



(51) International Patent Classification:
Not classified

SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(21) International Application Number:
PCT/US2024/037862

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(22) International Filing Date:
12 July 2024 (12.07.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/513,497 13 July 2023 (13.07.2023) US

(71) Applicants: **CZ BIOHUB SF, LLC** [US/US]; 499 Illinois Street, 14th Floor, San Francisco, CA 94158 (US). **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, CA 94607-5200 (US).

(72) Inventors: **LANDRY, Markita**; 1111 Franklin Street, 12th Floor, Oakland, CA 94607-5200 (US). **WANG, Jeffrey, Wei-Ting**; 1111 Franklin Street, 12th Floor, Oakland, CA 94607-5200 (US). **SQUIRE, Henry, J.**; 1111 Franklin Street, 12th Floor, Oakland, CA 94607-5200 (US).

(74) Agent: **BRUSCA, Eric, M.**; Marshall, Gerstein & Borun LLP, 233 S. Wacker Drive, 6300 Willis Tower, Chicago, IL 60606-6357 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE,

(54) Title: CELL-PENETRATING PEPTIDES FOR NUCLEIC ACID AND PROTEIN DELIVERY IN PLANTS

(57) Abstract: The present disclosure provides plant-derived cell-penetrating peptides to enable biomolecule cargo delivery in plants with efficiencies that surpass prior CPPs. It further discloses a method for sequence alignment of plant homeodomain proteins to discover plant-specific CPPs that function superior to conventional CPPs. In particular, fragments of plant homeodomain proteins, including transcription factors, are described that are useful for transporting biomolecules to plant cells without the use of biolistics or other exogenous delivery agents.

CELL-PENETRATING PEPTIDES FOR NUCLEIC ACID AND PROTEIN DELIVERY IN PLANTS**CROSS-REFERENCE TO RELATED APPLICATIONS**

[0001] The present application claims priority to U.S. Provisional Patent Application No. 63/513,497, filed on July 13, 2023, the entirety of which is incorporated by reference herein.

REFERENCE TO THE SEQUENCE LISTING

[0002] This application includes a sequence listing submitted electronically, in a file entitled "59166_Seqlisting.xml," created on July 12, 2024 and having a size of 143,243 bytes, which is incorporated by reference herein.

FIELD

[0003] The present disclosure relates to plant-derived, cell-penetrating peptides and uses thereof.

BACKGROUND

[0004] Plant genetic engineering is an important tool for improving crop yield, quality, and resistance to abiotic/biotic stresses for sustainable agriculture, among other uses. Plant genetic engineering involves two main steps: plant transformation and the regeneration of transformed plants. Transformation is the process of introducing DNA, RNA, and proteins into plant cell/tissue. However, plant cells have cell walls that especially impede the delivery of cargoes such as protein. The development of CRISPR-Cas9 (Jinek, M et al. (2012) *Science*. 337(6096):816-821) and other DNA editing tools (Li, H et al. (2020) *Mol Plant*. 13(5):671-674) has only increased the need for working protein delivery tools in plants, which could accelerate basic research, spawn novel agricultural biologic agents, or potentiate DNA-free gene editing of plants. Recent discoveries in morphogenic transcription factors that accelerate plant regeneration evince a new class of possible protein cargoes if these proteins could be delivered (Lowe, K et al. (2016) *Plant Cell*. 28(9):1998-2015). These motivations have led researchers to develop novel nanoparticle-based strategies for the delivery of biomacromolecules to walled plant cells. For example, multiple technologies have been developed to deliver siRNA to plants using diverse vehicles such as single walled carbon nanotubes (Demirer, GS et al. (2020) *Sci Adv*. 6(26): eaaz0495), DNA nanostructures (Demirer, GS et al. (2019) *Nat Nanotechnol*. 14(5):456-464), carbon dots (Schwartz, SH et al. (2020) *Plant Physiol*. 184(2):647-657), and gold nanoparticles (Zhang, H et al. (2022) *Nat Nanotechnol*. 17(2):197-205).

