVPSCSTCGLRQYSQPQFRIKGGLFADI
ASHPWQAAIFAKHRRSPGERFLCGGILISSCWILSAAHCFQERFPPHHLTVILGRTYRV
VPGEEEQKFEVEKYIVHKEFDDDTYDNDIALLQLKSDSSRCAQESSVVRTVCLPPAD
LQLPDWTECELSGYGKHEALSPFYSERLKEAHVRLYPSSRCTSQHLLNRTVTDNMLC
AGDTRSGGPQANLHDACQGDSGGPLVCLNDGRMTLVGIISWGLGCGQKDVPGVYT
KVTNYLDWIRDNMRP

SEQ ID NO:4 Human pro-insulin amino acid sequence (86 amino acids in length)

FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPL
ALEGSLQKRGIVEQCCTSICSLYQLENYCN

SEQ ID NO:15 TTFC, Tetanus Toxin Fragment C (451 amino acids in length)

KNLDCWVDNEEDIDVILKKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIH
LVNNESSEVIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYDTNEYSIISSMK
KYSLSIGSGWSVSLKGNNLIWTLKDSAGEVRQITFRDLSDKFNAYLANKWVFITITND
RLSSANLYINGVLMGSAEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALN
PKEIEKLYTSYLSITFLRDFWGNPLRYDTEYYLIPVAYSSKDVQLKNITDYMYLTNAP
SYTNGKLNIYYRRLYSGLKFIIKRYTPNNEIDSFVRSGDFIKLYVSYNNNEHIVGYPKD
GNAFNNLDRILRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDKDASLGLVGT
HNGQIGNDPNRDILIASNWYFNHLKDKTLTCDWYFVPTDEGWTND

SEQ ID NO:16 Receptor Binding Domain (RBD) of the S1-spike protein from the SARS-CoV-2 virus (223 amino acids in length)

RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFST
FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCV
IAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYF
PLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNF

SEQ ID NO:17 The S1-spike protein from the SARS-CoV-2 virus (673 amino acids in length

SQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHV
SGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIK
VCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGN
FKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALH
RSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCT
LKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNC
VADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIA
DYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG
STPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNL
VKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCS
FGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRA
GCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRAR

SEQ ID NO:18 CtxB, *Cholera* Toxin B (103 amino acids in length)

TPQNITDLCAEYHNTQIHTLNDKIFSYTESLAGKREMAIITFKNGATFQVEVPGSQHID
SQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAISMAN

<u>Illustrative expression construct sequences</u>

1. *cpc_us…optIFN-cmR…cpcA* construct (see, Fig. 1, panel B)

CTCGAG - XhoI DNA restriction site
AGATCT - BglII DNA restriction site
GGATCC - BamHI DNA restriction site

Lower case - *cpc* upstream
5' RECOMBINATION

UPPER CASE - Codon-optimized human interferon (501nt)
lower case - intergenic sequence in construct
lower case bold - *cmR*
<u>lower case underlined</u> - Transcription terminator

3' RECOMBINATION
lower case italics - *cpcB-cpcA* intergenic sequence
lower case bold - *cpcA* (partial)

SEQ ID NO:5 ***cpc_us…optIFN-cmR…cpcA*** **(2336nt) nucleic acid sequence**

CTCGAG*taggctgtggttccctaggcaacagtcttccctaccccactggaaactaaaaaaaacgagaaaagttcgcaccgaa*
catcaattgcataatttttagccctaaaaacataagctgaacgaaactggttgtcttcccttcccaatccaggacaatctgagaatcccc
tgcaacattacttaacaaaaaagcaggaataaaattaacaagatgtaacagacataagtcccatcaccgttgtataaagttaact
gtgggattgcaaaagcattcaagcctaggcgctgagctgtttgagcatcccggtggcccttgtcgctgcctccgtgtttctccctggat
ttatttaggtaatatctctcataaatccccgggtagttaacgaaagttaatggagatcagtaacaataactctagggtcattactttgg
*actccctcagtttatccgggggaattgtgtttaagaaaatcccaactcataaagtcaagtaggagattaattca*ATGTGTGA
CTTGCCTCAGACGCATTCTTTGGGAAGCCGACGCACACTGATGCTGCTCGCCCAA
ATGCGCCGGATCTCCTTATTCTCCTGTCTCAAGGATCGGCATGACTTCGGCTTCCC
TCAGGAGGAGTTTGGAAATCAGTTCCAAAAGGCCGAAACCATTCCGGTCCTCCAT
GAAATGATTCAACAGATCTTTAACTTATTCAGTACCAAAGACAGCAGTGCGGCCT
GGGACGAAACATTACTCGATAAATTCTACACGGAATTATACCAACAGTTGAACG
ACTTAGAAGCCTGTGTAATCCAAGGTGTTGGTGTCACTGAGACTCCATTAATGAA
AGAAGACTCTATTCTGGCCGTCCGCAAGTATTTCCAGCGAATCACACTGTATTTG
AAAGAGAAAAAGTATTCTCCGTGTGCGTGGGAGGTAGTACGGGCTGAAATCATG
CGGTCCTTCTCTTTAAGCACAAACCTCCAGGAATCTCTGCGCTCCAAAGAATGAA
GATCTgcggccgcgttgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtatttttgagtta
tcgagattttcaggagctaaggaagctaaa**atggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgta**
aagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagacc
gtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccggaattccgtatgg
caatgaaagacggtgagctggtgatatgggatagtgttcaccttgttacaccgttttccatgagcaaactgaaacgttttcatc
gctctggagtgaataccacgacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggc
ctatttccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggcc
aatatggacaacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgatt
caggttcatcatgccgtctgtgatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggc
ggggcgtaa<u>ttttttaaggcagttattggtgcccttaaacgcctgg</u>*GGATCCtctggttattttaaaaaccaacttactcaggttcc*
*atacccgagaaaatccagcttaaagctgacatatctaggaaaatttcacattctaacgggagataccagaaca***atgaaaaccc**
ctttaactgaagccgtttccaccgctgactctcaaggtcgctttctgagcagcaccgaattgcaaattgctttcggtcgtctacgt
caagctaatgctggtttgcaagccgctaaagctctgaccgacaatgcccagagcttggtaaatggtgctgcccaagccgtttat
aacaaattcccctacaccacccaaacccaaggcaacaactttgctgcggatcaacggggtaaagacaagtgtgccccgggac
atcggctactacctccgcatcgttacctactgcttagttgctggtggtaccggtcctttggatgagtacttgatcgccggtattgat
gaaatcaaccgcaccttgacctctccccagctggtatgttCTCGAG

2. cpcB…optIFN-cmR…cpcA construct (see, Fig. 1, panel C):

CTCGAG	- XhoI DNA restriction site
AGATCT	- BglII DNA restriction site
GGATCC	- BamHI DNA restriction site
Lower case	- partial *cpcB*
UPPER CASE	- intergenic sequence *cpcB-cpcA*
5' RECOMBINATION	
UPPER CASE	- Codon-optimized human interferon (501nt)
lower case	- intergenic sequence
lower case bold	- *cmR*
lower case underlined	- Transcription terminator
3' RECOMBINATION	
lower case italics	- *cpcB-cpcA* intergenic sequence
lower case bold	- *cpcA* (partial)

SEQ ID NO:6 cpcB…optIFN-cmR…cpcA (2340nt) nucleic acid sequence

CTCGAGccgcatcaccggtaatgcttccgctatcgtttccaacgctgctcgtgctttgttcgccgaacagccccaattaatccaacc
cggtggaaacgcctacaccagccgtcgtatggctgcttgtttgcgtgacatggaaatcatcctccgctatgttacctacgcaaccttcac
cggcgacgcttccgttctagaagatcgttgcttgaacggtctccgtgaaacctacgttgccctgggtgttcccggtgcttccgtagctgct
ggcgttcaaaaaatgaaagaagctgccctggacatcgttaacgatcccaatggcatcacccgtggtgattgcagtgctatcgttgctga
aatcgctggttacttcgaccgcgccgctgctgccgtagcctag**TCTGGTTATTTTAAAAACCAACTTTAC**
TCAGGTTCCATACCCGAGAAAATCCAGCTTAAAGCTGACATATCTAGGAAAA
TTTTCACATTCTAACGGGAGATACCAGAACAATGTGTGACTTGCCTCAGACGC
ATTCTTTGGGAAGCCGACGCACACTGATGCTGCTCGCCCAAATGCGCCGGATCTC
CTTATTCTCCTGTCTCAAGGATCGGCATGACTTCGGCTTCCCTCAGGAGGAGTTTG
GAAATCAGTTCCAAAAGGCCGAAACCATTCCGGTCCTCCATGAAATGATTCAAC
AGATCTTTAACTTATTCAGTACCAAAGACAGCAGTGCGGCCTGGGACGAAACATT
ACTCGATAAATTCTACACGGAATTATACCAACAGTTGAACGACTTAGAAGCCTGT
GTAATCCAAGGTGTTGGTGTCACTGAGACTCCATTAATGAAAGAAGACTCTATTC
TGGCCGTCCGCAAGTATTTCCAGCGAATCACACTGTATTTGAAAGAGAAAAAGT
ATTCTCCGTGTGCGTGGGAGGTAGTACGGGCTGAAATCATGCGGTCCTTCTCTTT
AAGCACAAACCTCCAGGAATCTCTGCGCTCCAAAGAATGAAGATCTgcggccgcgttga
tcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtatttttgagttatcgagattttcaggagctaagg
aagctaaa**atggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatttc**
agtcagttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaaagaaaaataagcaca
agttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccggaattccgtatggcaatgaaagacggtgagctg

gtgatatgggatagtgttcaccattgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaataccacga
cgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattg
agaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgccc
ccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtctgtg
atggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaa<u>ttttttaaggcagt</u>
<u>tattggtgcccttaaacgcctgg</u>*GGATCCtctggttattttaaaaaccaactttactcaggttccataccgagaaaatccagctta*
*aagctgacatatctaggaaaattttcacattctaacgggagataccagaaca***atgaaaaccccttttaactgaagccgtttccacc**
gctgactctcaaggtcgctttctgagcagcaccgaattgcaaaattgctttcggtcgtctacgtcaagctaatgctggtttgcaagc
cgctaaagctctgaccgacaatgcccagagcttggtaaatggtgctgcccaagccgtttataacaaattcccctacaccacccca
aacccaaggcaacaactttgctgcggatcaacggggtaaagacaagtgtgcccgggacatcggctactacctccgcatcgtt
acctactgcttagttgctggtggtaccggtcctttggatgagtacttgatcgccggtattgatgaaatcaaccgcaccttgacct
ctccccagctggtatgttCTCGAG

3. *cpc us … cpcB**Xa**IFN-cmR…cpcA* construct (see, Fig. 1, panel D):

CTCGAG - XhoI DNA restriction site
AGATCT - BglII DNA restriction site
GGATCC - BamHI DNA restriction site

Lower case - *cpcB*
5' RECOMBINATION

UPPER CASE - Factor Xa cleavage site (IEGR)
UPPER CASE - codon-optimized human interferon
lower case - intergenic sequence
lower case bold - cmR
lower case underline - Transcription terminator

3' RECOMBINATION
lower case italics - *cpcB-cpcA* intergenic sequence
lower case bold - *cpcA* (partial)

SEQ ID NO:7 ***cpc us … cpcB**Xa*IFN-cmR…cpcA* (2361 nt) nucleic aic squence**

CTCGAGatgttcgacgtattcactcgggttgtttcccaagctgatgctcgcggcgagtacctctctggttctcagttagatgctttgag
cgctaccgttgctgaaggcaacaaacggattgattctgttaaccgcatcaccggtaatgcttccgctatcgtttccaacgctgctcgtgct
ttgttcgccgaacagccccaattaatccaaccccggtggaaacgcctacaccagccgtcgtatggctgcttgtttgcgtgacatggaaat
catcctccgctatgttacctacgcaaccttcaccggcgacgcttccgttctagaagatcgttgcttgaacggtctccgtgaaacctacgtt
gccctgggtgttcccggtgcttccgtagctgctggcgttcaaaaaatgaaagaagctgccctggacatcgttaacgatcccaatggcat
cacccgtggtgattgcagtgctatcgttgctgaaatcgctggttacttcgaccgcgccgctgctgccgtagccATCGAAGGGC
GATGTGACTTGCCTCAGACGCATTCTTTGGGAAGCCGACGCACACTGATGCTGCT
CGCCCAAATGCGCCGGATCTCCTTATTCTCCTGTCTCAAGGATCGGCATGACTTC
GGCTTCCCTCAGGAGGAGTTTGGAAATCAGTTCCAAAAGGCCGAAACCATTCCG
GTCCTCCATGAAATGATTCAACAGATCTTTAACTTATTCAGTACCAAAGACAGCA
GTGCGGCCTGGGACGAAACATTACTCGATAAATTCTACACGGAATTATACCAAC
AGTTGAACGACTTAGAAGCCTGTGTAATCCAAGGTGTTGGTGTCACTGAGACTCC
ATTAATGAAAGAAGACTCTATTCTGGCCGTCCGCAAGTATTTCCAGCGAATCACA
CTGTATTTGAAAGAGAAAAAGTATTCTCCGTGTGCGTGGGAGGTAGTACGGGCT
GAAATCATGCGGTCCTTCTCTTTAAGCACAAACCTCCAGGAATCTCTGCGCTCCA
AAGAATGAAGATCTgcggccgcgttgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgg
gcgtatttttgagttatcgagattttcaggagctaaggaagctaaa**atggagaaaaaaatcactggatataccaccgttgatatatc**
ccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacg
gccttttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatcc
ggaattccgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaact

gaaacgttttcatcgctctggagtgaataccacgacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacg
gtgaaaaacctggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttg
atttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctga
tgccgctggcgattcaggttcatcatgccgtctgtgatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcga
tgagtggcagggcggggcgtaattttttttaaggcagttattggtgcccttaaacgcctggGGATCCtctggttattttaaaaacca
actttactcaggttccatacccgagaaaatccagcttaaagctgacatatctaggaaaatttttcacattctaacgggagataccaga
acaatgaaaaccccttttaactgaagccgtttccaccgctgactctcaaggtcgctttctgagcagcaccgaattgcaaattgctt
tcggtcgtctacgtcaagctaatgctggtttgcaagccgctaaagctctgaccgacaatgcccagagcttggtaaatggtgctg
cccaagccgtttataacaaattcccctacaccacccaaacccaaggcaacaactttgctgcggatcaacggggtaaagacaa
gtgtgcccgggacatcggctactacctccgcatcgttacctactgcttagttgctggtggtaccggtcctttggatgagtacttga
tcgccggtattgatgaaatcaaccgcaccttgacctctccccagctggtatgttCTCGAG

4. *cpc us ... cpcB*Xa*IFN'-cmR... cpcA* construct (see, Fig. 5):

CTCGAG - XhoI DNA restriction site
AGATCT - BglII DNA restriction site
GGATCC - BamHI DNA restriction site

Lower case - *cpcB*
5' RECOMBINATION

UPPER CASE - Factor Xa cleavage site (IEGR)
UPPER CASE - Native human interferon
lower case - intergenic sequence in Cinzia's construct
lower case bold - *cmR*
lower case underlined - Transcription terminator

3' RECOMBINATION
lower case italics - *cpcB-cpcA* intergenic sequence
lower case bold - *cpcA* (partial)

SEQ ID NO:8 ***cpc us ... cpcB*Xa*IFN'-cmR...cpcA* (2361 nt) nucleic acid sequence**

CTCGAGatgttcgacgtattcactcgggttgtttcccaagctgatgctcgcggcgagtacctctctggttctcagttagatgctttgag
cgctaccgttgctgaaggcaacaaacggattgattctgttaaccgcatcaccggtaatgcttccgctatcgtttccaacgctgctcgtgct
ttgttcgccgaacagccccaattaatccaaccccggtggaaacgcctacaccagccgtcgtatggctgcttgtttgcgtgacatggaaat
catcctccgctatgttacctacgcaaccttcaccggcgacgcttccgttctagaagatcgttgcttgaacggtctccgtgaaacctacgtt
gccctgggtgttcccggtgcttccgtagctgctggcgttcaaaaaatgaaagaagctgccctggacatcgttaacgatcccaatggcat
cacccgtggtgattgcagtgctatcgttgctgaaatcgctggttacttcgaccgcgccgctgctgccgtagccATCGAAGGGC
GATGTGATCTGCCTCAAACCCACAGCCTGGGTAGCAGGAGGACCTTGATGCTCCT
GGCACAGATGAGGAGAATCTCTCTTTTCTCCTGCTTGAAGGACAGACATGACTTT
GGATTTCCCCAGGAGGAGTTTGGCAACCAGTTCCAAAAGGCTGAAACCATCCCT
GTCCTCCATGAGATGATCCAGCAGATCTTCAATCTCTTCAGCACAAAGGACTCAT
CTGCTGCTTGGGATGAGACCCTCCTAGACAAATTCTACACTGAACTCTACCAGCA
GCTGAATGACCTGGAAGCCTGTGTGATACAGGGGGTGGGGGTGACAGAGACTCC
CCTGATGAAGGAGGACTCCATTCTGGCTGTGAGGAAATACTTCCAAAGAATCACT
CTCTATCTGAAAGAGAAGAAATACAGCCCTTGTGCCTGGGAGGTTGTCAGAGCA
GAAATCATGAGATCTTTTTCTTTGTCAACAAACTTGCAAGAAAGTTTAAGAAGTA
AGGAATGAAGATCTgcggccgcgttgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgg
gcgtatttttgagttatcgagattttcaggagctaaggaagctaaa**atggagaaaaaaatcactggatataccaccgttgatatatc**
ccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacg
gccttttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatcc
ggaattccgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaact

gaaacgttttcatcgctctggagtgaataccacgacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacg
gtgaaaaacctggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttg
atttaaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctga
tgccgctggcgattcaggttcatcatgccgtctgtgatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcga
tgagtggcagggcggggcgtaatttttttaaggcagttattggtgcccttaaacgcctggGGATCCtctggttattttaaaaacca
actttactcaggttccatacccgagaaaatccagcttaaagctgacatatctaggaaaatttttcacattctaacgggagataccaga
acaatgaaaaccccttttaactgaagccgtttccaccgctgactctcaaggtcgctttctgagcagcaccgaattgcaaattgctt
tcggtcgtctacgtcaagctaatgctggtttgcaagccgctaaagctctgaccgacaatgcccagagcttggtaaatggtgctg
cccaagccgtttataacaaattcccctacaccacccaaacccaaggcaacaactttgctgcggatcaacggggtaaagacaa
gtgtgcccgggacatcggctactacctccgcatcgttacctactgcttagttgctggtggtaccggtcctttggatgagtacttga
tcgccggtattgatgaaatcaaccgcaccttgacctctccccagctggtatgttCTCGAG

5. *cpc us ... cpcB*HisTag*Xa*IFN-cmR... cpcA* construct (see, Fig. 6):

CTCGAG - XhoI DNA restriction site
AGATCT - BglII DNA restriction site
GGATCC - BamHI DNA restriction site

Lower case - *cpcB*
5' RECOMBINATION

UPPER CASE - Histag 6x
UPPER CASE - Factor Xa cleavage site (IEGR)
UPPER CASE - synechocystis-optimized human interferon
lower case - intergenic sequence in Cinzia's construct
lower case bold - cmR
lower case underlined - Transcription terminator

3' RECOMBINATION
lower case italics - *cpcB-cpcA* intergenic sequence
lower case bold - *cpcA* (partial)

SEQ ID NO:9 *cpc us ... cpcB*HisTag*Xa*IFN-cmR...cpcA* (2379 nt) nucleic acid sequence