[0005] Cargo delivery can also be accomplished using cell-penetrating peptides (CPPs), which are short peptides that facilitate the transport of cargo molecules through the plasma membrane to the cytosol. In most cases, CPPs are coupled to cargo molecules through non-covalent conjugation, forming CPP-cargo complexes. To date, DNA, RNA, and proteins such as antibodies have been reported as cargo molecules that can be combined with CPP for delivery across the cell plasma membrane. Protein delivery is particularly challenging as their function is dependent on secondary and tertiary structures which can be disturbed by delivery vehicles. Most studies of CPP-protein complexes have contributed to applications in mammalian cells, which do not have a cell wall, whereas few studies have focused on plant cells. Although the cell wall may be permissible to materials below the size exclusion limit of 5-10nm or proteins around 50-100 kDa (Read and Bacic 1996), few have demonstrated delivery of proteins using cell penetrating peptides in plants (Chang, M et al. (2007) *New Phytol.* 174(1):46-56; Guo, B et al. (2019) *PLoS One.* 14(7): e0214033). Additionally, the surface charges of proteins differ from one another, such that the delivery of some protein cargoes with existing charged CPPs may be possible but not other protein cargoes. Furthermore, existing CPPs have been developed and optimized for cargo delivery to mammalian cells, making it challenging to use these CPPs for delivery in plants. Therefore, despite these proof-of-principle advances, protein delivery to walled plant cells remains largely dependent on biolistic delivery, which requires protein dehydration (and thus potential inactivation) to a gold particle surface and forceful and injurious rupture of plant membranes to accomplish delivery in a low throughput and low efficiency manner (Hamada, H et al. (2018) *Sci Rep.* 8(1):14422). Thus, protein delivery to plants presents a major bottleneck for plant genetic engineering and there is need to develop new strategies for plant transformation with ease, robustness, and high efficiency that protein delivery could provide.

[0006] Another key barrier to the use of nanotechnologies for plant biomolecule delivery, and specifically cell penetrating peptides for protein delivery to plants, is the lack of quantitative validation of successful intracellular protein delivery. The lack of protein delivery validation methods makes it difficult to unilaterally distinguish successful protein delivery from artefact, lytic sequestration, or quantitatively compare delivery efficiency of different peptides (Wang, JW et al. (2021) *Curr Opin Plant Biol.* 60:102052).

[0007] This lack of tools to deliver and successfully quantify protein delivery in plants is due to the near universal dependence of confocal microscopy to validate delivery of fluorescent proxy cargoes. However, confocal microscopy in plant tissues poses a set of unique problems that make it challenging to distinguish artefact from signal and make absolute quantification of signal impossible. Aerial tissues of plants are heterogeneous, highly light scattering, and possess intrinsic auto-fluorescence (Donaldson L. (2020)

Molecules. 25(10):2393.), which makes it difficult to distinguish signal from noise. Furthermore, unlike mammalian cells, the plant cell cytosol in the majority of cell types is highly compressed against the cell wall by the plant's large central vacuole, making unambiguous imaging of cytosolic contents challenging due to the small surface area of cytosolic contents (Serna L. (2005) *New Phytol.* 165(3):947-952). In addition, the plant cell is surrounded by a porous and adsorbent cellulosic wall that is 100-500nm thick (Sugiura D, Terashima I, Evans JR. (2020) *Plant Physiol.* 183(4):1600-1611) which spans the Rayleigh diffraction resolution limit of visible light imaging and the axial resolution of most confocal microscopes. Together, the small cytosolic volume which is proximal to the cell wall makes it impossible to distinguish – with the necessary spatial precision – the location of fluorescent cargoes near versus imbedded in the cell wall, or inside the cell cytosol (Zhang, H et al. (2022) *Nat Nanotechnol.* 17(2):197-205.), without super-resolution microscopy (Pawley, J. B. (2006). *Handbook of Biological Confocal Microscopy*. Boston, MA, Springer). Additionally, free fluorophore from cargo degradation (Lacroix, A et al. (2019) *ACS Cent Sci.* 5(5), 882–891.) or endosomal entrapment of cargoes would contribute to measured fluorescence intensity and intracellular colocalization in plants but fail to correlate with successful delivery. For these reasons, gauging cellular uptake of cargoes based solely on confocal microscopy data of fluorophore-tagged cargo in plants does not confirm successful intracellular delivery nor provide quantitative data for effective uptake. These barriers have made biomacromolecule delivery in plants, particularly protein delivery, exceptionally challenging.

[0008] There is therefore a need for new materials and methods to deliver biomolecules to plant cells. There is also a need to develop a versatile, unambiguous platform to confirm the delivery of proteins of varying sizes in walled plant tissues.

SUMMARY

[0009] The present disclosure provides compositions and methods for delivering cargoes to plant cells.

[0010] In one embodiment, the present disclosure provides a composition comprising a plant-derived cell-penetrating peptide (CPP).

[0011] In another embodiment, the composition comprising a plant-derived CPP comprises an amino acid sequence derived from a plant homeodomain protein. In one embodiment, the amino acid sequence comprises 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 amino acids. In one embodiment, the homeodomain protein is a transcription factor. In some embodiments, the homeodomain protein comprises a sequence derived from WUSCHEL, CLSY2, PINTOX, LUMI, NDX, PHD, RLT2, BEL1, KNAT1, ZHD3, ATB16, HDG7, HAT1, REV, WOX3B, HOX12, Q8LLD8, WOX, Q40238,

Q69G85 or CMJ244C. In some embodiments, the CPP is selected from the group consisting of: AtWUS_1 (SEQ ID NO: 1), AtWUS_2 (SEQ ID NO: 2), AtWUS_3 (SEQ ID NO: 3), At_SAW (SEQ ID NO: 4), At_PINTOX (SEQ ID NO: 5), At_LD (SEQ ID NO: 6), At_NDX (SEQ ID NO: 7), At_PHD (SEQ ID NO: 8), At_DDT (SEQ ID NO: 9), At_BEL (SEQ ID NO: 10), At_KNOX (SEQ ID NO: 11), At_PLINC (SEQ ID NO: 12), At_ZIPI (SEQ ID NO: 13), At_ZIPIV (SEQ ID NO: 14), At_ZIPII (SEQ ID NO: 15), At_ZIPIII (SEQ ID NO: 16), Zm_WOX (SEQ ID NO: 17), Os_HOX12 (SEQ ID NO: 18), St_PHD (SEQ ID NO: 19), Nt_WOX (SEQ ID NO: 20), Sp_DDT (SEQ ID NO: 21), Lj_NDX (SEQ ID NO: 22), Cm_BEL (SEQ ID NO: 23), Pp_ZIPIII (SEQ ID NO: 24), Cr_KNOX (SEQ ID NO: 25), Ot_WOX (SEQ ID NO: 26), Sm_ZIPIII (SEQ ID NO: 27), Pa_KNOX (SEQ ID NO: 28), Cl_WOX (SEQ ID NO: 29), Gg_BEL (SEQ ID NO: 30), Pt_WOX (SEQ ID NO: 31), and Cp_ZIPIII (SEQ ID NO: 32).

[0012] In one embodiment, the CPP comprises an amino acid sequence selected from the group consisting of SEQ ID Nos: 1-23 or a fragment thereof, analog or derivative thereof.

[0013] In one embodiment, the composition comprising a CPP comprises an amino acid sequence selected from the group consisting of: (SEQ ID NO: 1) KNVFWFQNHKARERQKKRFN; (SEQ ID NO: 2) KNVFWFQNHKARERQ; and (SEQ ID NO: 3) HKARERQ.

[0014] The present disclosure also provides a polynucleotide encoding a CPP. In one embodiment, the polynucleotide is selected from the group consisting of: a polynucleotide encoding a peptide having at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID Nos: 1-23; and a polynucleotide encoding a peptide comprising the amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 2 or SEQ ID NO: 3.

[0015] The present disclosure further provides a complex for delivery of a biomolecule inside a cell comprising: a CPP; and a biomolecule wherein the biomolecule is fused at the N-terminal or C-terminal of the CPP. In one embodiment, the biomolecule is selected from the group consisting of a chemical compound, a protein or fragment thereof, a recombinant protein or fragment thereof, a glycoprotein or fragment thereof, a peptide, an antibody or fragment thereof, an enzyme or fragment thereof, a nuclease or fragment thereof, a hormone or fragment thereof, a cytokine or fragment thereof, a transcription factor or fragment thereof, a toxin or fragment thereof, a nucleic acid, a carbohydrate, a lipid, a glycolipid, a drug, a fluorophore, a fluorescent protein or fragment thereof, an antibiotic, a recombinase, and a plant hormone.