CTCGAGatgttcgacgtattcactcgggttgtttcccaagctgatgctcgcggcgagtacctctctggttctcagttagatgctttgag
cgctaccgttgctgaaggcaacaaacggattgattctgttaaccgcatcaccggtaatgcttccgctatcgtttccaacgctgctcgtgct
ttgttcgccgaacagccccaattaatccaacccggtggaaacgcctacaccagccgtcgtatggctgcttgtttgcgtgacatggaaat
catcctccgctatgttacctacgcaaccttcaccggcgacgcttccgttctagaagatcgttgcttgaacggtctccgtgaaacctacgtt
gccctgggtgttcccggtgcttccgtagctgctggcgttcaaaaaatgaaagaagctgccctggacatcgttaacgatcccaatggcat
cacccgtggtgattgcagtgctatcgttgctgaaatcgctggttacttcgaccgcgccgctgctgccgtagccCACCATCACC
ATCACCATATCGAAGGGCGATGTGACTTGCCTCAGACGCATTCTTTGGGAAGCCG
ACGCACACTGATGCTGCTCGCCCAAATGCGCCGGATCTCCTTATTCTCCTGTCTCA
AGGATCGGCATGACTTCGGCTTCCCTCAGGAGGAGTTTGGAAATCAGTTCCAAAA
GGCCGAAACCATTCCGGTCCTCCATGAAATGATTCAACAGATCTTTAACTTATTC
AGTACCAAAGACAGCAGTGCGGCCTGGGACGAAACATTACTCGATAAATTCTAC
ACGGAATTATACCAACAGTTGAACGACTTAGAAGCCTGTGTAATCCAAGGTGTTG
GTGTCACTGAGACTCCATTAATGAAAGAAGACTCTATTCTGGCCGTCCGCAAGTA
TTTCCAGCGAATCACACTGTATTTGAAAGAGAAAAAGTATTCTCCGTGTGCGTGG
GAGGTAGTACGGGCTGAAATCATGCGGTCCTTCTCTTTAAGCACAAACCTCCAGG
AATCTCTGCGCTCCAAAGAATGAAGATCTgcggccgcgttgatcggcacgtaagaggttccaactttcacc

ataatgaaataagatcactaccgggcgtattttttgagttatcgagattttcaggagctaaggaagctaaa**atggagaaaaaaaatcact**
ggatataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataacc
agaccgttcagctggatattacggccttttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattctt
gcccgcctgatgaatgctcatccggaattccgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgt
tacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaataccacgacgatttccggcagtttctacacatata
ttcgcaagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaatc
cctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaatattat
acgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtctgtgatggcttccatgtcggcagaatgctt
aatgaattacaacagtactgcgatgagtggcagggcggggcgtaattttttaaggcagttattggtgcccttaaacgcctgg*GG*
ATCCtctggttatttttaaaaaccaacttttactcaggttccataccgagaaaatccagcttaaagctgacatatctaggaaaatttt
*cacattctaacgggagataccagaaca***atgaaaaccccttttaactgaagccgtttccaccgctgactctcaaggtcgctttctga**
gcagcaccgaattgcaaattgctttcggtcgtctacgtcaagctaatgctggtttgcaagccgctaaagctctgaccgacaatg
cccagagcttggtaaatggtgctgcccaagccgtttataacaaattcccctacaccacccaaacccaaggcaacaactttgctg
cggatcaacggggtaaagacaagtgtgcccgggacatcggctactacctccgcatcgttacctactgcttagttgctggtggta
ccggtcctttggatgagtacttgatcgccggtattgatgaaatcaaccgcaccttgacctctcccccagctggtatgttCTCG
AG

6. *cpc-US… nptI*IFN… cpcA+cpc genes-DS* construct (see, Fig. 13):

UPPER CASE - upstream *cpc* operon FLANKING SITE (506 nt)
*nptI**(His$_{6x}$*Xa)**IFN* (acts also as the resistance cassette) (1,341 nt)
lower case underlined - Transcription terminator
UPPER CASE - *cpcB-cpcA* intergenic sequence
UPPER CASE - ***cpcA* gene** FLANKING SITE (517 nt including UPPER CASE intergenic sequence)
CTCGAG - XhoI restriction site
AGATCT - BglII restriction site
GGATCC - BamHI restriction site

SEQ ID NO:10 *cpc-US… nptI*IFN… cpcA+cpc genes-DS* (2420 nt) nucleic acid sequence

CTCGAGGGAAAGTAGGCTGTGGTTCCCTAGGCAACAGTCTTCCCTACCCCACTGG
AAACTAAAAAAACGAGAAAAGTTCGCACCGAACATCAATTGCATAATTTTAGCC
CTAAAACATAAGCTGAACGAAACTGGTTGTCTTCCCTTCCCAATCCAGGACAATC
TGAGAATCCCCTGCAACATTACTTAACAAAAAAGCAGGAATAAAATTAACAAGA
TGTAACAGACATAAGTCCCATCACCGTTGTATAAAGTTAACTGTGGGATTGCAAA
AGCATTCAAGCCTAGGCGCTGAGCTGTTTGAGCATCCCGGTGGCCCTTGTCGCTG
CCTCCGTGTTTCTCCCTGGATTTATTTAGGTAATATCTCTCATAAATCCCCGGGTA
GTTAACGAAAGTTAATGGAGATCAGTAACAATAACTCTAGGGTCATTACTTTGGA
CTCCCTCAGTTTATCCGGGGGAATTGTGTTTAAGAAAATCCCAACTCATAAAGTC
AAGTAGGAGATTAATTCA*atgagtcacatccagagagaaactagttgttcccgacctcgtttgaatagcaatatgg*
atgcagatctgtacggatataaatgggcgcgagataacgtaggccaatctggggccactatttatcggttatatggcaaaccagat
gctcccgaactgtttctcaaacatggcaaagggtctgtggccaatgatgttaccgatgaaatggtgcggttgaactggttgacaga
atttatgcccctcccgaccatcaaacattttatcaggactccagacgatgcatggctattaactacggccattcctgggaaaactgcc
tttcaggtgttggaagaatatcccgattctggtgagaatatcgtcgatgcgttagcggttttttctaagacgtctacatagcattcccgttt
gcaattgtcccttaattcggaccgggtgttccgcttggcgcaggctcagtcccggatgaataacggtttggtagatgcctcggacttt
gatgatgaacggaacggctggcccgttgaacaggtttggaaagagatgcataagctgctgccctttctccccgacagcgttgttac
tcatggagatttttctctcgataatctgatttcgacgaaggcaagctaatggctgtatcgatgtgggacgggtagggattgcggac
cggtatcaagacctagcaatttgtggaactgcctaggtgaatttcccccagcctacaaaaacggctgtttcaaaaatacggaatc
*gataatcccgacatgaacaaattacaatttcatctgatgctagatgagttctttc*accatcaccatcaccat**atcgaagggcgatgtg**
acttgcctcagacgcattctttgggaagccgacgcacactgatgctgctgcccaaatgcgccggatctccttattctcctgtctc
aaggatcggcatgacttcggcttcctcaggaggagtttggaaatcagttccaaaaggccgaaaccattccggtcctccatga
aatgattcaacagatctttaacttattcagtaccaaagacagcagtgcggcctgggacgaaacattactcgataaattctacac
ggaattataccaacagttgaacgacttagaagcctgtgtaatccaaggtgttggtgtcactgagactccattaatgaaagaag

actctattctggccgtccgcaagtatttccagcgaatcacactgtatttgaaagagaaaaagtattctccgtgtgcgtgggaggt
agtacgggctgaaatcatgcggtccttctctttaagcacaaacctccaggaatctctgcgctccaaagaatgattttttaaggca
gttattggtgcccttaaacgcctggg**GATCCTCTGGTTATTTTAAAAACCAACTTTACTCAGGTTC**
CATACCCGAGAAAATCCAGCTTAAAGCTGACATATCTAGGAAAATTTTCACATTC
TAACGGGAGATACCAGAACAATGAAAACCCCTTTAACTGAAGCCGTTTCCACC
GCTGACTCTCAAGGTCGCTTTCTGAGCAGCACCGAATTGCAAATTGCTTTCG
GTCGTCTACGTCAAGCTAATGCTGGTTTGCAAGCCGCTAAAGCTCTGACCGA
CAATGCCCAGAGCTTGGTAAATGGTGCTGCCCAAGCCGTTTATAACAAATTC
CCCTACACCACCCAAACCCAAGGCAACAACTTTGCTGCGGATCAACGGGGT
AAAGACAAGTGTGCCCGGGACATCGGCTACTACCTCCGCATCGTTACCTACT
GCTTAGTTGCTGGTGGTACCGGTCCTTTGGATGAGTACTTGATCGCCGGTAT
TGATGAAATCAACCGCACCTTTGACCTCTCCCCCAGCTGGTATGTTGAAGCT
CTGAAATACATCTCGAG

7. *glgA1-US…* P_{TRC}*-nptI*IFN…glgA1-DS* construct: (see, FIG. 19, panel A)

UPSTREAM *glgA1*	FLANKING SITE (540 nt)
UPPER CASE	**lower case combination P_{TRC} (101 nt)**
*nptI**(His$_{6x}$*Xa)**IFN*	(acts also as the resistance cassette) (1,341 nt)
UPPER CASE T*psbA2*	**(terminator *psbA2*)** (193)
DOWNSTREAM *glgA1*	FLANKING SITE (512)
CTCGAG	- XhoI restriction site
GGATCC	- BamHI restriction site

SEQ ID NO:11 *glgA1-US…PTRC-nptI*IFN…glgA1-DS* (2705 nt) nucleic acid sequence

CTCGAGGCCATGTCCCAAATTCTTGATCCCATCCCCAACAACCAGCCATCAGCCT
TATTCTGTTGCTACGTCAATGCCACCAATCAAATCCAAGTGGCCCGCATTACCAA
TGTCCCTAATTGGTATTTTGAAAGAGTTGTGTTCCCTGGTCAACGGTTAGTATTTG
AGGCAGTGCCCAGCGCTCAGTTAGAAATTCATACTGGCATGATGGCCAGCTCGAT
TATTTCGGACACCATTCCCTGCGAACAACTGAGTATTGATCCCGACGGATTAGCA
GCGGGCGGTTTCATCTCTCCAGAAAAAGAACACGAGTCCGAGGATATGACTTCC
CAATCCTTAGTGGCTTAGCAATGAATTAATGAATTGGAATACTTAGGCCATGCCA
CCGGCCGGCAATGGATAGTCCACGGACAAAGCACTAAGAAAAAGGTATAGGGAT
GGAAAGCAGAAACTGTTAATTACTCTCTCCGATGGGTAACCACCACCGTCATATA
ATTGAGCGGAAAGTATGGCAACCAGGCCCTGAACTCAATTAGTGGAATAACG**CG**
GTCCTGCAGGATTCTGAAATGAGCTGTTGACAATTAATCATCCGGCTCGTAT
AAtgtgtggaAATTGTGAGCGGATAACAATTAGGAGGTTAATTAACA*atgagtcacatcc*
agagagaaactagttgttcccgacctcgtttgaatagcaatatggatgcagatctgtacggatataaatgggcgcgagataacgta
ggccaatctggggccactatttatcggttatatggcaaaccagatgctcccgaactgtttctcaaacatggcaaagggtctgtggcc
aatgatgttaccgatgaaatggtgcggttgaactggttgacagaatttatgcccctcccgaccatcaaacattttatcaggactccag
acgatgcatggctattaactacggccattcctgggaaaactgcctttcaggtgttggaagaatatcccgattctggtgagaatatcgt
cgatgcgttagcggttttttctaagacgtctacatagcattcccgtttgcaattgtccctttaattcggaccgggtgttccgcttggcgcag
gctcagtcccggatgaataacggtttggtagatgcctcggactttgatgatgaacggaacggctggcccgttgaacaggtttggaa
agagatgcataagctgctgccctctcccccgacagcgttgttactcatggagatttttctctcgataatctgattttcgacgaaggca
agctaattggctgtatcgatgtgggacgggtagggattgcggaccggtatcaagacctagcaatttttgtggaactgcctaggtgaat
tttccccagcctacaaaaacggctgtttcaaaaatacggaatcgataatcccgacatgaacaaattacaatttcatctgatgctag
*atgagttctttc*accatcaccatcaccatatcgaagggcg*atgtgacttgcctcagacgcattctttgggaagccgacgcacactgat*
gctgctcgcccaaatgcgccggatctccttattctcctgtctcaaggatcggcatgacttcggcttccctcaggaggagtttggaaat
cagttccaaaaggccgaaaccattccggtcctccatgaaatgattcaacagatctttaacttattcagtaccaaagacagcagtgc
ggcctgggacgaaacattactcgataaattctacacggaattataccaacagttgaacgacttagaagcctgtgtaatccaaggtg
ttggtgtcactgagactccattaatgaaagaagactctattctggccgtccgcaagtatttccagcgaatcacactgtatttgaaaga

gaaaaagtattctccgtgtgcgtgggaggtagtacgggctgaaatcatgcggtccttctctttaagcacaaacctccaggaatctct
*gcgctccaaagaatga*GGATCC**TCCTTGGTGTAATGCCAACTGAATAATCTGCAAATT**
GCACTCTCCTTCAATGGGGGGTGCTTTTTGCTTGACTGAGTAATCTTCTGAT
TGCTGATCTTGATTGCCATCGATCGCCGGGGAGTCCGGGGCAGTTACCATT
AGAGAGTCTAGAGAATTAATCCATCTTCGATAGAGGAATTATGGGGGAAGA
ACCCTAGGCAATTGATGGCCATGCGTTATGGCTGTATCCCCATTGTGCGGCGGAC
AGGGGGTTTGGTGGATACGGTATCCTTCTACGATCCTATCAATGAAGCCGGCACC
GGCTATTGCTTTGACCGTTATGAACCCCTGGATTGCTTTACGGCCATGGTGCGGG
CCTGGGAGGGTTTCCGTTTCAAGGCAGATTGGCAAAAATTACAGCAACGGGCCA
TGCGGGCAGACTTTAGTTGGTACCGTTCCGCCGGGGAATATATCAAAGTTTATAA
GGGCGTGGTGGGGAAACCGGAGGAATTAAGCCCCATGGAAGAGGAAAAAATCG
CTGAGTTAACTGCTTCCTATCGCTAACAATCTCCCGGCAGTGAAGTAAAATCCTG
AACCCTAATCCCGCTCCACTGCCGACCCCAATTCTCCTTGCCTAGGCAAATTTGA
AAATTTTTTCTGATCAATGCTTGTGGTGAAGCAAAAGCTATGTTAACGTTATAAA
TCGTGCCAATGAAGCACAACGGGCTCGAG

8. *cpc us ... cpcB*HisTag*Xa*optK2S-cmR... cpcA* construct (see, Fig. 15)

5' RECOMBINATION

CTCGAG - XhoI DNA restriction site
Lower case - CpcB
UPPER CASE - Histag 6x
UPPER CASE - Factor Xa cleavage site (IEGR)
UPPER CASE - *Synechocystis*-optimized K2S (without first methionine, plus stop codon)
AGATCT - BglII DNA restriction site
GGATCC - BamHI DNA restriction site
lower case - intergenic sequence in Cinzia's construct
Lower case bold - cmR
lower case underlined - Transcription terminator
GGATCC - BamHI DNA restriction site
lower case italics - CpcB-CpcA intergenic sequence
lower case bold - CpcA (partial)
CTCGAG - XhoI DNA restriction site

3' RECOMBINATION

SEQ ID NO:12 *cpc us ... cpcB*HisTag*Xa*optK2S-cmR...cpcA* (2949 nt) nucleic acid sequence

CTCGAGatgttcgacgtattcactcgggttgtttcccaagctgatgctcgcggcgagtacctctctggttctcagttagatgctttgag
cgctaccgttgctgaaggcaacaaacggattgattctgttaaccgcatcaccggtaatgcttccgctatcgtttccaacgctgctcgtgct
ttgttcgccgaacagccccaattaatccaacccggtggaaacgcctacaccagccgtcgtatggctgcttgtttgcgtgacatggaaat
catcctccgctatgttacctacgcaaccttcaccggcgacgcttccgttctagaagatcgttgcttgaacggtctccgtgaaacctacgtt
gccctgggtgttcccggtgcttccgtagctgctggcgttcaaaaaatgaaagaagctgccctggacatcgttaacgatcccaatggcat
cacccgtggtgattgcagtgctatcgttgctgaaatcgctggttacttcgaccgcgccgctgctgccgtagccCACCATCACC
ATCACCATATCGAAGGGCGATCCTATCAAGGCAATTCCGATTGTTATTTTGGCAA
TGGCTCCGCCTATCGGGGCACCCATTCCTTGACCGAATCCGGCGCCTCCTGTTTG
CCCTGGAATTCCATGATTTTGATTGGCAAAGTGTATACCGCCCAAAATCCCTCCG
CCCAAGCCTTGGGCTTGGGCAAACATAATTATTGTCGGAATCCCGATGGCGATGC
CAAACCCTGGTGTCATGTGTTGAAGAATCGGCGGTTGACCTGGGAATATTGTGAT
GTGCCCTCCTGTTCCACCTGTGGCTTGCGGCAATATTCCCAACCCCAATTTCGGAT
TAAAGGCGGCTTGTTTGCCGATATTGCCTCCCATCCCTGGCAAGCCGCCATCTTT
GCCAAACATCGGCGGTCTCCCGGCGAACGGTTCTTGTGTGGCGGCATTTTGATTT
CCTCCTGTTGGATTTTGTCCGCCGCCCATTGTTTTCAAGAACGGTTTCCTCCCCAT
CATTTGACCGTGATTTTGGGCCGGACCTATCGGGTGGTGCCCGGCGAAGAAGAA
CAGAAATTTGAAGTGGAGAAATATATTGTGCATAAAGAATTTGATGATGATACCT
ATGATAATGATATTGCCTTGTTGCAATTGAAATCCGATTCCTCCCGGTGTGCCCA

AGAATCCTCCGTGGTGCGGACCGTGTGTTTGCCTCCCGCCGATTTGCAATTGCCC
GATTGGACCGAATGTGAATTGTCCGGCTATGGCAAACATGAAGCCTTGTCTCCCT
TTTATTCCGAACGGTTGAAAGAAGCCCATGTGCGGTTGTATCCCTCCTCCCGGTG
TACCTCCCAACATTTGTTGAATCGGACCGTGACCGATAATATGTTGTGTGCCGGC
GATACCCGGTCCGGCGGCCCCCAAGCCAATTTGCATGATGCCTGTCAAGGCGATT
CCGGCGGCCCCTTGGTGTGTTTGAATGATGGCCGGATGACCTTGGTGGGCATTAT
TTCCTGGGGCTTGGGCTGTGGCCAGAAAGATGTGCCCGGCGTGTATACCAAAGTG
ACCAATTATTTGGATTGGATTCGGGATAATATGCGGCCCTAAAGATCTgcggccgcgtt
gatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtatttttgagttatcgagatttcaggagctaa
ggaagctaaa**atggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatt**
tcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaaagaaaaataagca
caagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccggaattccgtatggcaatgaaagacggtgagc
tggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaataccac
gacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttat
tgagaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgc
ccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtctgt
gatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaatttttttaaggca
gttattggtgcccttaaacgcctgg*GGATCCtctggttattttaaaaaccaacttactcaggttccataccgagaaaatccagct*
*taaagctgacatatctaggaaaattttcacattctaacgggagataccagaaca***atgaaaaccccttttaactgaagccgtttcca**
ccgctgactctcaaggtcgctttctgagcagcaccgaattgcaaattgctttcggtcgtctacgtcaagctaatgctggtttgcaa
gccgctaaagctctgaccgacaatgcccagagcttggtaaatggtgctgcccaagccgtttataacaaattcccctacaccacc
caaacccaaggcaacaactttgctgcggatcaacggggtaaagacaagtgtgcccgggacatcggctactacctccgcatcg
ttacctactgcttagttgctggtggtaccggtcctttggatgagtacttgatcgccggtattgatgaaatcaaccgcaccttgac
ctctccccagctggtatgttCTCGAG

9. *cpc-US... nptI*HisTag*Xa*K2S... cpcA+cpc genes-DS* construct: (see, Fig. 19, panel B)

UPPER CASE - upstream *cpc* operon FLANKING SITE (506 nt)
*nptI**(His$_{6x}$*Xa)**K2S* (acts also as the resistance cassette) (2,478 nt)
lower case underlined - Transcription terminator
UPPER CASE - *cpcB-cpcA* intergenic sequence **plus *cpcA* gene** FLANKING SITE (517 nt)

CTCGAG - XhoI restriction site
AGATCT - BglII restriction site
GGATCC - BamHI restriction site

SEQ ID NO:13 *cpc-US... nptI*HisTag*Xa*K2S... cpcA+cpc genes-DS* (2990 nt) nucleic acid sequence

CTCGAGGGAAAGTAGGCTGTGGTTCCCTAGGCAACAGTCTTCCCTACCCCACTGG
AAACTAAAAAAACGAGAAAAGTTCGCACCGAACATCAATTGCATAATTTTAGCC
CTAAAACATAAGCTGAACGAAACTGGTTGTCTTCCCTTCCCAATCCAGGACAATC
TGAGAATCCCCTGCAACATTACTTAACAAAAAAGCAGGAATAAAATTAACAAGA
TGTAACAGACATAAGTCCCATCACCGTTGTATAAAGTTAACTGTGGGATTGCAAA
AGCATTCAAGCCTAGGCGCTGAGCTGTTTGAGCATCCCGGTGGCCCTTGTCGCTG
CCTCCGTGTTTCTCCCTGGATTTATTTAGGTAATATCTCTCATAAATCCCCGGGTA
GTTAACGAAAGTTAATGGAGATCAGTAACAATAACTCTAGGGTCATTACTTTGGA
CTCCCTCAGTTTATCCGGGGGAATTGTGTTTAAGAAAATCCCAACTCATAAAGTC
AAGTAGGAGATTAATTCA*atgagtcacatccagagagaaactagttgttccgacctcgtttgaatagcaatatgg*
atgcagatctgtacggatataaatgggcgcgagataacgtaggccaatctggggccactatttatcggttatatggcaaaccagat
gctcccgaactgtttctcaaacatggcaaagggtctgtggccaatgatgttaccgatgaaatggtgcggttgaactggttgacaga
atttatgcccctcccgaccatcaaacatttatcaggactccagacgatgcatggctattaactacggccattcctgggaaaactgcc
tttcaggtgttggaagaatatcccgattctggtgagaatatcgtcgatgcgttagcggttttttctaagacgtctacatagcattcccgttt
gcaattgtccctttaattcggaccgggtgttccgcttggcgcaggctcagtcccggatgaataacggtttggtagatgcctcggacttt
gatgatgaacggaacggctggcccgttgaacaggtttggaaagagatgcataagctgctgccctttctcccccgacagcgttgttac
tcatggagatttttctctcgataatctgatttcgacgaaggcaagctaattggctgtatcgatgtgggacgggtagggattgcggac
cggtatcaagacctagcaattttgtggaactgcctaggtgaatttttcccccagcctacaaaaacggctgtttcaaaaatacggaatc
*gataatcccgacatgaacaaattacaatttcatctgatgctagatgagttctttc*accatcaccatcaccat**atcgaagggcga***TCC*
TATCAAGGCAATTCCGATTGTTATTTTGGCAATGGCTCCGCCTATCGGGGCACCCATTCCTTGACCGAATC
CGGCGCCTCCTGTTTGCCCTGGAATTCCATGATTTTGATTGGCAAAGTGTATACCGCCCAAAATCCCTCC
GCCCAAGCCTTGGGCTTGGGCAAACATAATTATTGTCGGAATCCCGATGGCGATGCCAAACCCTGGTGTC
ATGTGTTGAAGAATCGGCGGTTGACCTGGGAATATTGTGATGTGCCCTCCTGTTCCACCTGTGGCTTGCG
GCAATATTCCCAACCCCAATTTCGGATTAAAGGCGGCTTGTTTGCCGATATTGCCTCCCATCCCTGGCAAG

CCGCCATCTTTGCCAAACATCGGCGGTCTCCCGGCGAACGGTTCTTGTGTGGCGGCATTTTGATTTCCTC
CTGTTGGATTTTGTCCGCCGCCCATTGTTTTCAAGAACGGTTTCCTCCCCATCATTTGACCGTGATTTTGG
GCCGGACCTATCGGGTGGTGCCCGGCGAAGAAGAACAGAAATTTGAAGTGGAGAAATATATTGTGCATAA
AGAATTTGATGATGATACCTATGATAATGATATTGCCTTGTTGCAATTGAAATCCGATTCCTCCCGGTGTGC
CCAAGAATCCTCCGTGGTGCGGACCGTGTGTTTGCCTCCCGCCGATTTGCAATTGCCCGATTGGACCGA
ATGTGAATTGTCCGGCTATGGCAAACATGAAGCCTTGTCTCCCTTTTATTCCGAACGGTTGAAAGAAGCCC
ATGTGCGGTTGTATCCCTCCTCCCGGTGTACCTCCCAACATTTGTTGAATCGGACCGTGACCGATAATATG
TTGTGTGCCGGCGATACCCGGTCCGGCGGCCCCCAAGCCAATTTGCATGATGCCTGTCAAGGCGATTCC
GGCGGCCCCTTGGTGTGTTTGAATGATGGCCGGATGACCTTGGTGGGCATTATTTCCTGGGGCTTGGGC
TGTGGCCAGAAAGATGTGCCCGGCGTGTATACCAAAGTGACCAATTATTTGGATTGGATTCGGGATAATAT
GCGGCCCTAA<u>ttttttaaggcagttattggtgcccttaaacgcctggg</u>GATCCTCTGGTTATTTTAAAAACC
AACTTTACTCAGGTTCCATACCCGAGAAAATCCAGCTTAAAGCTGACATATCTAG
GAAAATTTTCACATTCTAACGGGAGATACCAGAACAATGAAAACCCCTTTAACT
GAAGCCGTTTCCACCGCTGACTCTCAAGGTCGCTTTCTGAGCAGCACCGAAT
TGCAAATTGCTTTCGGTCGTCTACGTCAAGCTAATGCTGGTTTGCAAGCCGC
TAAAGCTCTGACCGACAATGCCCAGAGCTTGGTAAATGGTGCTGCCCAAGC
CGTTTATAACAAATTCCCCTACACCACCCAAACCCAAGGCAACAACTTTGCT
GCGGATCAACGGGGTAAAGACAAGTGTGCCCGGGACATCGGCTACTACCTC
CGCATCGTTACCTACTGCTTAGTTGCTGGTGGTACCGGTCCTTTGGATGAGT
ACTTGATCGCCGGTATTGATGAAATCAACCGCACCTTTGACCTCTCCCCCAG
CTGGTATGTTGAAGCTCTGAAATACATCTCGAG

10. *cpc us ... cpcB*INS - cmR + cpc genes ... cpc ds* construct (see, Fig. 16):

CTCGAG - XhoI DNA restriction site
AGATCT - BglII DNA restriction site
GGATCC - BamHI DNA restriction site
Lower case - cpcB
5' RECOMBINATION
UPPER CASE - Factor Xa cleavage site (IEGR)
lower case - Human proinsulin, codon-optimized for expression in *Synechocystis* PCC.6803
lower case - intergenic sequence in Cinzia's construct
lower case - cmR
lower case underlined - Transcription terminator
3' RECOMBINATION
lower case italics - *cpcB-cpcA* intergenic sequence
lower case bold - *cpcA* (partial)

SEQ ID NO:14 *cpc us ... cpcB*INS - cmR + cpc genes ... cpc ds* (2112 nt) nucleic acid sequence

atgttcgacgtattcactcgggttgtttcccaagctgatgctcgcggcgagtacctctctggttctcagttagatgctttgagcgctaccgtt
gctgaaggcaacaaacggattgattctgttaaccgcatcaccggtaatgcttccgctatcgtttccaacgctgctcgtgctttgttcgccg
aacagccccaattaatccaacccggtggaaacgcctacaccagccgtcgtatggctgcttgtttgcgtgacatggaaatcatcctccgc
tatgttacctacgcaaccttcaccggcgacgcttccgttctagaagatcgttgcttgaacggtctccgtgaaacctacgttgccctgggtg
ttcccggtgcttccgtagctgctggcgttcaaaaaatgaaagaagctgccctggacatcgttaacgatcccaatggcatcacccgtggt
gattgcagtgctatcgttgctgaaatcgctggttacttcgaccgcgccgctgctgccgtagccATCGAAGGGCGAttcgtga
accagcacttgtgcggtagtcacttagtcgaagcgctctatctagtctgtggtgaacgaggtttcttctatactcctaagactcgacgtga
ggctgaggacctccaagtaggacaggtagaactaggaggcggaccaggagccgggtctttgcagccgttggcactagaagggagc
ctccagaagcgagggatcgtggagcagtgctgcacatccatctgtagcttataccaattagagaattactgcaattagAGATCTgc
ggccgcgttgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtatttttgagttatcgagattttc
aggagctaaggaagctaaa**atggagaaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaagaacattt**
tgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaaagaaa
aataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccggaattccgtatggcaatgaaag
acggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagt
gaataccacgacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttccccta
aagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaatatggac
aacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatc
atgccgtctgtgatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaat
ttttttaaggcagttattggtgcccttaaacgcctgg*GGATCCtctggttatttaaaaaccaacttactcaggttccataccgaga*
*aaatccagcttaaagctgacatatctaggaaaatttcacattctaacgggagataccagaaca***atgaaaaccccttttaactgaa**

gccgtttccaccgctgactctcaaggtcgctttctgagcagcaccgaattgcaaattgctttcggtcgtctacgtcaagctaatgc
tggtttgcaagccgctaaagctctgaccgacaatgcccagagcttggtaaatggtgctgcccaagccgtttataacaaattccc
ctacaccacccaaacccaaggcaacaactttgctgcggatcaacggggtaaagacaagtgtgcccgggacatcggctactac
ctccgcatcgttacctactgcttagttgctggtggtaccggtcctttggatgagtacttgatcgccggtattgatgaaatcaaccg
caccttttgacctctccccccagctggtatgtt

*11. cpcB*L7*His*TTFC-smR +cpc* (3243 nt) (see Fig. 17, panel A)
TTFC: Tetanus Toxin Fragment C

UPPER CASE, *cpcB* gene + L7 linker (underlined) for homologous recombination (537 nt)
Lower case <caccatcaccatcaccatgataatttgtatttacaaggc>: His-tag + TEV cleavage site (39 nt)
UPPER CASE BOLD, Tetanus Toxin Fragment C (TTFC) + **STOP CODON** (1356 nt)
Lower case bold RBS (18 nt)
UPPER CASE ITALICS, *smR* gene for antibiotic selection (792 nt)
Lower case italics, transcription terminator + intergenic seq + partial *cpcA* gene for homologous recombination (501 nt)

ATGTTCGACGTATTCACTCGGGTTGTTTCCCAAGCTGATGCTCGCGGCGAGTACCTCTCTG
GTTCTCAGTTAGATGCTTTGAGCGCTACCGTTGCTGAAGGCAACAAACGGATTGATTCTG
TTAACCGCATCACCGGTAATGCTTCCGCTATCGTTTCCAACGCTGCTCGTGCTTTGTTCGC
CGAACAGCCCCAATTAATCCAACCCGGTGGAAACGCCTACACCAGCCGTCGTATGGCTG
CTTGTTTGCGTGACATGGAAATCATCCTCCGCTATGTTACCTACGCAACCTTCACCGGCG
ACGCTTCCGTTCTAGAAGATCGTTGCTTGAACGGTCTCCGTGAAACCTACGTTGCCCTGG
GTGTTCCCGGTGCTTCCGTAGCTGCTGGCGTTCAAAAAATGAAAGAAGCTGCCCTGGACA
TCGTTAACGATCCCAATGGCATCACCCGTGGTGATTGCAGTGCTATCGTTGCTGAAATCG
CTGGTTACTTCGACCGCGCCGCTGCTGCCGTAGCCCCCATGCCTTGGCGCGTGATTcaccatc
accatcaccatgataatttgtatttacaaggc**AAGAACTTAGACTGTTGGGTCGATAATGAGGAGGATAT**
CGATGTCATTCTAAAGAAGTCTACCATCCTAAATCTGGACATTAACAATGATATCAT
TAGTGATATTTCTGGTTTTAATTCTTCTGTTATCACATACCCCGACGCCCAATTAGTT
CCAGGAATTAATGGGAAGGCTATTCATCTAGTAAATAATGAGAGCAGCGAAGTGAT
CGTCCACAAGGCGATGGACATTGAGTATAATGATATGTTCAACAACTTTACTGTGTC
CTTTTGGTTGCGCGTCCCCAAAGTGTCTGCCAGTCACCTGGAACAATACGACACGA
ATGAATATAGTATCATTAGCAGTATGAAAAAGTATAGTTTAAGTATTGGGTCTGGGT
GGTCCGTCTCTCTCAAAGGAAACAACCTCATCTGGACCCTCAAGGATTCTGCAGGC
GAAGTGCGTCAAATTACATTCCGCGACTTGTCCGATAAATTCAATGCGTACCTCGCT
AACAAATGGGTTTTCATCACCATCACGAACGACCGGCTGAGTAGCGCTAACCTCTA
CATTAATGGCGTGTTGATGGGGAGTGCGGAGATCACCGGCCTGGGGGCAATTCGC
GAGGACAACAACATCACACTCAAGTTGGACCGTTGCAATAACAACAACCAATATGT
CTCTATCGACAAATTTCGTATTTTCTGTAAGGCGCTAAACCCAAAGGAGATCGAAAA
GTTATATACTAGTTATTTGAGCATCACGTTTTTACGCGATTTTTGGGGCAACCCACT
GCGTTATGACACTGAATATTATCTCATTCCCGTTGCGTACAGCAGTAAAGACGTCCA
ATTAAAGAATATCACGGATTATATGTATCTGACTAATGCTCCCAGTTACACGAACGG
GAAATTAAACATTTACTACCGCCGTCTGTACTCTGGTCTGAAGTTTATTATCAAACG
CTACACCCCCAACAATGAAATCGACTCTTTTGTTCGGTCTGGTGACTTTATTAAACT
GTACGTAAGTTACAACAACAATGAACACATCGTGGGATACCCTAAAGACGGGAATG
CGTTCAATAACTTAGATCGGATCCTCCGAGTAGGGTATAATGCACCCGGTATTCCTC
TGTATAAGAAGATGGAAGCGGTAAAGCTCCGTGACCTCAAAACTTATAGCGTGCAA
CTCAAACTGTACGACGACAAAGATGCGTCTCTAGGGTTGGTGGGTACCCACAACGG
ACAAATCGGGAATGACCCTAACCGCGATATTCTAATCGCTTCTAATTGGTATTTTAA
CCACTTAAAAGATAAGACCCTCACCTGCGACTGGTATTTCGTCCCAACCGACGAGG
GATGGACTAATGATTGAggaattaggaggtaatat*ATGAGGGAAGCGGTGATCGCCGAAGTATCGA*
CTCAACTATCAGAGGTAGTTGGCGTCATCGAGCGCCATCTCGAACCGACGTTGCTGGCCGTAC
ATTTGTACGGCTCCGCAGTGGATGGCGGCCTGAAGCCACACAGTGATATTGATTTGCTGGTTAC
GGTGACCGTAAGGCTTGATGAAACAACGCGGCGAGCTTTGATCAACGACCTTTTGGAAACTTC
GGCTTCCCCTGGAGAGAGCGAGATTCTCCGCGCTGTAGAAGTCACCATTGTTGTGCACGACGA
CATCATTCCGTGGCGTTATCCAGCTAAGCGCGAACTGCAATTTGGAGAATGGCAGCGCAATGA
CATTCTTGCAGGTATCTTCGAGCCAGCCACGATCGACATTGATCTGGCTATCTTGCTGACAAAA
GCAAGAGAACATAGCGTTGCCTTGGTAGGTCCAGCGGCGGAGGAACTCTTTGATCCGGTTCCT
GAACAGGATCTATTTGAGGCGCTAAATGAAACCTTAACGCTATGGAACTCGCCGCCCGACTGG
GCTGGCGATGAGCGAAATGTAGTGCTTACGTTGTCCCGCATTTGGTACAGCGCAGTAACCGGC

AAAATCGCGCCGAAGGATGTCGCTGCCGACTGGGCAATGGAGCGCCTGCCGGCCCAGTATCA
GCCCGTCATACTTGAAGCTAGACAGGCTTATCTTGGACAAGAAGAAGATCGCTTGGCCTCGCG
CGCAGATCAGTTGGAAGAATTTGTCCACTACGTGAAAGGCGAGATCACCAAGGTAGTCGGCAA
ATAAttttttaaggcagttattggtgccttaaacgcctgggGATCCtctggttattttaaaaaccaacttactcaggttccataccgagaa
aatccagcttaaagctgacatatctaggaaaatttttcacattctaacgggagataccagaacaatgaaaaccccttttaactgaagccgtttcca
ccgctgactctcaaggtcgctttctgagcagcaccgaattgcaaattgctttcggtcgtctacgtcaagctaatgctggtttgcaagccgctaaag
ctctgaccgacaatgcccagagcttggtaaatggtgctgcccaagccgtttataacaaattcccctacaccacccaaacccaaggcaacaact
ttgctgcggatcaacggggtaaagacaagtgtgcccgggacatcggctactacctccgcatcgttacctactgcttagttgctggtggtaccggt
cctttggatgagtacttgatcgccggtattga

12. *cpcB*L7*His*TEV*RBDS1-smR +cpc* (2559 nt) (see, Fig. 18, panel A)
RBDS1 of S protein from SARS-CoV-2,
website http covid-19.uniprot.org/uniprotkb/P0DTC2

UPPER CASE, *cpcB* gene + L7 linker (underlined) for homologous recombination (537 nt)
Lower case <caccatcaccatcaccatgataatttgtatttacaaggc>: His-tag + TEV cleavage site (39 nt)
UPPER CASE BOLD, Receptor Binding Domain (RBD) of the S1-protein from SARS-CoV-2 + **STOP CODON** (672 nt)
Lower case bold RBS (ggaattaggaggtaatat), (18 nt)
UPPER CASE ITALICS, smR gene for antibiotic selection (792 nt)
Lower case italics, transcription terminator + intergenic seq + partial *cpcA* gene for homologous recombination (501 nt)

ATGTTCGACGTATTCACTCGGGTTGTTTCCCAAGCTGATGCTCGCGGCGAGTACCTCTCTG
GTTCTCAGTTAGATGCTTTGAGCGCTACCGTTGCTGAAGGCAACAAACGGATTGATTCTG
TTAACCGCATCACCGGTAATGCTTCCGCTATCGTTTCCAACGCTGCTCGTGCTTTGTTCGC
CGAACAGCCCCAATTAATCCAACCCGGTGGAAACGCCTACACCAGCCGTCGTATGGCTG
CTTGTTTGCGTGACATGGAAATCATCCTCCGCTATGTTACCTACGCAACCTTCACCGGCG
ACGCTTCCGTTCTAGAAGATCGTTGCTTGAACGGTCTCCGTGAAACCTACGTTGCCCTGG
GTGTTCCCGGTGCTTCCGTAGCTGCTGGCGTTCAAAAAATGAAAGAAGCTGCCCTGGACA
TCGTTAACGATCCCAATGGCATCACCCGTGGTGATTGCAGTGCTATCGTTGCTGAAATCG
CTGGTTACTTCGACCGCGCCGCTGCTGCCGTAGCCCCCATGCCTTGGCGCGTGATTcaccatc
accatcaccatgataatttgtatttacaaggc**CGGGTGCAACCCACCGAATCCATTGTGCGGTTTCCCAAT**
ATTACCAATTTGTGTCCCTTTGGCGAAGTGTTTAATGCCACCCGGTTTGCCTCCGTG
TATGCCTGGAATCGGAAACGGATTTCCAATTGTGTGGCCGATTATTCCGTGTTGTAT
AATTCCGCCTCCTTTTCCACCTTTAAATGTTATGGCGTGTCCCCCACCAAATTGAAT
GATTTGTGTTTTACCAATGTGTATGCCGATTCCTTTGTGATTCGGGGCGATGAAGTG
CGGCAAATTGCCCCCGGCCAAACCGGCAAAATTGCCGATTATAATTATAAATTGCC
CGATGATTTTACCGGCTGTGTGATTGCCTGGAATTCCAATAATTTGGATTCCAAAGT
GGGCGGCAATTATAATTATTTGTATCGGTTGTTTCGGAAATCCAATTTGAAACCCTT
TGAACGGGATATTTCCACCGAAATTTATCAAGCCGGCTCCACCCCCTGTAATGGCG
TGGAAGGCTTTAATTGTTATTTTCCCTTGCAATCCTATGGCTTTCAACCCACCAATG
GCGTGGGCTATCAACCCTATCGGGTGGTGGTGTTGTCCTTTGAATTGTTGCATGCC
CCCGCCACCGTGTGTGGCCCCAAAAAATCCACCAATTTGGTGAAAAATAAATGTGT
GAATTTTTGAggaattaggaggtaatat*ATGAGGGAAGCGGTGATCGCCGAAGTATCGACTCAACTAT*
CAGAGGTAGTTGGCGTCATCGAGCGCCATCTCGAACCGACGTTGCTGGCCGTACATTTGTACG
GCTCCGCAGTGGATGGCGGCCTGAAGCCACACAGTGATATTGATTTGCTGGTTACGGTGACCG
TAAGGCTTGATGAAACAACGCGGCGAGCTTTGATCAACGACCTTTTGGAAACTTCGGCTTCCCC
TGGAGAGAGCGAGATTCTCCGCGCTGTAGAAGTCACCATTGTTGTGCACGACGACATCATTCC
GTGGCGTTATCCAGCTAAGCGCGAACTGCAATTTGGAGAATGGCAGCGCAATGACATTCTTGC
AGGTATCTTCGAGCCAGCCACGATCGACATTGATCTGGCTATCTTGCTGACAAAAGCAAGAGAA
CATAGCGTTGCCTTGGTAGGTCCAGCGGCGGAGGAACTCTTTGATCCGGTTCCTGAACAGGAT
CTATTTGAGGCGCTAAATGAAACCTTAACGCTATGGAACTCGCCGCCCGACTGGGCTGGCGAT
GAGCGAAATGTAGTGCTTACGTTGTCCCGCATTTGGTACAGCGCAGTAACCGGCAAAATCGCG
CCGAAGGATGTCGCTGCCGACTGGGCAATGGAGCGCCTGCCGGCCCAGTATCAGCCCGTCAT
ACTTGAAGCTAGACAGGCTTATCTTGGACAAGAAGAAGATCGCTTGGCCTCGCGCGCAGATCA
GTTGGAAGAATTTGTCCACTACGTGAAAGGCGAGATCACCAAGGTAGTCGGCAAATAAtttttaag
gcagttattggtgcccttaaacgcctgggGATCCtctggttatttttaaaaaccaacttttactcaggttccataccgagaaaatccagcttaaa
gctgacatatctaggaaaatttttcacattctaacgggagataccagaacaatgaaaaacccctttaactgaagccgtttccaccgctgactctcaa
ggtcgctttctgagcagcaccgaattgcaaattgctttcggtcgtctacgtcaagctaatgctggtttgcaagccgctaaagctctgaccgacaat
gcccagagcttggtaaatggtgctgcccaagccgtttataacaaatttcccctacaccacccaaaacccaaggcaacaacttttgctgcggatcaa
cggggtaaagacaagtgtgcccgggacatcggctactacctccgcatcgttacctactgcttagttgctggtggtaccggtcctttggatgagta
cttgatcgccggtattga