[0016] In one embodiment, the biomolecule is fused via a linker. In one embodiment, the linker is a GSGS linker.

[0017] In still another embodiment, the biomolecule is selected from the group consisting of CRISPR associated protein 9 (CAS9), CAS12, CAS13, CAS14, CAS variants, CxxC-finger protein-1 (Cfp1), zinc-finger nucleases (ZEN) and transcription activator-like effector nuclease (TALEN). In another embodiment of the present disclosure, the nucleic acid is selected from the group consisting of DNA, RNA, antisense oligonucleotide (ASO), microRNA (miRNA), small interfering RNA (siRNA), aptamer, locked nucleic acid (LNA), peptide nucleic acid (PNA), and morpholino.

[0018] In one embodiment, the recombinant protein is selected from the group consisting of a morphogenic protein, a growth factor, a receptor, a signaling protein, a membrane protein, and a transmembrane protein. In one embodiment, the nuclease is an RNA-guided endonuclease, a CRISPR endonuclease, a type I CRISPR endonuclease, a type II CRISPR endonuclease, a type III CRISPR endonuclease, a type IV CRISPR endonuclease, a type V CRISPR endonuclease, a type VI CRISPR endonuclease, CRISPR endonuclease, CRISPR associated protein 9 (Cas9), Cpf1, a zinc-finger nuclease (ZFNs), a Transcription activator-like effector nucleases (TALENs), a homing endonuclease, or a meganuclease. In some embodiments, the nuclease is a CRISPR endonuclease further comprising a guide RNA, a crRNA, a tracrRNA, or both a crRNA and a tracrRNA.

[0019] In one aspect, the present disclosure provides a method for delivering a biomolecule into a plant cell. In one embodiment, the method comprises contacting said plant cell with a composition or complex of the present disclosure under conditions that allow the CPP to penetrate the plant cell.

[0020] In still another embodiment, the present disclosure provides a method of identifying a CPP capable of transporting a biomolecule to a subcellular location in a plant.

[0021] In one embodiment, the method comprises the steps of: a) fusing a biomolecule of interest to a CPP; b) contacting a plant cell with a composition comprising the CPP of (a); c) performing an assay to determine the ability of the CPP to translocate the biomolecule to a subcellular location of the cell.

[0022] In one embodiment, the biomolecule is a fluorophore and the assay is a delivered complementation *in planta* (DCIP) assay.

[0023] The present disclosure further provides a method of quantitating protein delivery comprising the steps of: a) fusing a protein of interest to a CPP; b) contacting a plant cell with a composition comprising the CPP-protein fusion of (a); and c) performing an assay to quantitate the amount or number of CPP-protein fusion to translocate a subcellular location of the cell.

[0024] In one embodiment, the protein is a fluorophore and the assay is a delivered complementation *in planta* (DCIP) assay.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] **Figure 1** is a schematic diagram of the of DCIP platform and cytoDCIP constructs. In both constructs, expression of mCherry-sfGFP1-10 fusion protein is driven by the 35S constitutive promoter from the cauliflower mosaic virus (CaMV) and terminated by the tNOS terminator from the *Agrobacterium tumefaciens* nopaline synthase gene. The DCIP construct further possess a SV40 nuclear localization signal (NLS) for targeting the fusion protein to the nucleus. The cytoDCIP construct produces a similar fluorescent signal yet does not contain a nuclear localization signal and thus localizes to the cytosol. Both constructs were assembled as level-1 assemblies in GoldenbraiD 2.0.