*13. cpcB*L7*His*TEV*S1-smR +cpc (3909 nts)* (see, Fig. 19, panel C)
S1 domain of S protein from SARS-CoV-2
website http covid-19.uniprot.org/uniprotkb/P0DTC2

UPPER CASE, *cpcB* gene + L7 linker (underlined) for homologous recombination (537 nt)
Lower case <caccatcaccatcaccatgataatttgtatttacaaggc>: His-tag + TEV cleavage site (39 nt)
UPPER CASE BOLD, S1 domain of spike S-protein from SARS-CoV-2 virus + **STOP CODON** (2022 nt)
Lower case bold RBS (ggaattaggaggtaatat) (18 nt)
UPPER CASE ITALICS, smR gene for antibiotic selection (792 nt)
Lower case italics, transcription terminator + intergenic seq + partial *cpcA* gene for homologous recombination (501 nt)

ATGTTCGACGTATTCACTCGGGTTGTTTCCCAAGCTGATGCTCGCGGCGAGTACCTCTCTG
GTTCTCAGTTAGATGCTTTGAGCGCTACCGTTGCTGAAGGCAACAAACGGATTGATTCTG
TTAACCGCATCACCGGTAATGCTTCCGCTATCGTTTCCAACGCTGCTCGTGCTTTGTTCGC
CGAACAGCCCCAATTAATCCAACCCGGTGGAAACGCCTACACCAGCCGTCGTATGGCTG
CTTGTTTGCGTGACATGGAAATCATCCTCCGCTATGTTACCTACGCAACCTTCACCGGCG
ACGCTTCCGTTCTAGAAGATCGTTGCTTGAACGGTCTCCGTGAAACCTACGTTGCCCTGG
GTGTTCCCGGTGCTTCCGTAGCTGCTGGCGTTCAAAAAATGAAAGAAGCTGCCCTGGACA
TCGTTAACGATCCCAATGGCATCACCCGTGGTGATTGCAGTGCTATCGTTGCTGAAATCG
CTGGTTACTTCGACCGCGCCGCTGCTGCCGTAGCCCCCATGCCTTGGCGCGTGATTcaccatc
accatcaccatgataatttgtatttacaaggcTCCCAATGTGTGAATTTGACCACCCGGACCCAATTGCCC
CCCGCCTATACCAATTCCTTTACCCGGGGCGTGTATTATCCCGATAAAGTGTTTCGG
TCCTCCGTGTTGCATTCCACCCAAGATTTGTTTTGCCCTTTTTTTCCAATGTGACCT
GGTTTCATGCCATTCATGTGTCCGGCACCAATGGCACCAAACGGTTTGATAATCCC
GTGTTGCCCTTTAATGATGGCGTGTATTTTGCCTCCACCGAAAAATCCAATATTATT
CGGGGCTGGATTTTTGGCACCACCTTGGATTCCAAAACCCAATCCTTGTTGATTGTG
AATAATGCCACCAATGTGGTGATTAAAGTGTGTGAATTTCAATTTTGTAATGATCCC
TTTTTGGGCGTGTATTATCATAAAAATAATAAATCCTGGATGGAATCCGAATTTCGG
GTGTATTCCTCCGCCAATAATTGTACCTTTGAATATGTGTCCCAACCCTTTTTGATG
GATTTGGAAGGCAAACAAGGCAATTTTAAAAATTTGCGGGAATTTGTGTTTAAAAAT
ATTGATGGCTATTTTAAAATTTATTCCAAACATACCCCCATTAATTTGGTGCGGGAT
TTGCCCCAAGGCTTTTCCGCCTTGGAACCCTTGGTGGATTTGCCCATTGGCATTAAT
ATTACCCGGTTTCAAACCTTGTTGGCCTTGCATCGGTCCTATTTGACCCCCGGCGAT
TCCTCCTCCGGCTGGACCGCCGGCGCCGCCGCCTATTATGTGGGCTATTTGCAACC
CCGGACCTTTTTGTTGAAATATAATGAAAATGGCACCATTACCGATGCCGTGGATTG
TGCCTTGGATCCCTTGTCCGAAACCAAATGTACCTTGAAATCCTTTACCGTGGAAAA
AGGCATTTATCAAACCTCCAATTTTCGGGTGCAACCCACCGAATCCATTGTGCGGTT
TCCCAATATTACCAATTTGTGTCCCTTTGGCGAAGTGTTTAATGCCACCCGGTTTGC
CTCCGTGTATGCCTGGAATCGGAAACGGATTTCCAATTGTGTGGCCGATTATTCCGT
GTTGTATAATTCCGCCTCCTTTTCCACCTTTAAATGTTATGGCGTGTCCCCCACCAA
ATTGAATGATTTGTGTTTTACCAATGTGTATGCCGATTCCTTTGTGATTCGGGGCGA
TGAAGTGCGGCAAATTGCCCCCGGCCAAACCGGCAAAATTGCCGATTATAATTATA
AATTGCCCGATGATTTTACCGGCTGTGTGATTGCCTGGAATTCCAATAATTTGGATT
CCAAAGTGGGCGGCAATTATAATTATTTGTATCGGTTGTTTCGGAAATCCAATTTGA
AACCCTTTGAACGGGATATTTCCACCGAAATTTATCAAGCCGGCTCCACCCCCTGTA
ATGGCGTGGAAGGCTTTAATTGTTATTTTCCCTTGCAATCCTATGGCTTTCAACCCA
CCAATGGCGTGGGCTATCAACCCTATCGGGTGGTGGTGTTGTCCTTTGAATTGTTG
CATGCCCCCGCCACCGTGTGTGGCCCCAAAAAATCCACCAATTTGGTGAAAAATAA
ATGTGTGAATTTTAATTTTAATGGCTTGACCGGCACCGGCGTGTTGACCGAATCCAA
TAAAAAATTTTTGCCCTTTCAACAATTTGGCCGGGATATTGCCGATACCACCGATGC
CGTGCGGGATCCCCAAACCTTGGAAATTTTGGATATTACCCCCTGTTCCTTTGGCGG

CGTGTCCGTGATTACCCCCGGCACCAATACCTCCAATCAAGTGGCCGTGTTGTATC
AAGATGTGAATTGTACCGAAGTGCCCGTGGCCATTCATGCCGATCAATTGACCCCC
ACCTGGCGGGTGTATTCCACCGGCTCCAATGTGTTTCAAACCCGGGCCGGCTGTTT
GATTGGCGCCGAACATGTGAATAATTCCTATGAATGTGATATTCCCATTGGCGCCG
GCATTTGTGCCTCCTATCAAACCCAAACCAATTCCCCCCGGCGGGCCCGGTGAggaatt
aggaggtaatat*ATGAGGGAAGCGGTGATCGCCGAAGTATCGACTCAACTATCAGAGGTAGTTGGC*
GTCATCGAGCGCCATCTCGAACCGACGTTGCTGGCCGTACATTTGTACGGCTCCGCAGTGGAT
GGCGGCCTGAAGCCACACAGTGATATTGATTTGCTGGTTACGGTGACCGTAAGGCTTGATGAA
ACAACGCGGCGAGCTTTGATCAACGACCTTTTGGAAACTTCGGCTTCCCCTGGAGAGAGCGAG
ATTCTCCGCGCTGTAGAAGTCACCATTGTTGTGCACGACGACATCATTCCGTGGCGTTATCCAG
CTAAGCGCGAACTGCAATTTGGAGAATGGCAGCGCAATGACATTCTTGCAGGTATCTTCGAGCC
AGCCACGATCGACATTGATCTGGCTATCTTGCTGACAAAAGCAAGAGAACATAGCGTTGCCTTG
GTAGGTCCAGCGGCGGAGGAACTCTTTGATCCGGTTCCTGAACAGGATCTATTTGAGGCGCTA
AATGAAACCTTAACGCTATGGAACTCGCCGCCCGACTGGGCTGGCGATGAGCGAAATGTAGTG
CTTACGTTGTCCCGCATTTGGTACAGCGCAGTAACCGGCAAAATCGCGCCGAAGGATGTCGCT
GCCGACTGGGCAATGGAGCGCCTGCCGGCCCAGTATCAGCCCGTCATACTTGAAGCTAGACA
GGCTTATCTTGGACAAGAAGAAGATCGCTTGGCCTCGCGCGCAGATCAGTTGGAAGAATTTGT
CCACTACGTGAAAGGCGAGATCACCAAGGTAGTCGGCAAATAAttttttaaggcagttattggtgcccttaaacg
cctgggGATCCtctggttatttttaaaaaccaactttactcaggttccataccgagaaaatccagcttaaagctgacatatctaggaaaatttt
cacattctaacgggagataccagaacaatgaaaaccccttttaactgaagccgtttccaccgctgactctcaaggtcgctttctgagcagcaccg
aattgcaaattgctttcggtcgtctacgtcaagctaatgctggtttgcaagccgctaaagctctgaccgacaatgcccagagcttggtaaatggtg
ctgcccaagccgtttataacaaattcccctacaccacccaaacccaaggcaacaactttgctgcggatcaacggggtaaagacaagtgtgcc
cgggacatcggctactacctccgcatcgttacctactgcttagttgctggtggtaccggtcctttggatgagtacttgatcgccggtattga

*14. cpcB*L7*His*TEV*ctxB-smR +cpc* (2199 nt) (see, Fig. 19, panel D)
ctxB from *Vibrio cholerae*, web site http www.uniprot.org/uniprot/Q57193

UPPER CASE, *cpcB* gene + L7 linker (underlined) for homologous recombination (537 nt)
Lower case <caccatcaccatcaccatgataatttgtatttacaaggc>: His-tag + TEV cleavage site (39 nt)
UPPER CASE BOLD, *ctxB* gene + **STOP CODON** (312 nt)
Lower case bold RBS (ggaattaggaggtaatat) (18 nt)
UPPER CASE ITALICS, *smR* gene for antibiotic selection (792 nt)
Lower case italics, transcription terminator + intergenic seq + partial *cpcA* gene for homologous recombination (501 nt)

ATGTTCGACGTATTCACTCGGGTTGTTTCCCAAGCTGATGCTCGCGGCGAGTACCTCTCTG
GTTCTCAGTTAGATGCTTTGAGCGCTACCGTTGCTGAAGGCAACAAACGGATTGATTCTG
TTAACCGCATCACCGGTAATGCTTCCGCTATCGTTTCCAACGCTGCTCGTGCTTTGTTCGC
CGAACAGCCCCAATTAATCCAACCCGGTGGAAACGCCTACACCAGCCGTCGTATGGCTG
CTTGTTTGCGTGACATGGAAATCATCCTCCGCTATGTTACCTACGCAACCTTCACCGGCG
ACGCTTCCGTTCTAGAAGATCGTTGCTTGAACGGTCTCCGTGAAACCTACGTTGCCCTGG
GTGTTCCCGGTGCTTCCGTAGCTGCTGGCGTTCAAAAAATGAAAGAAGCTGCCCTGGACA
TCGTTAACGATCCCAATGGCATCACCCGTGGTGATTGCAGTGCTATCGTTGCTGAAATCG
CTGGTTACTTCGACCGCGCCGCTGCTGCCGTAGCCCCCATGCCTTGGCGCGTGATTcaccatc
accatcaccatgataatttgtatttacaaggc**ACCCCCCAAAATATTACCGATTTGTGTGCCGAATATCAT**
AATACCCAAATTCATACCTTGAATGATAAAATTTTTTCCTATACCGAATCCTTGGCC
GGCAAACGGGAAATGGCCATTATTACCTTTAAAAATGGCGCCACCTTTCAAGTGGA
AGTGCCCGGCTCCCAACATATTGATTCCCAAAAAAAAGCCATTGAACGGATGAAAG
ATACCTTGCGGATTGCCTATTTGACCGAAGCCAAAGTGGAAAAATTGTGTGTGTGG
AATAATAAAACCCCCCATGCCATTGCCGCCATTTCCATGGCCAATTGAggaattaggaggta
atat*ATGAGGGAAGCGGTGATCGCCGAAGTATCGACTCAACTATCAGAGGTAGTTGGCGTCATC*
GAGCGCCATCTCGAACCGACGTTGCTGGCCGTACATTTGTACGGCTCCGCAGTGGATGGCGG
CCTGAAGCCACACAGTGATATTGATTTGCTGGTTACGGTGACCGTAAGGCTTGATGAAACAACG
CGGCGAGCTTTGATCAACGACCTTTTGGAAACTTCGGCTTCCCCTGGAGAGAGCGAGATTCTC
CGCGCTGTAGAAGTCACCATTGTTGTGCACGACGACATCATTCCGTGGCGTTATCCAGCTAAGC
GCGAACTGCAATTTGGAGAATGGCAGCGCAATGACATTCTTGCAGGTATCTTCGAGCCAGCCA
CGATCGACATTGATCTGGCTATCTTGCTGACAAAAGCAAGAGAACATAGCGTTGCCTTGGTAGG
TCCAGCGGCGGAGGAACTCTTTGATCCGGTTCCTGAACAGGATCTATTTGAGGCGCTAAATGAA
ACCTTAACGCTATGGAACTCGCCGCCCGACTGGGCTGGCGATGAGCGAAATGTAGTGCTTACG
TTGTCCCGCATTTGGTACAGCGCAGTAACCGGCAAAATCGCGCCGAAGGATGTCGCTGCCGAC
TGGGCAATGGAGCGCCTGCCGGCCCAGTATCAGCCCGTCATACTTGAAGCTAGACAGGCTTAT
CTTGGACAAGAAGAAGATCGCTTGGCCTCGCGCGCAGATCAGTTGGAAGAATTTGTCCACTAC
GTGAAAGGCGAGATCACCAAGGTAGTCGGCAAATAAttttttaaggcagttattggtgcccttaaacgcctgggGAT
CCtctggttatttttaaaaaccaactttactcaggttccataccgagaaaatcagcttaaagctgacatatctaggaaaatttttcacattctaac
gggagataccagaacaatgaaaaccccttttaactgaagccgtttccaccgctgactctcaaggtcgctttctgagcagcaccgaattgcaaatt
gctttcggtcgtctacgtcaagctaatgctggtttgcaagccgctaaagctctgaccgacaatgcccagagcttggtaaatggtgctgcccaagc
cgtttataacaaattcccctacaccacccaaacccaaggcaacaactttgctgcggatcaacggggtaaagacaagtgtgcccgggacatcg
gctactacctccgcatcgttacctactgcttagttgctggtggtaccggtcctttggatgagtacttgatcgccggtattga

WHAT IS CLAIMED IS:

1. A cyanobacterial host cell comprising an expression unit comprising:

(i) a nucleic acid sequence comprising a transgene that encodes a biopharmaceutical protein, wherein the transgene is fused to the 3' end of a nucleic acid sequence that encodes a cyanobacteria β-subunit of phycocyanin (cpcB) polypeptide to produce a fusion polypeptide comprises cpcB and the biopharmaceutical protein of interest;

(ii) a nucleic acid sequence encoding a cyanobacteria α-subunit of phycocyanin (cpcA) polypeptide; and

(iii) a nucleic acid sequence encoding a cyanobacterial cpcC1, cpcC2 and cpcD polypeptide.

2. The cyanobacterial host cell of claim 1, wherein the recombinant expression unit is operably linked to an endogenous cyanobacteria *cpc* promoter.

3. The cyanobacterial host cell of claim 1, wherein the transgene encodes a native human interferon polypeptide.

4. The cyanobacterial host cell of claim 1, wherein the transgene encodes an interferon polypeptide having at least 95% identity to SEQ ID NO:1.

5. The cyanobacterial host cell of claim 3 or 4, wherein the fusion protein comprises a protease cleavage site between cpcB and the interferon polypeptide.

6. The cyanobacterial host cell of claim 5, wherein the protease cleavage site is a Factor Xa cleavage site.

7. The cyanobacterial host cell of claim 1, wherein the transgene encodes a native human tissue plasminogen activator (tPA) polypeptide or truncated native human tPA polypeptide.

8. The cyanobacterial host cell of claim 1, wherein the transgene encodes a tPA polypeptide having at least 95% identity to the region of SEQ ID NO:2 that lacks the signal peptide or having at least 95% identity to SEQ ID NO:3.

9. The cyanobacterial host cell of claim 7 or 8, wherein the fusion protein comprises a protease cleavage site between cpcB and the tPA polypeptide.

10. The cyanobacterial host cell of claim 1, wherein the transgene encodes a native Tetanus Toxin Fragment C (TTFC) polypeptide or a TTFC polypeptide having at least 95% identity to SEQ ID NO:15.

11. The cyanobacterial host cell of claim 10, wherein the fusion protein comprises a protease cleavage site between cpcB and the TTFC polypeptide.

12. The cyanobacterial host cell of claim 11, wherien the protease cleavage site is a Tobacco Etch Virus (TEV) cysteine protease cleavage site.

13. The cyanobacterial host cell of claim 1, wherein the transgene .encodes a native *Cholera* Toxin Fragment B polypeptide or a *Cholera* Toxin Fragment B polypeptide having at least 95% identity to SEQ ID NO:18.

14. The cyanobacterial host cell of claim 13, wherein the fusion protein comprises a protease cleavage site between cpcB and the *Cholera* Toxin Fragment B polypeptide.

15. The cyanobacterial host cell of claim 14, wherien the protease cleavage site is a Tobacco Etch Virus (TEV) cysteine protease cleavage site.

16. The cyanobacterial host cell of claim 1, wherein the transgene encodes a native human insulin polypeptide

17. The cyanobacterial host cell of claim 1, wherein the transgene encodes an insulin polypeptide having at least 95% identity to SEQ ID NO:4.

18. The cyanobacterial host cell of claim 16 or 17, wherein the fusion protein comprises a protease cleavage site between the cpcB and insulin polypeptide.

19. The cyanobacterial host cell of claim 1, wherein the transgene encodes a SARS-CoV2 polypeptide having at least 95% identity to SEQ ID NO:16 or 17.

20. The cyanobacterial host cell of claim 19, wherein the fusion protein comprises a protease cleavage site between cpcB and the SARS-CoV2 polypeptide.

21. The cyanobacerial host cell of any one of claims 1 to 20, wherein the expression unit comprises an antibiotic resistance gene between the transgene and cpcA.

22. The cyanobacterial host cell of any one of claims 1 to 21, wherein the cyanobacteria is a single celled cyanobacteria.

23. The cyanobacterial host cell of claim 22, where the cyanobacteria is a *Synechococcus sp.*, a *Thermosynechococcus elongatus*, a *Synechocystis* sp., or a *Cyanothece sp.*.

24. The cyanobacterial host cell of any one of claims 1 to 21, wherein the cyanobacteria are micro-colonial cyanobacteria.

25. The cyanobacterial host cell of claim 24, wherein the cyanobacteria is a *Gloeocapsa magma, Gloeocapsa phylum, Gloeocapsa alpicola, Gloeocpasa atrata, Chroococcus* spp., or *Aphanothece* sp.

26. The cyanobacterial host cell of any one of claims 1 to 21, wherein the cyanobacteria is a filamentous cyanobacteria.

27. The cyanobacterial host cell of claim 26, wherein the cyanobacteria is an *Oscillatoria* spp., a *Nostoc sp.*, an *Anabaena* sp., or an *Arthrospira* sp.

28. A cyanobacterial host cell culture comprising cyanobacteria of any one of claims 1 to 27.

29. A photobioreactor containing the cyanobacterial cell culture of claim 28.

30. A method of producing a biopharmaceutical protein, the method comprising culturing the cyanobacterial host cell culture of claim 27 to express the protein.

31. A method of engineering a cyanobacterial host cell to produce a biopharmaceutical protein, the method comprising introducing an expression cassette comprising a nucleic acid sequence comprising a transgene encoding the biopharmaceutical protein joined to the 3' end of a nucleic acid sequence encoding a cyanobacteria β-subunit of phycocyanin (cpcB) polypeptide to provide a polynucleotide sequence encoding a fusion polypeptide, into the cyaobacterial host cell by homologous recombination to generate an expression unit comprising the polynucleotide encoding the fusion polypeptide and the *cpcA*,

cpcC2, *cpcC1*, and *cpcD* operon genes; and selecting a cyanobacterial host cell that expresses a high level of the polypeptide.

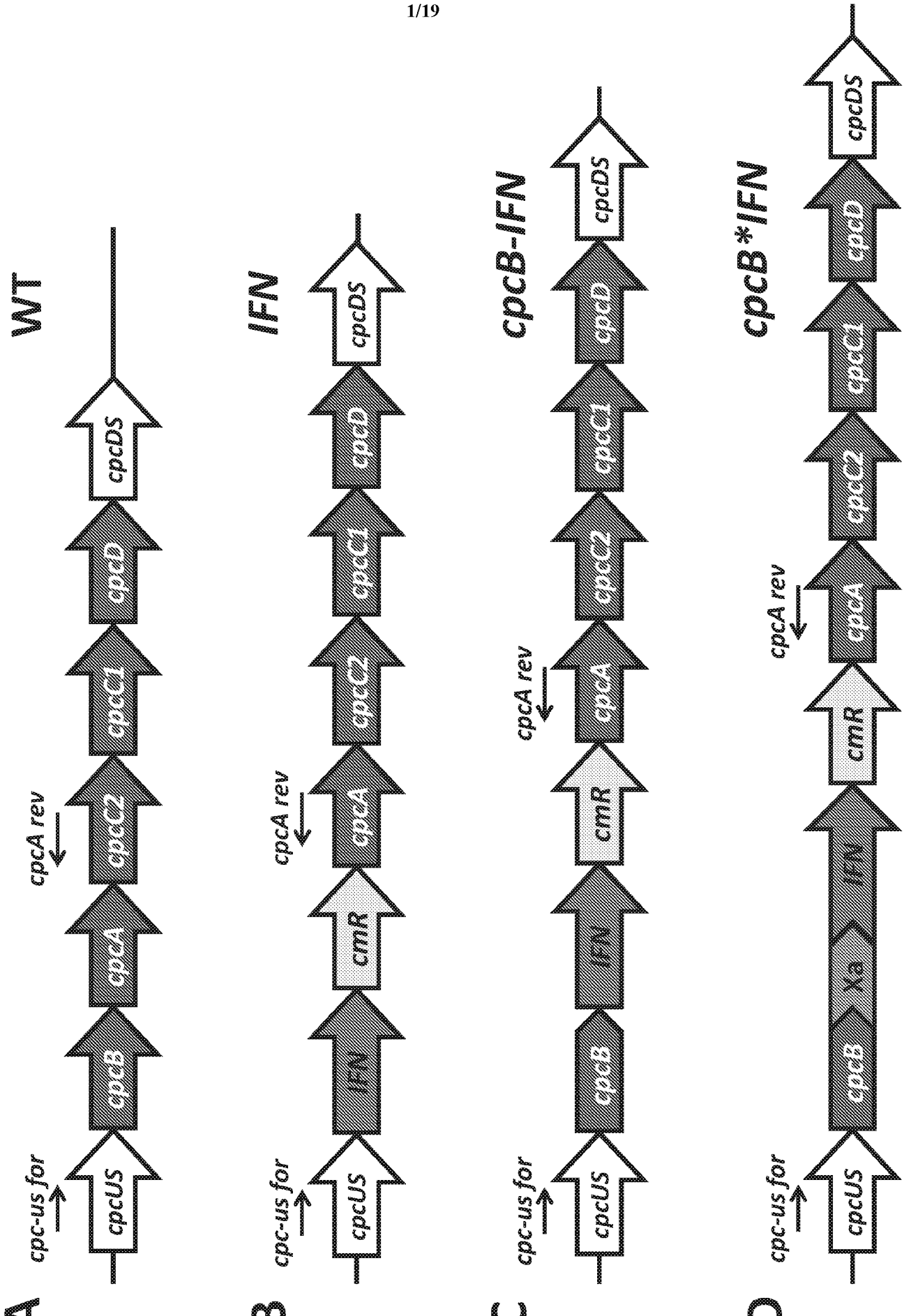
32. The method of claim 31, wherein the biopharmaceutical protein is selected from the group consisting of a native human interferon, a native human insulin, a native human tPA, a truncated form of the native human tPA, a SARS CoV-2 polypeptide receptor binding domain, a TTFC polypeptide, or a *Cholera* Toxin Fragment B polypeptide.

33. An isolated fusion protein comprising a biopharmaceutical protei fused to the 3' end of a nucleic acid sequence that encodes a cyanobacteria β-subunit of phycocyanin (cpcB) polypeptide.

34. The isolated fusion protein of claim 33, wherein the biopharmaceutical protein is selected from the group consisting of a native human interferon, a native human insulin, a native human tPA, a truncated form of the native human tPA, a SARS CoV-2 polypeptide receptor binding domain, a TTFC polypeptide, or a *Cholera* Toxin Fragment B polypeptide.

35. A nucleic acid encoding the fusion protein of claim 34.

Fig. 1


A
WT
cpc-us for
cpcUS
cpcB
cpcA
cpcC2
cpcC1
cpcD
cpcDS
cpcA rev
B
IFN
cpc-us for
cpcUS
IFN
cmR
cpcA
cpcC2
cpcC1
cpcD
cpcDS
cpcA rev
C
cpcB-IFN
cpc-us for
cpcUS
cpcB
IFN
cmR
cpcA
cpcC2
cpcC1
cpcD
cpcDS
cpcA rev
D
cpcB*IFN
cpc-us for
cpcUS
cpcB
Xa
IFN
cmR
cpcA
cpcC2
cpcC1
cpcD
cpcDS
cpcA rev

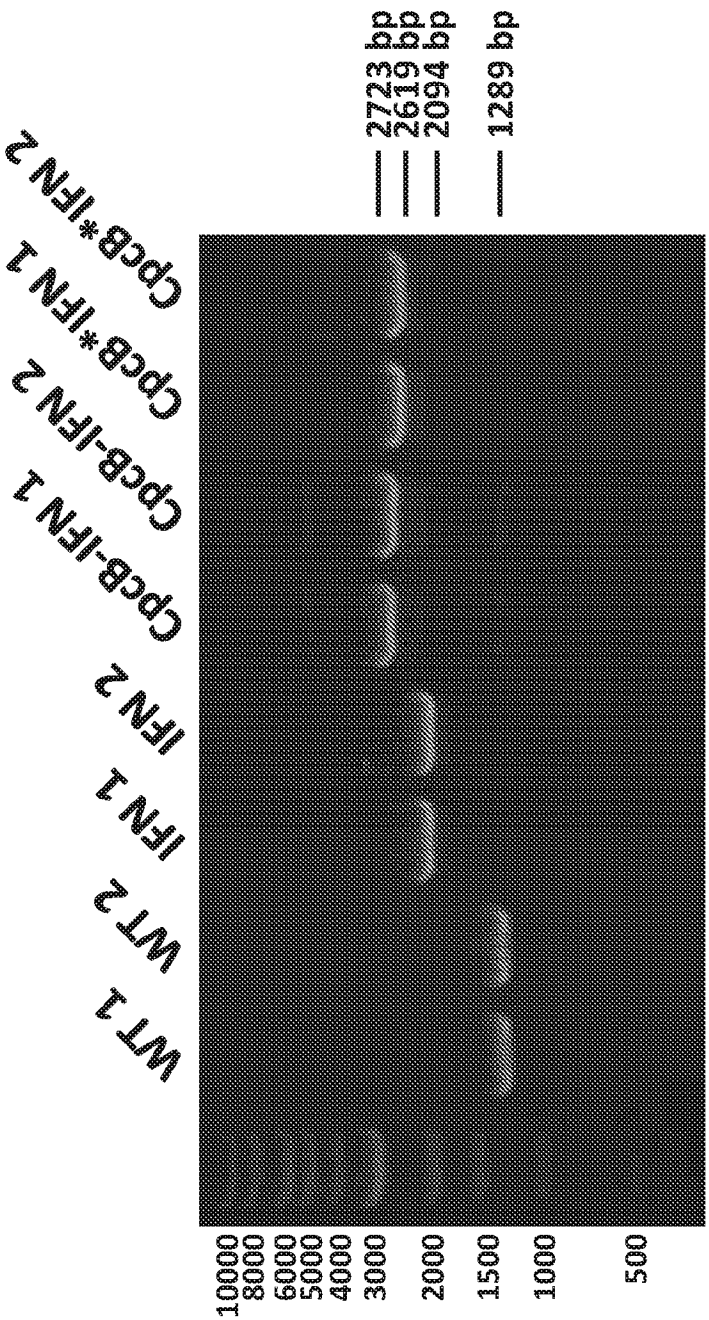

2723 bp
2619 bp
2094 bp
1289 bp
WT 1
WT 2
IFN 1
IFN 2
CpCB-IFN 1
CpCB-IFN 2
CpCB*IFN 1
CpCB*IFN 2
10000
8000
6000
5000
4000
3000
2000
1500
1000
500

Fig. 2

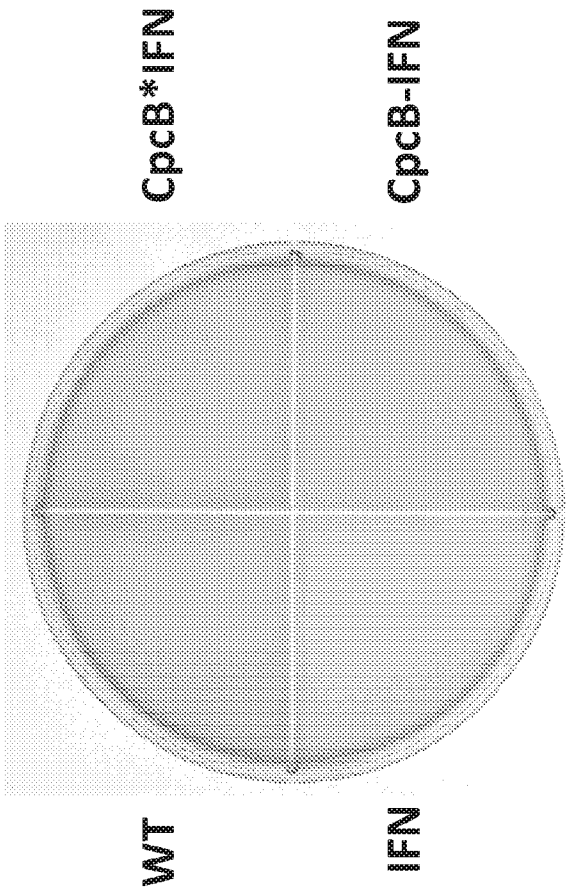


Fig. 3

Fig. 4

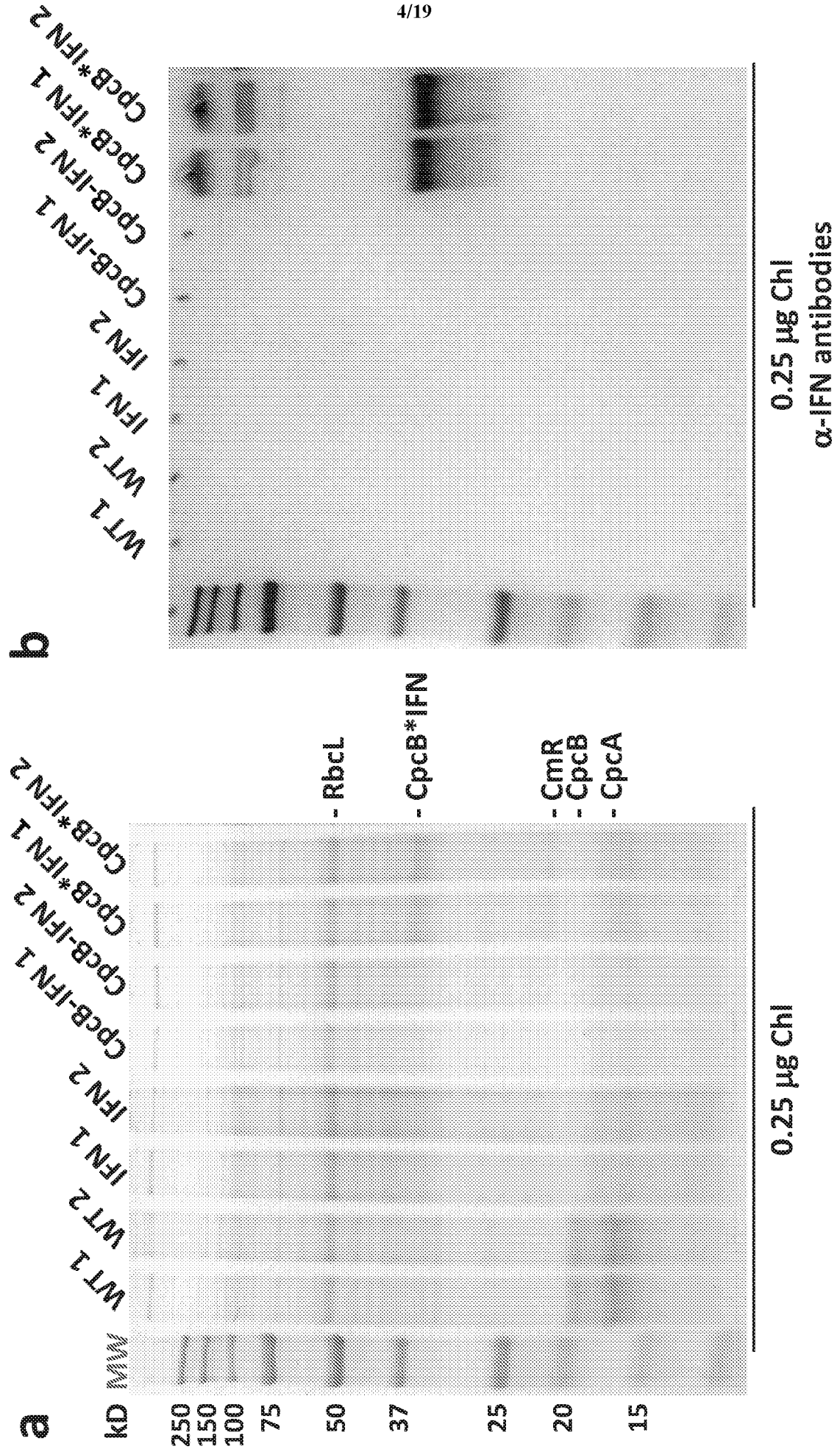
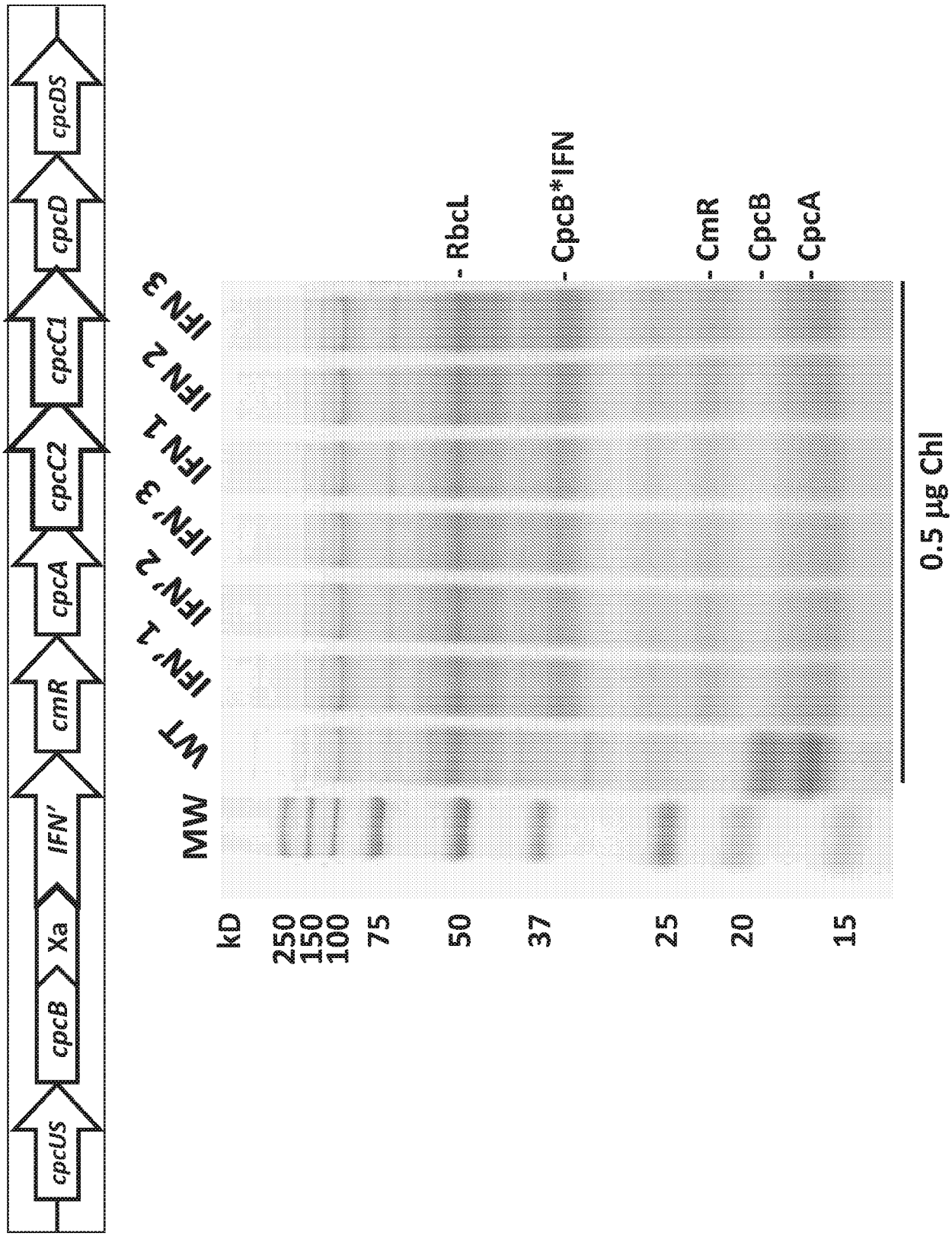

a
kD
250
150
100
75
50
37
25
20
15
WT 1
WT 2
IFN 1
IFN 2
CpcB-IFN 1
CpcB-IFN 2
CpcB*IFN 1
CpcB*IFN 2
- RbcL
- CpcB*IFN
- CmR
- CpcB
- CpcA
0.25 µg Chl
b
WT 1
WT 2
IFN 1
IFN 2
CpcB-IFN 1
CpcB-IFN 2
CpcB*IFN 1
CpcB*IFN 2
0.25 µg Chl
α-IFN antibodies

Fig. 5


cpcUS
cpcB
Xa
IFN'
cmR
cpcA
cpcC2
cpcC1
cpcD
cpcDS
MW
WT
IFN' 1
IFN' 2
IFN' 3
IFN 1
IFN 2
IFN 3
kD
250
150
100
75
50
37
25
20
15
- RbcL
- CpcB*IFN
- CmR
- CpcB
- CpcA
0.5 µg Chl