[0026] **Figure 2** demonstrates how DCIP can confirm and quantify delivery of proteins with cell penetrating peptides. **Figure 2A** is a schematic for a DCIP workflow as follows: (i) agroinfiltration of *N. benthamiana* leaves with the DCIP constructs, (ii) infiltrating the agroinfiltrated leaves from step (i) with the GFP11 peptide (that is fused to cargo and a CPP), three days post infiltration (d.p.i). During this incubation step, GFP11 is internalized into plant cells and if cytosolic delivery occurs, GFP11 is able to complement to GFP1-10 and sfGFP fluorescence is recovered (depicted here as step (iii)), (iv) Post-incubation, leaf discs are imaged and analyzed using Cell Profiler by using mCherry fluorescence to identify cells via their fluorescent nuclei, (v) The sfGFP fluorescence is normalized to mCherry fluorescence to account for variability in DCIP expression and the number of GFP positive cells relative to the total number of mCherry positive cells is determined as an analog to delivery efficiency. **Figure 2B** provides confocal micrographs of representative maximum intensity projection of a leaf disc expressing DCIP infiltrated with water. sfGFP fluorescence is pseudocolored green (left) and two-color overlay with mCherry fluorescence, pseudocolored magenta, resulting in a white appearance after overlay (right). mCherry expressing cells possess nuclei presenting as small, round fluorescent bodies amenable to automated image analysis. Orthogonal projections demonstrate depth of imaging in leaves. **Figure 2C** provides confocal micrographs of an equivalent DCIP expressing leaf infiltrated with 100 μ M R9-GFP11 (nona-arginine cell penetrating peptide fused to GFP11) showing robust sfGFP complementation 4 hours post infiltration with the sfGFP colocalizing with mCherry (left panel), moreover the delivery capability was shown to extend throughout the full thickness of the leaf tissue. Scale bar represents 100 μ m. **Figure 2D** is an image of a Western blot of *N. benthamiana* leaf lysates 3 d.p.i. probed with an anti-mCherry primary antibody and *N. benthamiana* leaf lysates 3 d.p.i. with either DCIP or cytoDCIP showing both protein fusions were expressed at the predicted molecular weight. **Figure 2E** provides

micrographs of a DCIP expressing plant infiltrated with either 100 μ M R9-GFP11 (top three panels) or water control (bottom three panels) and imaged using confocal microscopy at 4-5 hours using a 5x objective. mCherry fluorescence is pseudocolored magenta and sfGFP fluorescence is colored green. Chloroplast autofluorescence is colored blue. Scale bar is 1mm.

[0027] **Figure 3** shows validation of DCIP quantification of cargo delivery using a range of R9-GFP11 concentrations. **Figure 3A** provides column scatter plots of representative green/red ratio against concentrations of infiltrating R9-GFP11 in *N. benthamiana* expressing DCIP following a 4-5 hour incubation. Each data point represents the relative fluorescence from sfGFP caused by delivered GFP11-sfGFP1-10 complementation and mCherry expression in a single nucleus, successful delivery of the GFP11 cargo results in an increase of the green/red ratio. **Figure 3B** provides bar-dot plots of mean green/red ratio averaged across seven plants as experimental repeats (N=7). Error bars represent standard deviation of the group repeats. Statistical comparisons between each treatment condition and the non-treated control with a Kruskal-Wallis test in combination with Dunn's multiple comparison's test. **Figure 3C** left panel provides bar-dot plots of percentage of GFP positive cells averaged across N=7 biological repeats against R9-GFP11 concentrations, **Figure 3C** right panel provides bar-dot plots of the corresponding cargo delivery efficiency normalized to the 100 μ M R9-GFP11 treatment group. **Figure 3D** provides column scatter plots of representative green/red ratios after treating leaf discs for 4-5 hour against the treating R9-GFP11 concentrations. **Figure 3E** provides bar-dot plots of mean green/red ratio after treating leaf discs for 4-5 hour and averaged across six biological repeats (N=6) against concentrations of R9-GFP11. **Figure 3F** left panel provides bar-dot plots of average percent GFP positive cells against R9-GFP11 concentrations, **Figure 3F** right panel provides bar-dot plots of normalized delivery efficiency of leaves treated with 100-1000 μ M R9-GFP11 for 4-5 hours (N=6), * denotes $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$ and ns = $p > 0.05$. **Figure 3G** provides confocal micrographs of representative single color maximum intensity projections of leaves infiltrated with 0-1000 μ M R9-GFP11 and incubated for 4-5 hours. Nuclei exhibiting delivered complementation appear as round, green objects. Scale bar is 100 μ m.