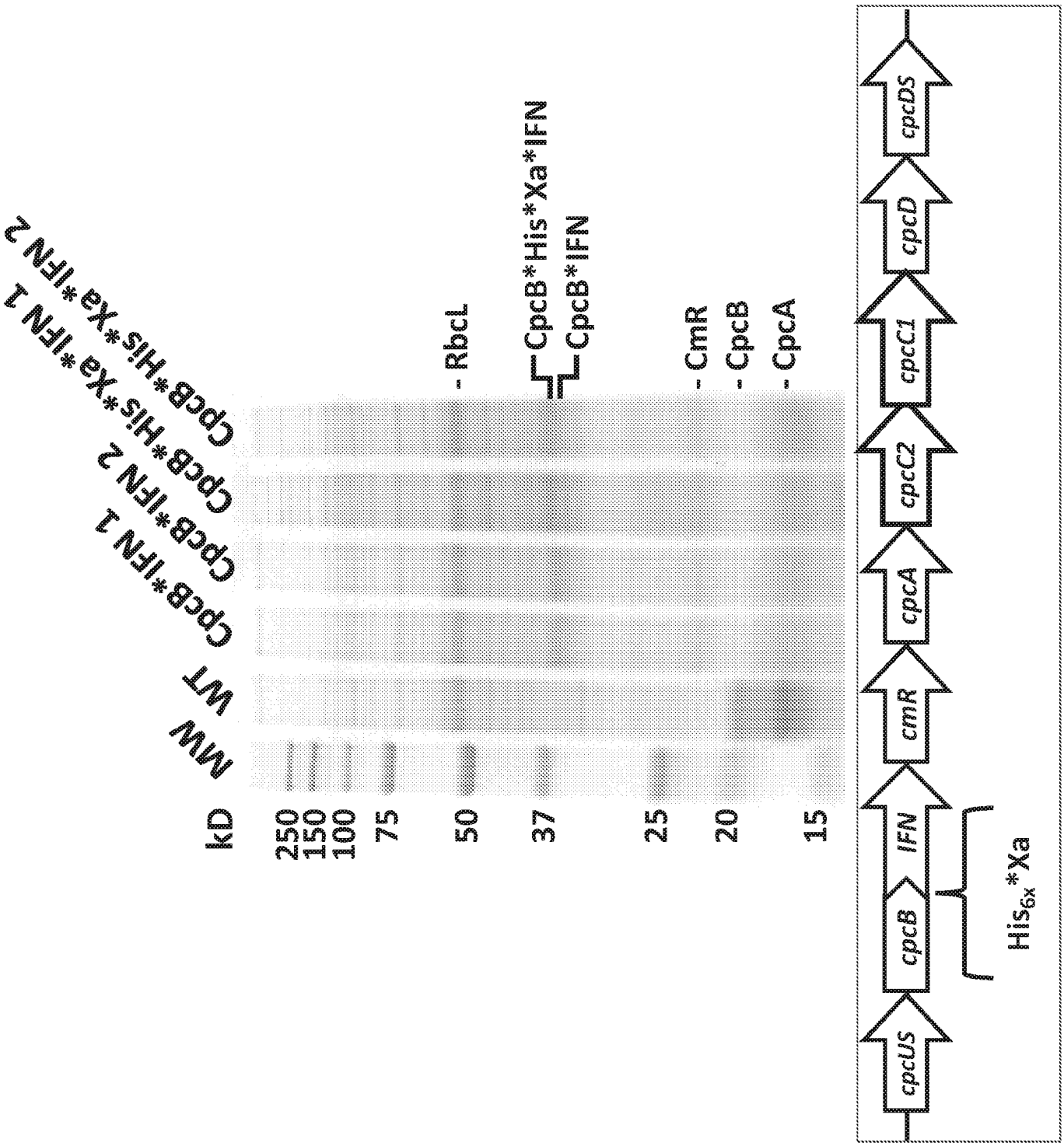

kD
250
150
100
75
50
37
25
20
15
MW
WT
CpcB*IFN 1
CpcB*IFN 2
CpcB*His*Xa*IFN 1
CpcB*His*Xa*IFN 2
- RbcL
- CpcB*His*Xa*IFN
- CpcB*IFN
- CmR
- CpcB
- CpcA
cpcUS
cpcB
IFN
cmR
cpcA
cpcC2
cpcC1
cpcD
cpcDS
His6x*Xa

Fig. 6

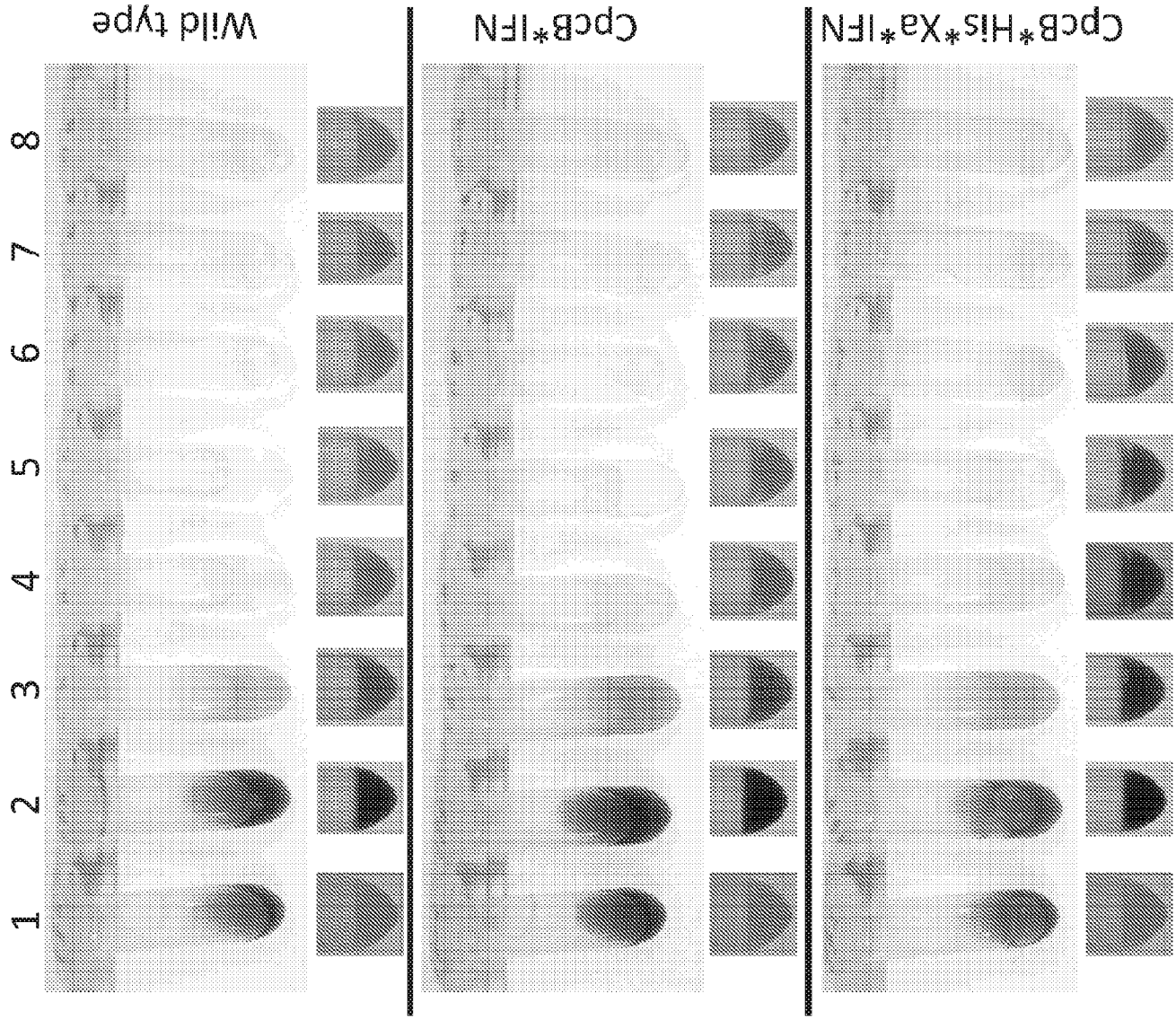
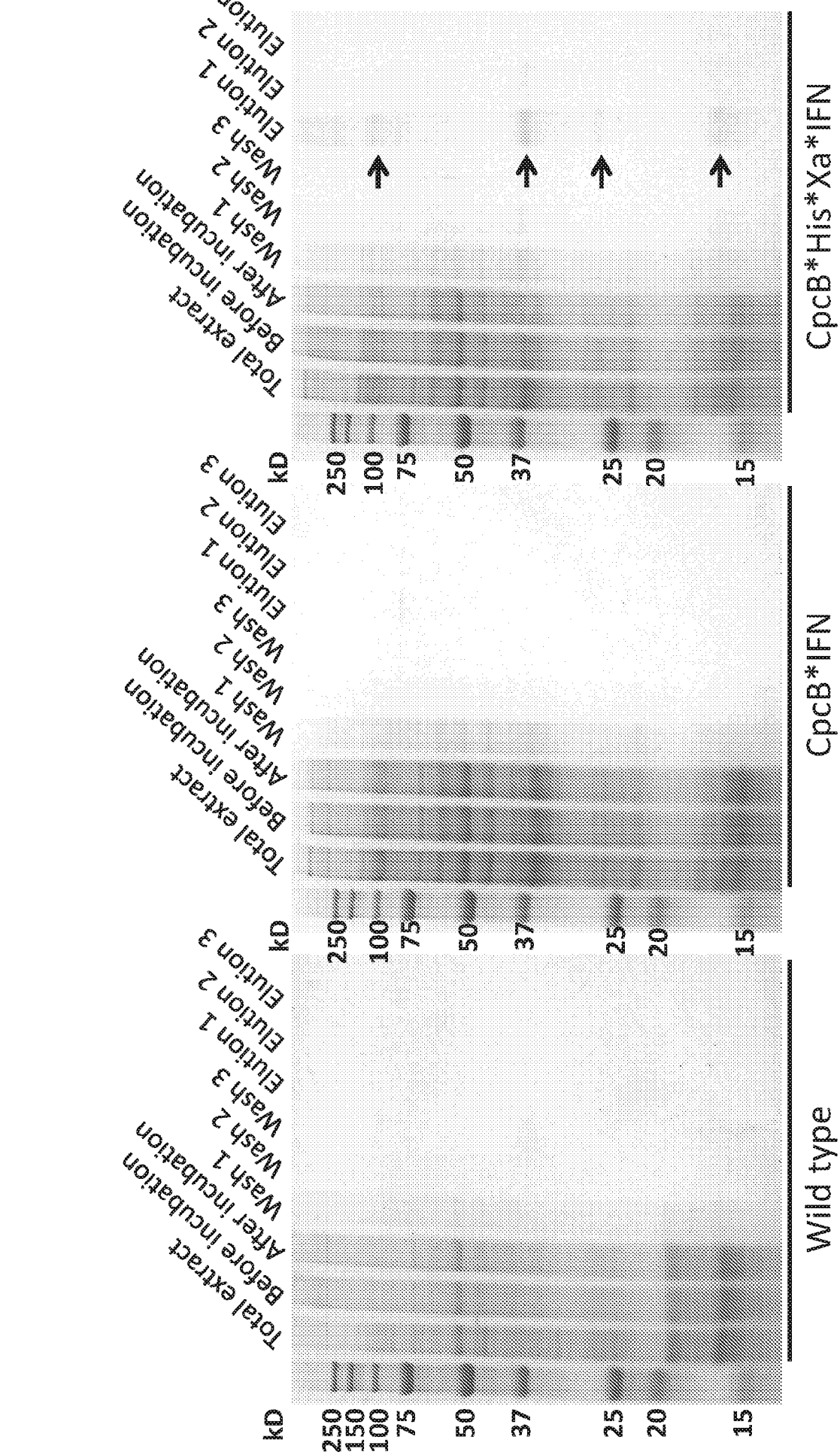


Fig. 7

Fig. 8


kD
250
150
100
75
50
37
25
20
15
Total extract
Before incubation
After incubation
Wash 1
Wash 2
Wash 3
Elution 1
Elution 2
Elution 3
Wild type
CpcB*IFN
CpcB*His*Xa*IFN
Total cell extract loading corresponding to 0.25 µg of Chl

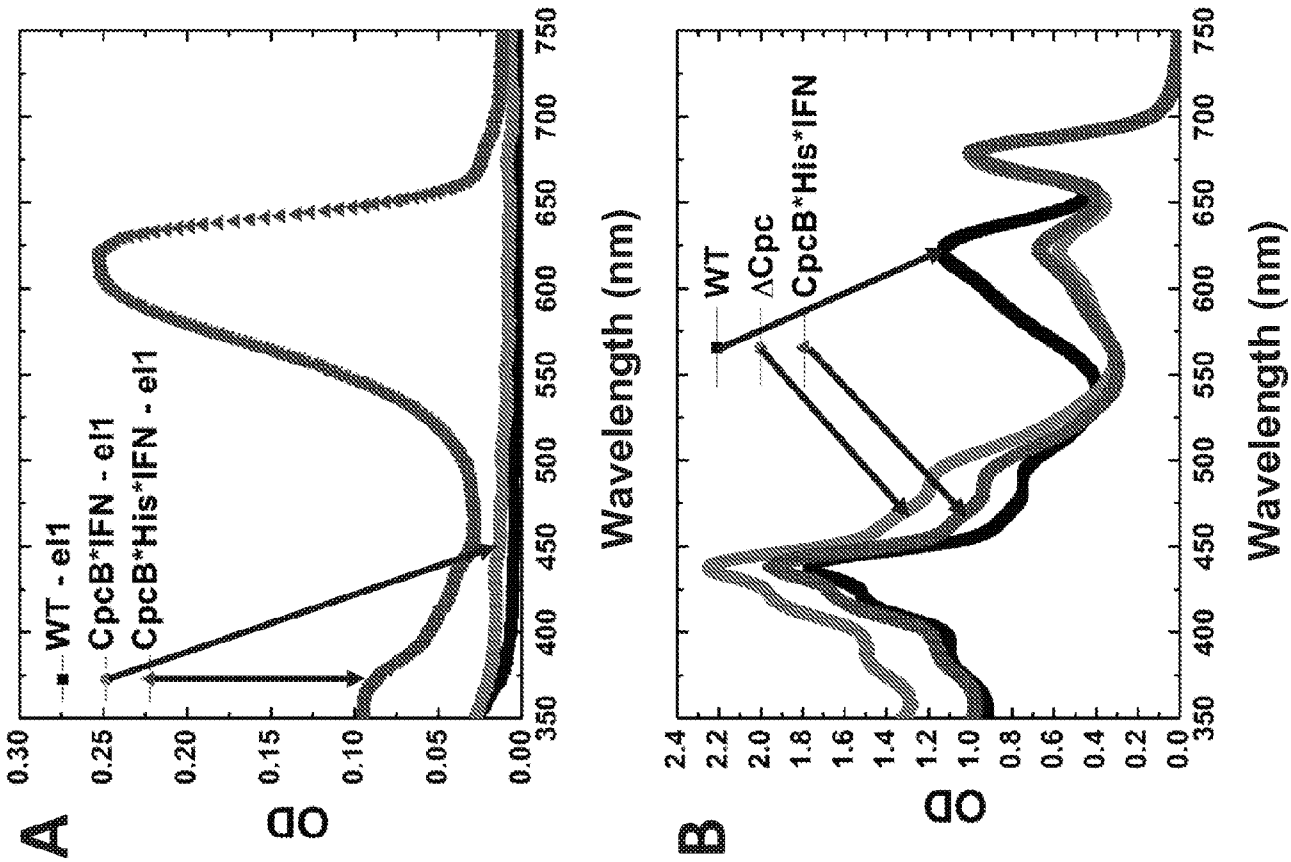

A
OD
Wavelength (nm)
WT - el1
CpcB*IFN - el1
CpcB*His*IFN - el1
B
OD
Wavelength (nm)
WT
ΔCpc
CpcB*His*IFN

Fig. 9

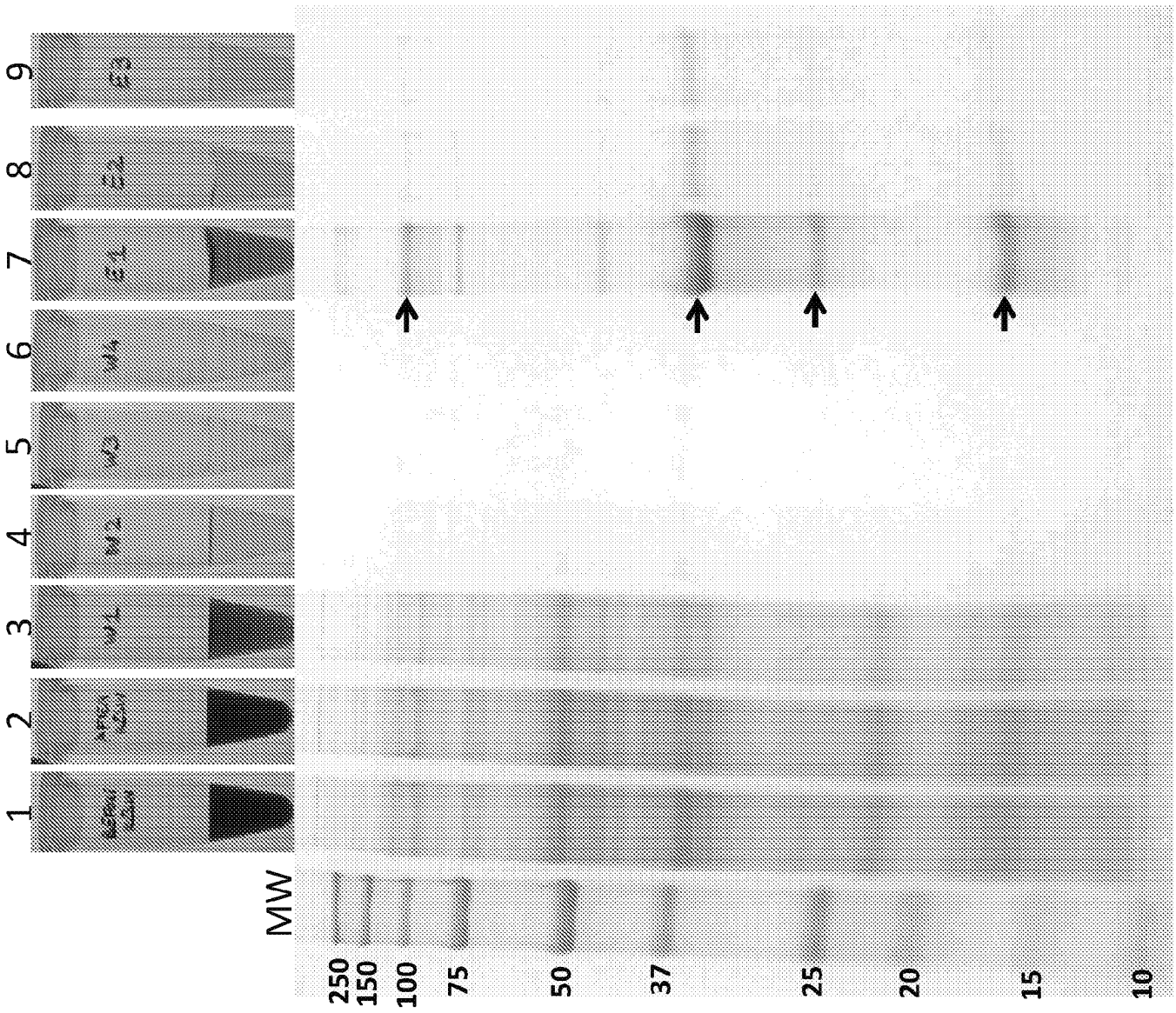


Fig. 10

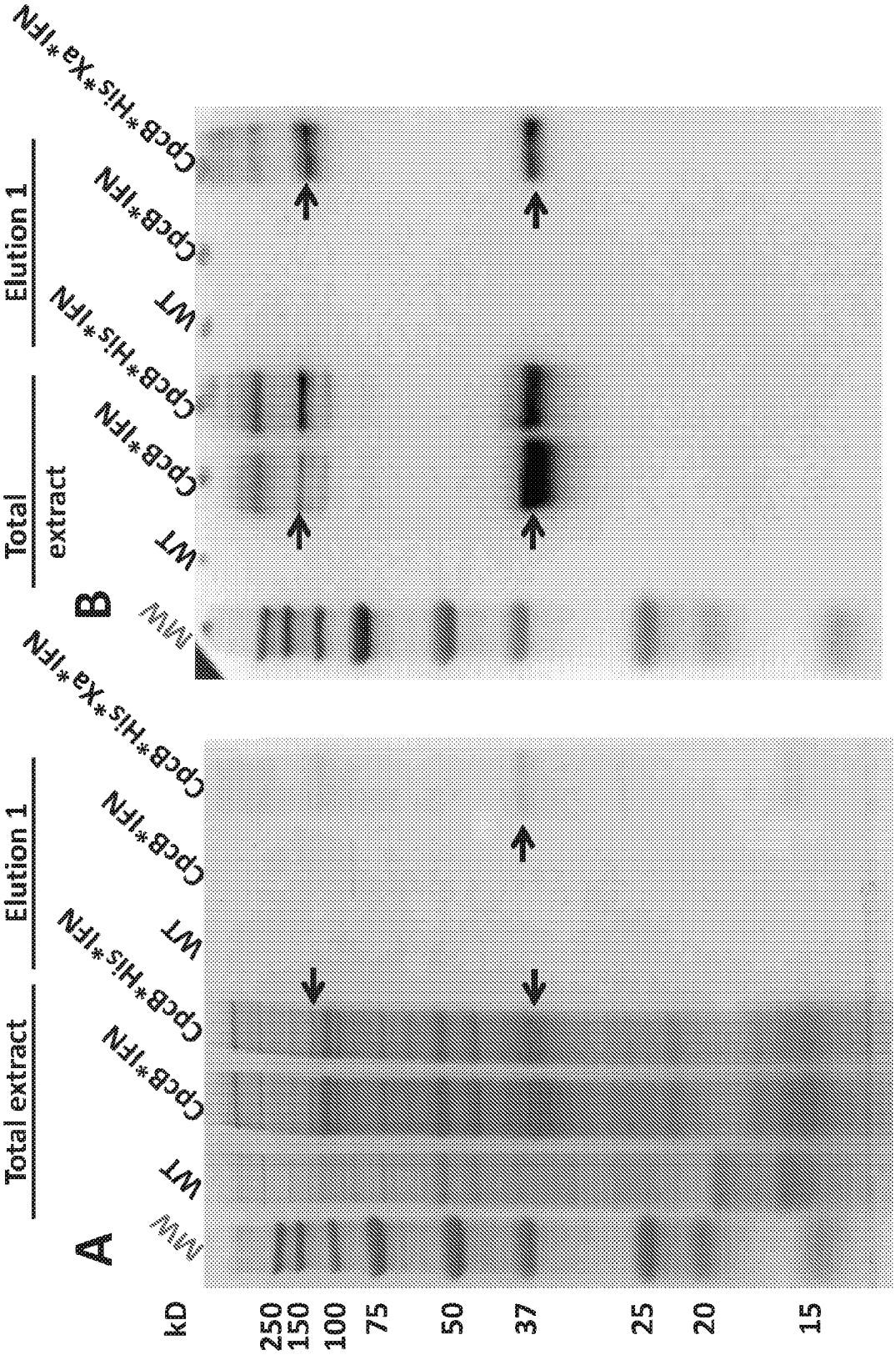

A
B
MW
kD
250
150
100
75
50
37
25
20
15
Total extract
Elution 1
WT
CpcB*IFN
CpcB*His*IFN
CpcB*His*Xa*IFN

Fig. 11

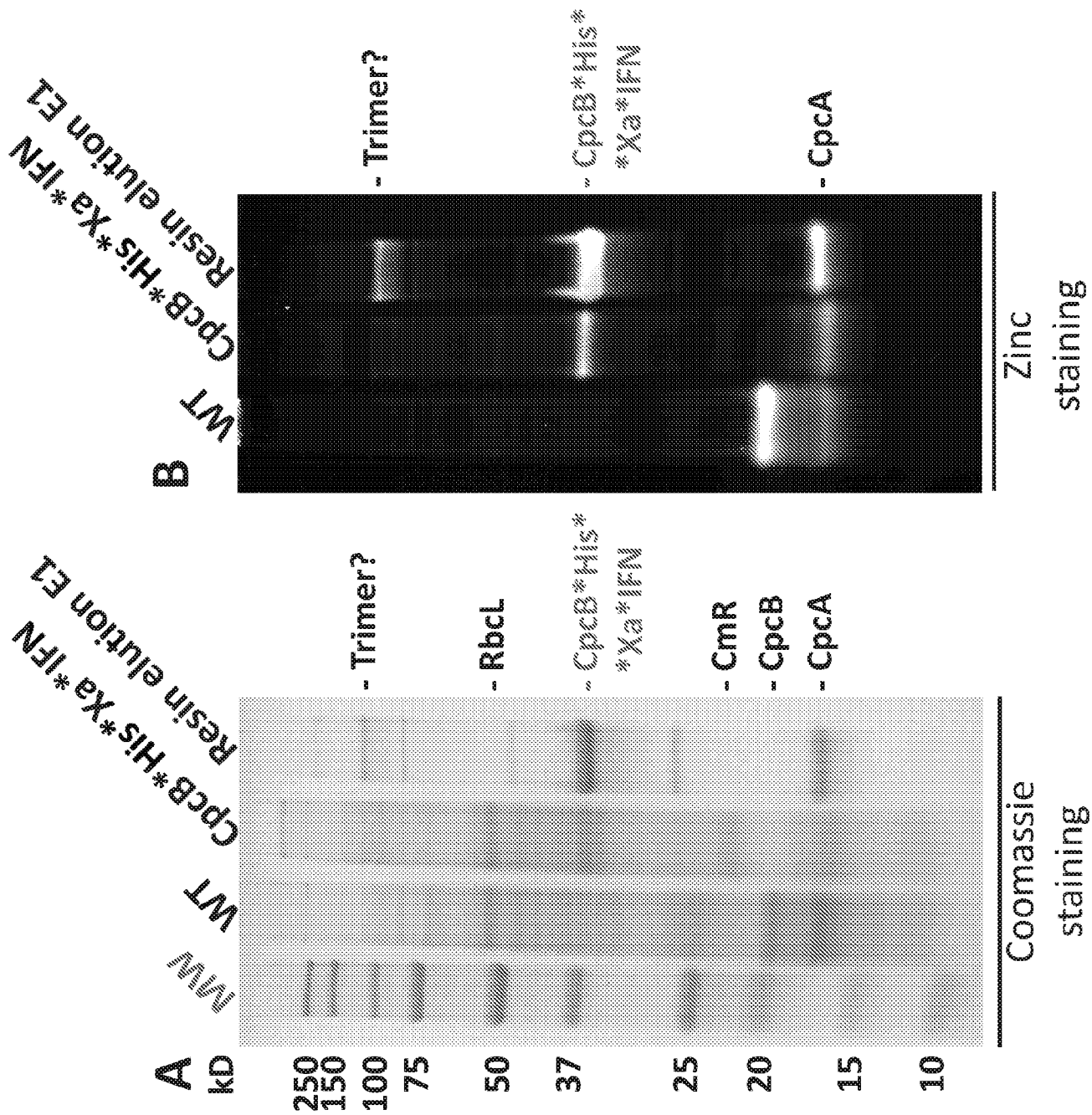

A
kD
250
150
100
75
50
37
25
20
15
10
MW
WT
CpcB*His*Xa*IFN
Resin elution E1
- Trimer?
- RbcL
- CpcB*His*
*Xa*IFN
- CmR
- CpcB
- CpcA
Coomassie
staining
B
WT
CpcB*His*Xa*IFN
Resin elution E1
- Trimer?
- CpcB*His*
*Xa*IFN
- CpcA
Zinc
staining

Fig. 12

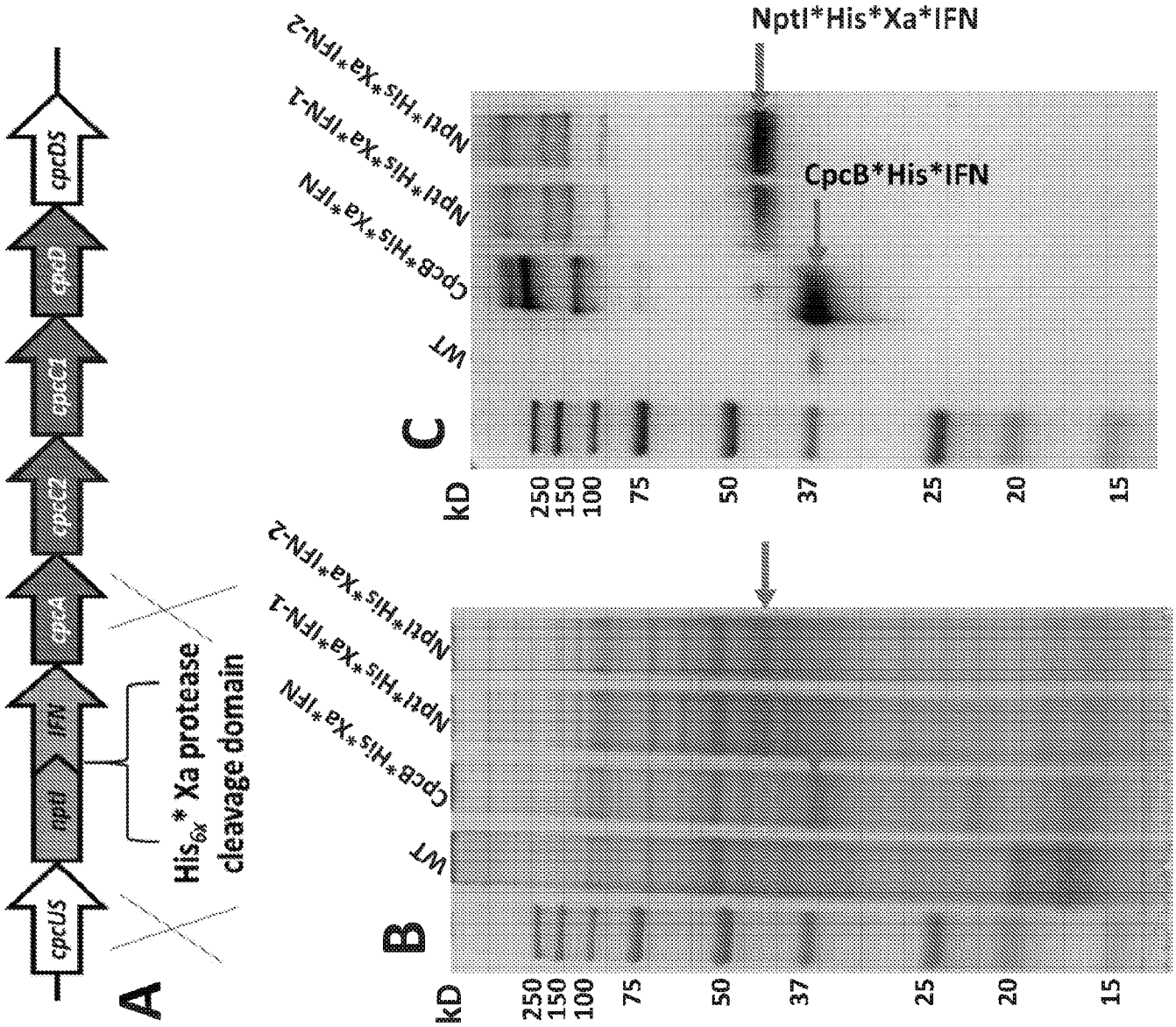

A
cpcUS
nptI
IFN
cpcA
cpcB2
cpcB1
cpcD
cpcDS
His6x* Xa protease cleavage domain
B
kD
250
150
100
75
50
37
25
20
15
WT
CpcB*His*Xa*IFN
NptI*His*Xa*IFN-1
NptI*His*Xa*IFN-2
C
NptI*His*Xa*IFN
CpcB*His*IFN

Fig. 13

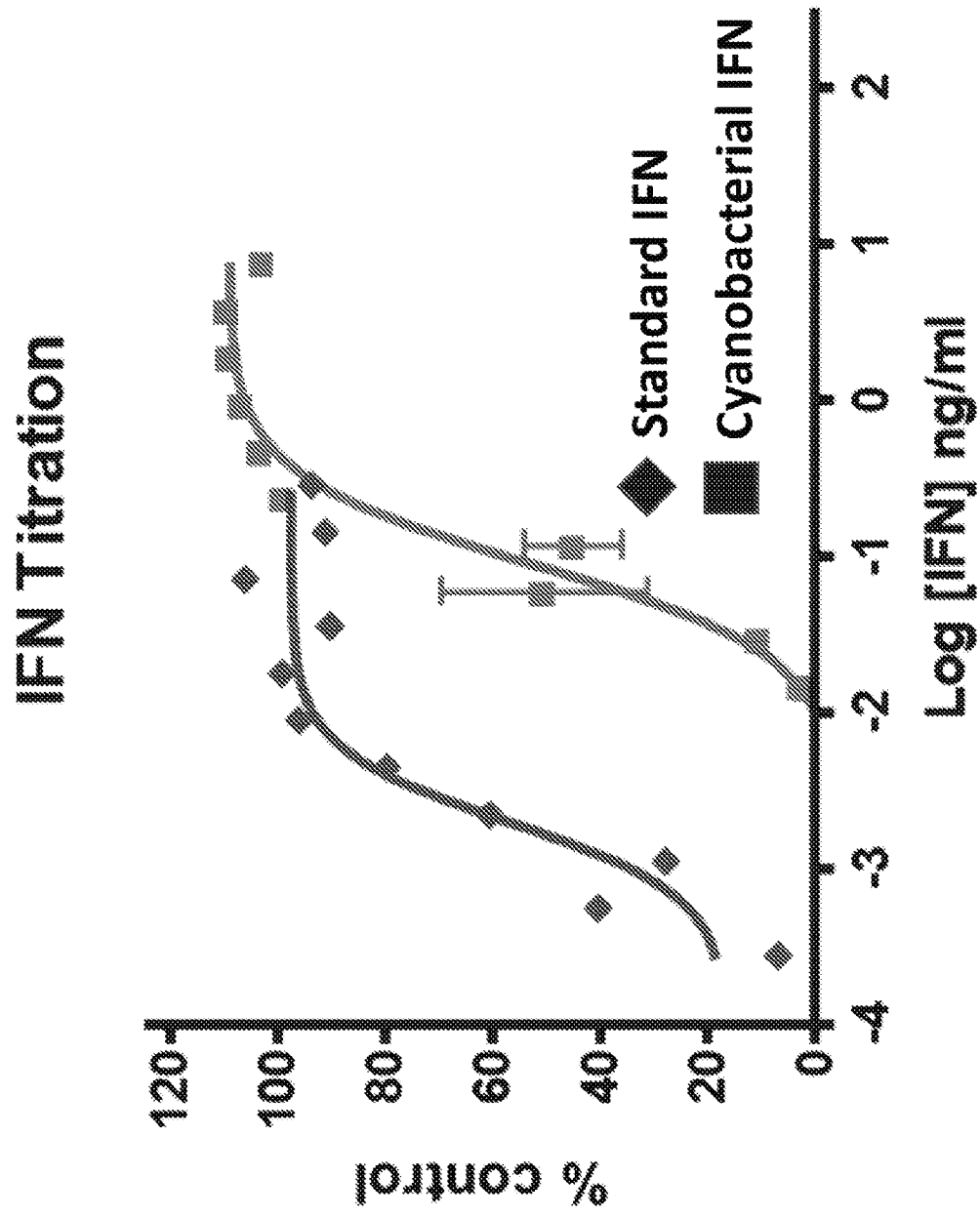
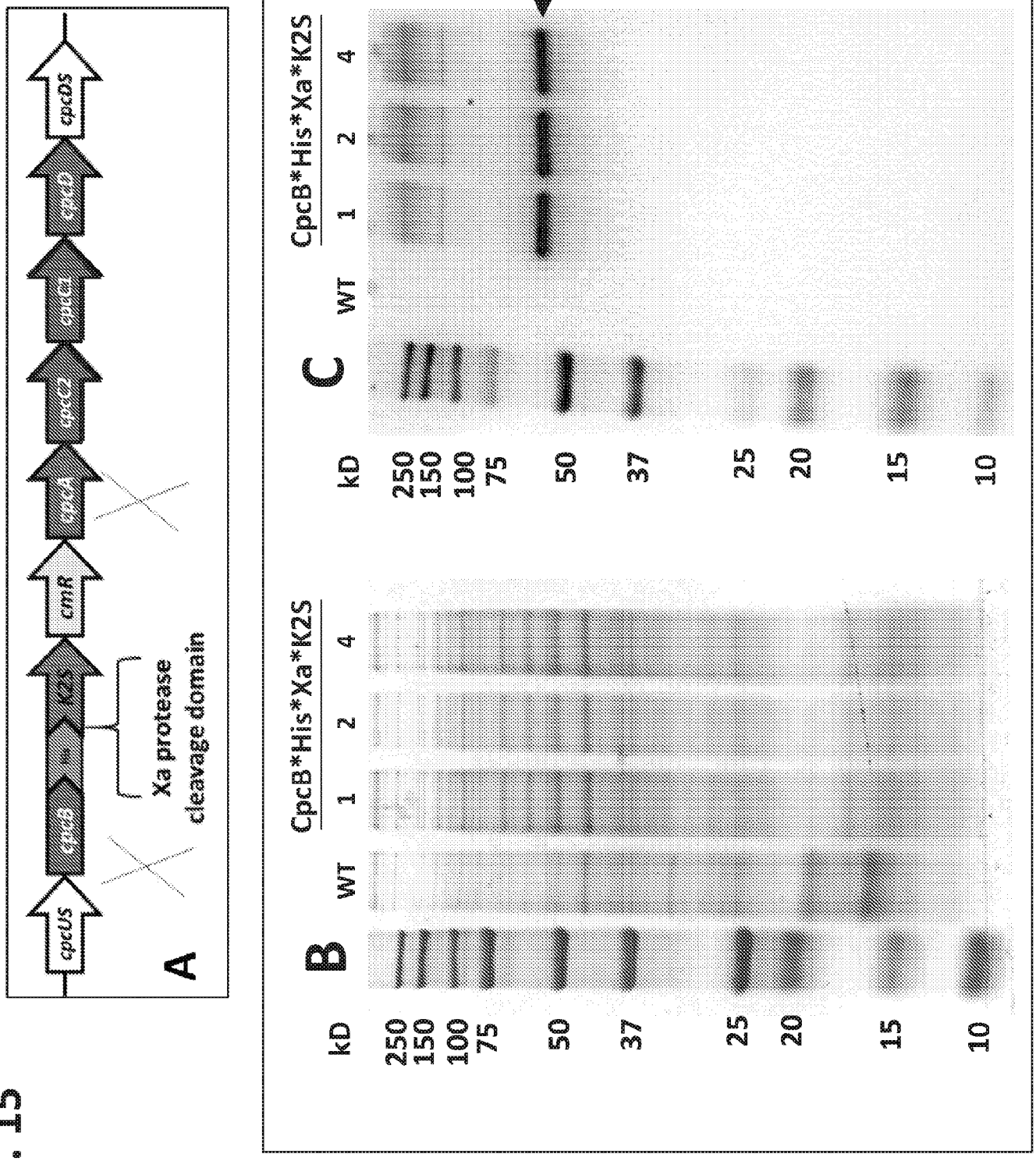

IFN Titration
Standard IFN
Cyanobacterial IFN
% control
120
100
80
60
40
20
0
-4
-3
-2
-1
0
1
2
Log [IFN] ng/ml

Fig. 14

Fig. 15


A
cpcUS
cpcB
K2S
cmR
cpcA
cpcG2
cpcG1
cpcD
cpcDS
Xa protease
cleavage domain
B
CpcB*His*Xa*K2S
WT
1
2
4
kD
250
150
100
75
50
37
25
20
15
10
C
CpcB*His*Xa*K2S
WT
1
2
4
58.9
kD
kD
250
150
100
75
50
37
25
20
15
10