[0028] **Figure 4** provides confocal micrographs of representative two-color maximum intensity projections of DCIP expressing leaves infiltrated with 0-1000 μ M R9-GFP11 following a 4-5 hour incubation. Scale bar is 100 μ m. mCherry is pseudocolored magenta and sfGFP is pseudocolored green. Overlay results in white coloration.

[0029] **Figure 5** demonstrates the use of DCIP as a CCP screening platform. **Figure 5A** is a table showing common and previously-known CPPs that were tested here as delivery tools for GFP11 and their corresponding sequence and molecular weights (kDa). CPP

sequences are bolded while a flexible GSGS linker connecting the CPP to the GFP11 cargo is underlined. **Figure 5B** left panel provides bar-dot plots of percent GFP positive nuclei of leaf discs incubated with 100 μ M of varying CPP-GFP11 conjugates for 4-5 hours (N=6). **Figure 5B** right panel provides bar-dot plots of the normalized delivery efficiency of leaf discs incubated with 100 μ M with varying CPP-GFP11 conjugates for 4-5 hours (N=6) normalized to R9 (nona-arginine). **Figure 5C** provides representative confocal micrographs of sfGFP fluorescent nuclei as the result of successful GFP11 delivery. White arrows point toward enhanced nucleolar localization of DCIP after delivery using arginine-rich CPP. **Figure 5D** left and right panels provide bar-dot plots of average percent GFP positive and normalized delivery efficiency in leaf discs treated with 100 μ M R9-GFP11 and either left at room temperature or kept at 4°C for 4-5 hours (N=5) respectively. **Figure 5E** provides a bar-dot plots of average normalized delivery intensity in leaf discs treated with 100 μ M R9-GFP11 and either left at room temperature or kept at 4°C for 4-5 hours (N=5). Delivery intensity was calculated by normalizing the mean green/red ratio of the 4°C treatment to that of the room temperature treatment. For normalized results of the low-temperature treatment, a one-sample t-test comparing to the ideal value of 1.0 was used. **Figure 5F** is a bar-dot plot of normalized delivery efficiency in leaves infiltrated with 100 μ M R9-GFP11 and co-infiltrated with either DMSO, 10 μ M ikarugamycin, or 40 μ M wortmannin for 4-5 hours (N=7). Unless otherwise indicated, Kruskal-Wallis test followed by Dunn's multiple comparisons test was performed for all statistical comparisons where ns = $p > 0.05$, * = $0.01 < p < 0.05$, and ** = $p < 0.01$. Scale bar is 100 μ m.

[0030] **Figure 6** demonstrates the low cell penetrating efficiency of the BP100-GFP11 construct, demonstrating that BP11, while a CPP that has shown high protein efficiency in mammalian cells, is ineffective for protein delivery in plants. This is a confocal micrograph of an example maximum intensity projection of sfGFP complementation as resulting from 100 μ M BP100-GFP11 incubation in a DCIP expressing leaf disc for 4-5 hours. Scale bar is 100 μ m. sfGFP fluorescent nucleus is marked by a white triangle. mCherry is pseudocolored magenta, sfGFP is pseudocolored green, and chloroplast autofluorescence is pseudocolored blue. Overlay results in white coloration.

[0031] **Figure 7** provides images of Coomassie-stained SDS polyacrylamide gels analyzing the purification of various proteins used in this study. **Figure 7A** is an image of an SDS-PAGE of purified GFP11-AtWUS-R9 exchanged into buffer P by dialysis or by desalting column. **Figure 7B** is an image of an SDS-PAGE of purified GFP11-AtWUS-R9 (batch #2) and GFP11-AtWUS with no CPP. **Figure 7C** is an image of an SDS-PAGE of first four recombinant proteins used in this study. Red triangles indicate the protein of interest.

Proteins were stained with Coomassie R-250. Figure 8D is an image of an SDS-PAGE of the helix-3 deletion WUS mutant, GFP11-AtWUS- $\Delta\alpha 3$.