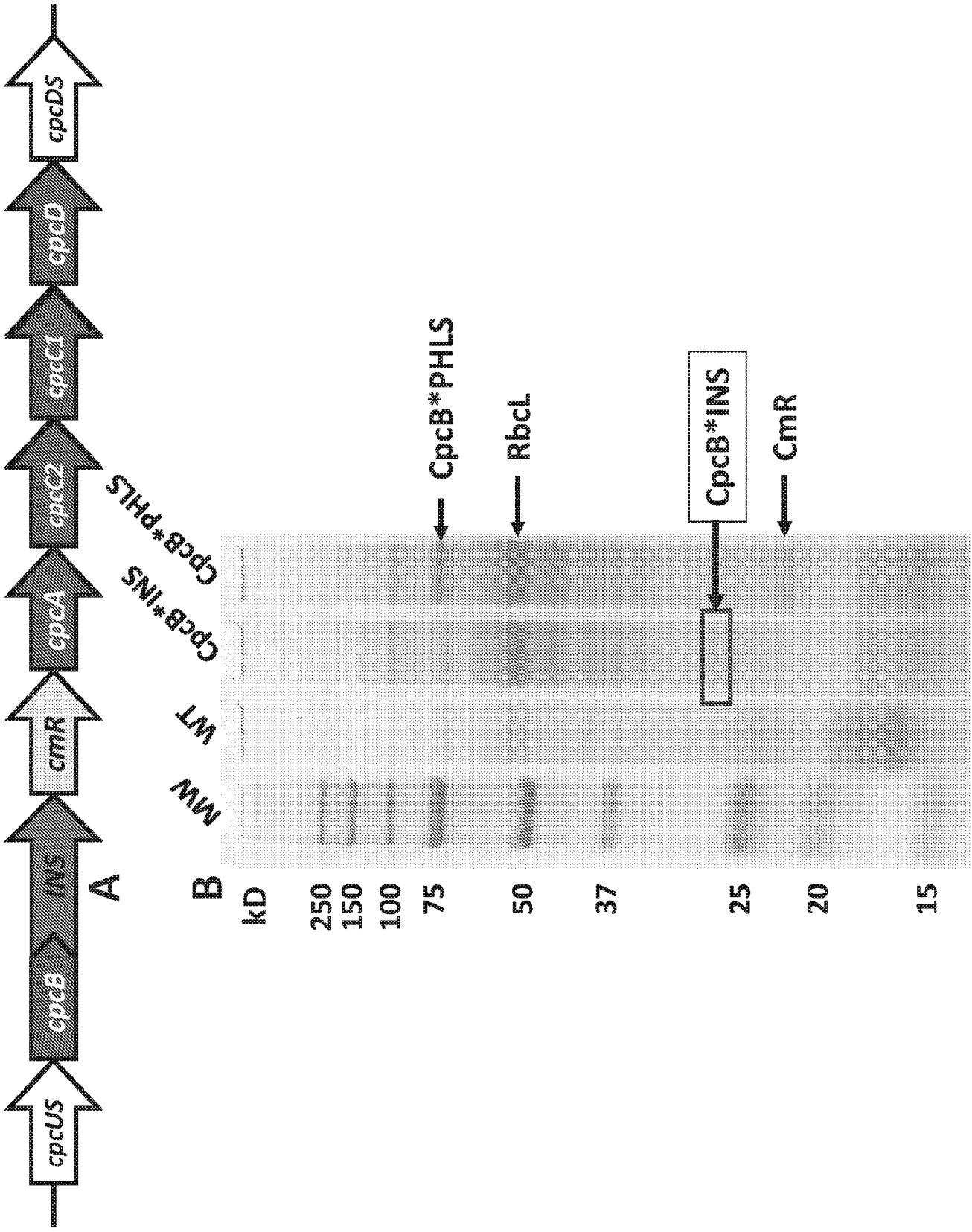
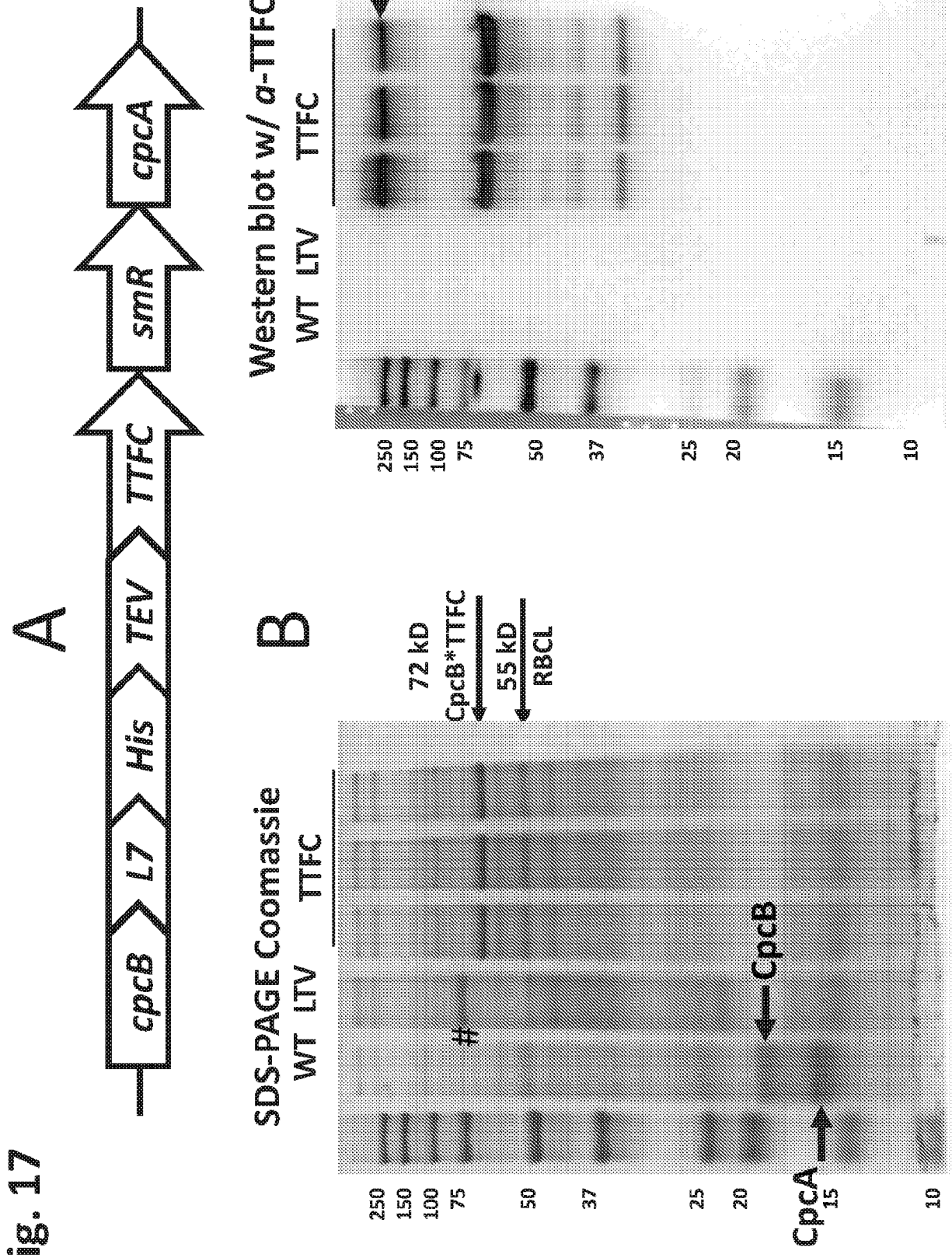

A
cpcUS
cpcB
INS
cmR
cpcA
cpcG2
cpcC1
cpcD
cpcDS
cpcB*PHLS
cpcB*INS
B
MW
WT
cpcB*INS
cpcB*PHLS
kD
250
150
100
75
50
37
25
20
15
RbcL
CpcB*PHLS
CpcB*INS
CmR

Fig. 16

Fig. 17


A
cpcB
L7
His
TEV
TTFC
smR
cpcA
B
SDS-PAGE Coomassie
WT LTV TTFC
72 kD
CpcB*TTFC
55 kD
RBCL
#
CpcB
CpcA
Western blot w/ α-TTFC
WT LTV TTFC
290 kD
72 kDa
CpcB*TTFC
250
150
100
75
50
37
25
20
15
10

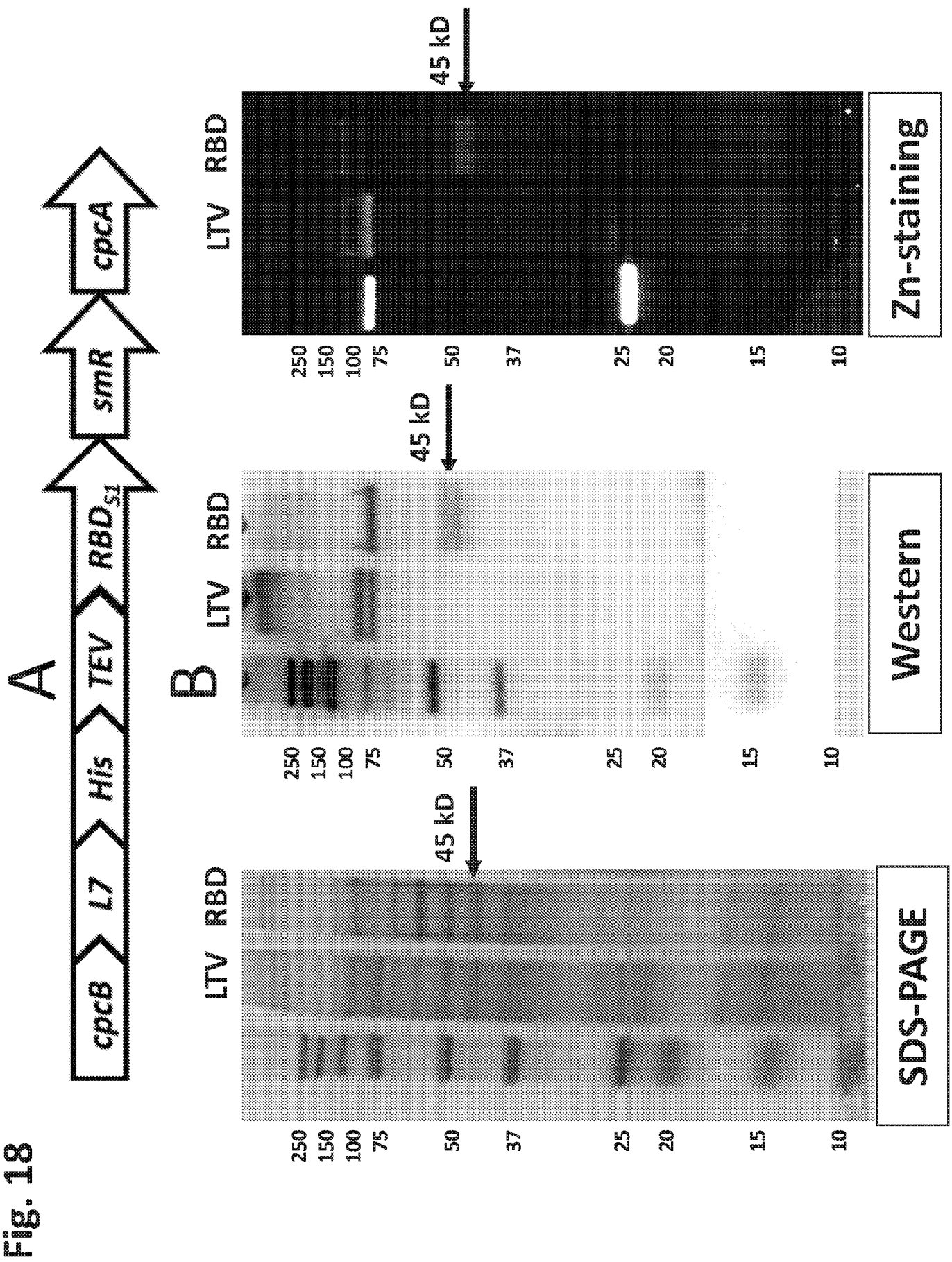

A
cpcB
L7
His
TEV
RBD$_{S1}$
smR
cpcA
B
LTV RBD
45 kD
250
150
100
75
50
37
25
20
15
10
SDS-PAGE
Western
Zn-staining

Fig. 18

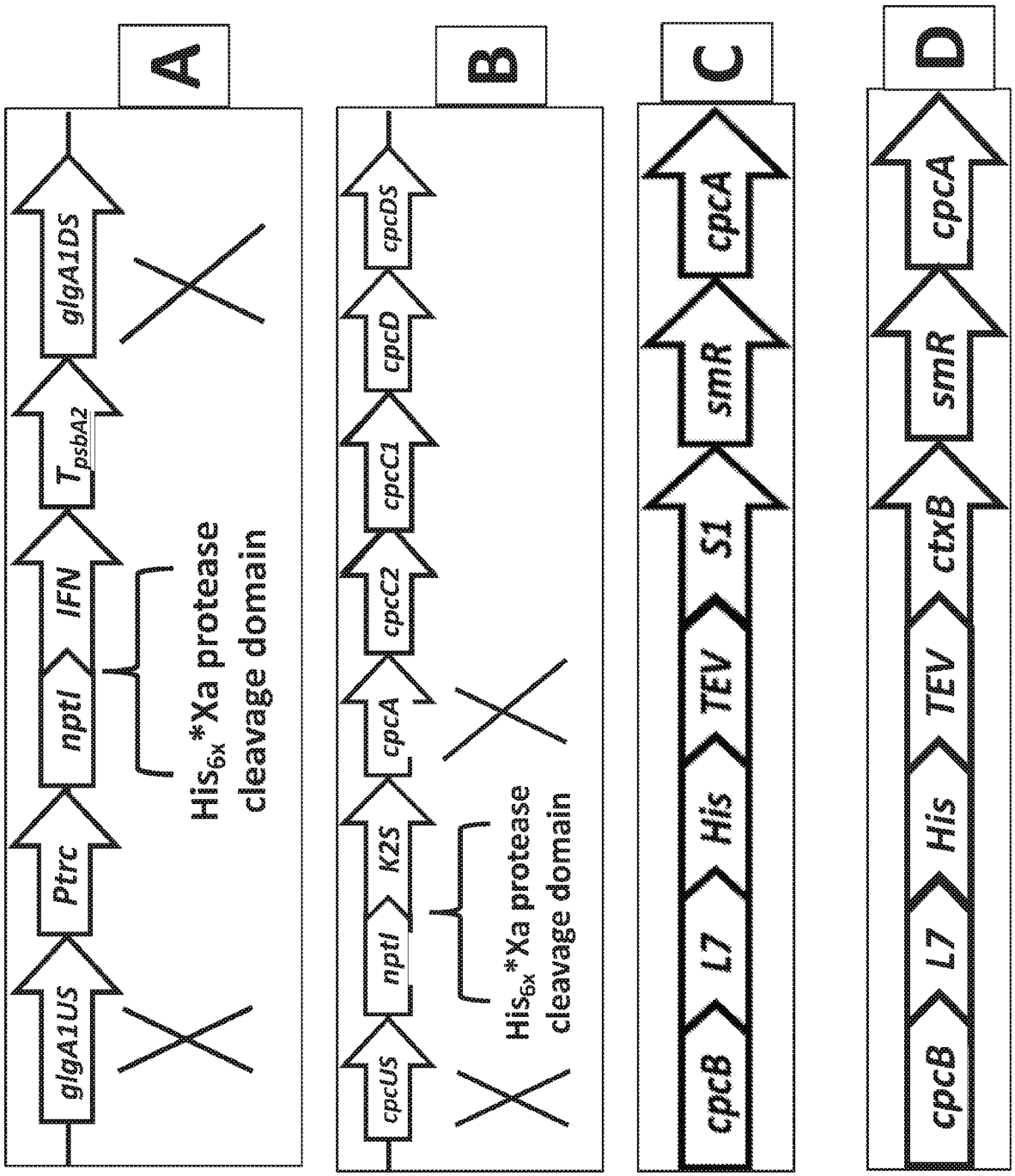

A
glgA1US
Ptrc
nptI
IFN
T_{psbA2}
glgA1DS
His_{6x}*Xa protease cleavage domain
B
cpcUS
nptI
K2S
cpcA
cpcC2
cpcC1
cpcD
cpcDS
His_{6x}*Xa protease cleavage domain
C
cpcB
L7
His
TEV
S1
smR
cpcA
D
cpcB
L7
His
TEV
ctxB
smR
cpcA

Fig. 19

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 20/50528

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12N 15/74, C12N 15/66, C12N 15/52 (2020.01)

CPC - C07K 14/4713, C07K 14/165, C07K 14/62, C12N 2510/02, C07K 14/195

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2018/0171342 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 21 June 2018 (21.06.2018). Especially para [0010]-[0011], [0020], [0034], [0048], [0070], [0097], Claim 36	1, 2, 33 --- 3-6, 34-35
Y	US 2004/0175359 A1 (DESJARLAIS et al.) 09 September 2004 (09.09.2004). Especially para [0022]-[0025], [0064]-[0069], [0075], Seq ID No: 2	3-6, 34-35
Y	US 2009/0011995 A1 (LEE et al.) 08 January 2009 (08.01.2009). Especially para [0158], [0239] -[0240], Abstract	6

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 January 2021	10 FEB 2021
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee Young Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 20/50528

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 21-30
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

----------Please see continuation in first extra sheet----------

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-6, 33-35, limited to an interferon polypeptide

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/US 20/50528

Continuation of Box No. III. Observations where unity of invention is lacking

Group I+, claims 1-20, 33-35, directed to a cyanobacterial host cell comprising an expression unit comprising a nucleic acid sequence comprising a transgene that encodes a biopharmaceutical protein, an isolated fusion protein, or a nucleic acid encoding the fusion protein. The cell, protein and nucleic acid will be searched to the extent that the biopharmaceutical transgene/protein encompasses an interferon polypeptide. It is believed that claims 1-6, 33-35 encompass this first named invention, and thus these claims will be searched without fee to the extent that the biopharmaceutical transgene/protein encompasses an interferon polypeptide. Additional biopharmaceutical transgene/protein(s) will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected biopharmaceutical transgene/protein(s). Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched. An exemplary election would be a biopharmaceutical transgene/protein comprising a tissue plasminogen activator (tPA) polypeptide, (claims 1, 2, 7-9, 33-35).

Group II, claims 31-32, directed to a method of engineering a cyanobacterial host cell to produce a biopharmaceutical protein.

The inventions listed as Groups I+ and II do not relate to a single special technical feature under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special technical features

Group I+ has the special technical feature of a composition comprising or consisting of a cyanobacterial host cell, an isolated fusion protein, or a nucleic acid encoding the fusion protein, that is not required by Group II.

Group II has the special technical feature of engineering a cyanobacterial host cell to produce a biopharmaceutical protein, the method comprising introducing an expression cassette into the cyaobacterial host cell by homologous recombination, that is not required by Group I+.

The inventions of Group I+ each include the special technical feature of a different biopharmaceutical transgene/protein, and is considered a distinct technical feature.

Common technical features

The inventions of Group I+ and Group II share the common technical feature of a cyanobacterial host cell comprising an expression unit comprising:
(i) a nucleic acid sequence comprising a transgene that encodes a biopharmaceutical protein, wherein the transgene is fused to the 3' end of a nucleic acid sequence that encodes a cyanobacteria beta-subunit of phycocyanin (cpcB) polypeptide to produce a fusion polypeptide comprises cpcB and the biopharmaceutical protein of interest;
(ii) a nucleic acid sequence encoding a cyanobacteria a -subunit of phycocyanin (cpcA) polypeptide; and
(iii) a nucleic acid sequence encoding a cyanobacterial cpcC1, cpcC2 and cpcD polypeptide.

The inventions of Group I+ further share the common technical feature of an isolated fusion protein comprising a biopharmaceutical protein fused to the 3' end of a nucleic acid sequence that encodes a cyanobacteria beta-subunit of phycocyanin (cpcB) polypeptide, and a nucleic acid encoding the fusion protein.

However, these shared technical features are previously disclosed by US 2018/0171342 A1 to the Regents of the University of California (hereinafter 'UC').

UC discloses a cyanobacterial host cell comprising an expression unit (para [0034] - "An "expression construct" or "expression cassette" as used herein refers to a recombinant nucleic acid construct, which, when introduced into a cyanobacterial host cell") comprising:
(i) a nucleic acid sequence comprising a transgene that encodes a biopharmaceutical protein (para [0070] - "A fusion construct of the invention may be employed to provide high level expression in cyanobacteria for any desired protein product...the transgene encodes a polypeptide pharmaceutical"), wherein the transgene is fused to the 3' end of a nucleic acid sequence that encodes a cyanobacteria beta-subunit of phycocyanin (cpcB) polypeptide to produce a fusion polypeptide comprises cpcB and the biopharmaceutical protein of interest (para [0011] - "an expression construct comprising a nucleic acid sequence encoding a transgene that is codon-optimized for expression in cyanobacteria fused to the 3' end of a leader nucleic acid sequence encoding a cyanobacteria protein that is expressed in cyanobacteria"; para [0010] - "the fusion of transgenic plant proteins to highly expressed endogenous cyanobacteria proteins, such as the CpcB beta-subunit of phycocyanin"; para [0070] - "A fusion construct...the transgene encodes a polypeptide pharmaceutical");
(ii) a nucleic acid sequence encoding a cyanobacteria alpha-subunit of phycocyanin (cpcA) polypeptide (para [0011] - "the leader nucleic acid sequence encodes a beta-subunit of phycocyanin (cpcB), an alpha-subunit of phycocyanin (cpcA)"); and
(iii) a nucleic acid sequence encoding a cyanobacterial cpcC1, cpcC2 and cpcD polypeptide (para [0020] - "RT-PCR products (cpcA-cpcC2, lanes b, d). Similarly, a cDNA fragment including the 3' end of cpcC1 and the 5' beginning sequence of cpcD was amplified with primers cpcC1-cpcD_Fw and cpcC1-cpcD_Rv (Table 1S)"; [0097]-[0098] "the cpcBPHLS fusion construct replaced the native cpcB sequence only, inserted upstream of the cpcA, cpcC2, cpcC1 and cpcD genes, denoted as cpcBPHLS+cpc").

As the technical features were known in the art at the time of the invention, they cannot be considered special technical features that would otherwise unify the groups.

Therefore, Group I+ and II inventions lack unity under PCT Rule 13 because they do not share the same or corresponding special technical feature.

NOTE, continuation of Item 4 above: claims 21-30 are held unsearchable because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).