[0032] **Figure 8** demonstrates DCIP mediated delivery of the morphogenic transcription factor, WUSCHEL in *N. benthamiana* leaves. **Figure 8A** are confocal micrographs of representative maximum intensity projection of a DCIP expressing *N. benthamiana* leaf infiltrated with 140 μ M GFP11-AtWUS and incubated as a leaf disc for 6H. Successful delivery presents as sfGFP (green pseudocolor) and mCherry (magenta pseudocolor) fluorescent nuclei as the result of delivered complementation. **Figure 8B** is a bar-dot plot showing quantification of GFP positive nuclei as a result of GFP11-AtWUS delivery in five plants (N=5) or buffer control. **Figure 8C** is a sequence alignment of AtWUS within several other plant and animal homeodomain transcription factors centered around putative conserved cell penetrating helix using Clustal Omega and ESPRIPT. Conserved residues are highlighted red, while similarly charged residues are boxed in blue. The tested AtWUS derived CPP (WUSP) is underlined in red. **Figure 8D** is a bar-dot plot of quantitative microscopy DCIP results for *N. benthamiana* treated with either 100 μ M R9-GFP11 or 100 μ M WUSP-GFP11 and imaged at 4-5 H post infiltration. Statistical analysis performed with T-test comparison and Holm-Šídák correction for multiple comparisons where ns = $p > 0.05$, * = $0.01 < p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

[0033] **Figure 9** demonstrates DCIP verified delivery of full-length AtWUS. **Figure 9A** shows confocal micrographs of representative maximum intensity projections of a cytoDCIP expressing plant infiltrated with either buffer or 140 μ M GFP11-AtWUS after 6 hours. Successful delivery and native nuclear-localization is indicated by presence of nuclear-localized sfGFP fluorescence (pseudocolored green) in contrast to mCherry (magenta) being localized everywhere within the cell. These results demonstrate that AtWUS can enter plant cells without a CPP like R9 and suggest that AtWUS itself or components or derivatives or fragments thereof are inherently cell penetrating in plants. Chloroplasts are pseudocolored blue. Scalebar is 100 μ m. **Figure 9B** shows confocal micrographs of maximum intensity projections of DCIP expressing plant infiltrated with 100 μ M WUSP-GFP11 and imaged at 4-5 hours post infiltration. Successful delivery is indicated by presence of nuclear-localized sfGFP fluorescence (pseudocolored green). mCherry and chloroplast autofluorescence are pseudocolored magenta and blue respectively. **Figure 9C** is a bar-dot plot of DCIP quantification for plants infiltrated with 140 μ M GFP11-AtWUS- $\Delta\alpha 3$ ($\Delta\alpha 3$) (N=5) and imaged at 6 hours post infiltration. Red highlighted point was identified as an outlier using Grubb's outlier test ($\alpha = 0.05$). Statistical analysis was done using an unpaired T-test with $p > 0.05$ = ns. **Figure 9D** is a bar-dot plot of data from panel C replotted without outlier with newly calculated mean and statistical comparison. **Figure 9E** is a bar-dot of an RT-qPCR analysis

of downstream AtWUS genes in 12-day old *Arabidopsis* seedlings treated with 3 μ M GFP11-AtWUS for 24H. 8-10 seedlings were treated per well and four wells were utilized for each treatment (N=4). These results show that AtWUS was successfully delivered and remains functional post-delivery such that it modulates gene expression as expected. **Figure 9F** is a bar-dot plot of a statistical comparison showing measured $\Delta\Delta$ CT values of GFP11-AtWUS treated seedlings compared to the buffer control. Statistical analysis performed with T-test comparison controlled for false discovery rate using the method by Benjamini, Krieger, and Yekutieli where ns = $p > 0.05$, * = $0.01 < p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$. **Figures 9G and H** are bar-dot plots from repeat experiments of the data displayed in **Figures 9E and F**. **Figure 9I** illustrates the phenotype of rice callus tissue mock treated, treated with 0.5 μ M of WUS2-R9, or treated with 1 μ M of WUS2-R9 at zero and twenty days post infiltration (dpi).

[0034] **Figure 10** demonstrates the efficiency of WUSCHEL derived CPPs. **Figure 10A** is a bar-dot plot showing quantification of GFP11 delivery to *N. benthamiana* leaves with 50 μ M of different CPP-GFP11 constructs, where the CPPs tested here are sequences from WUSCHEL. **Figure 10B** is a table showing the WUSCHEL based CPPs tested and their amino acid sequence.

[0035] **Figure 11** demonstrates the evolutionary diversity of the species from which the disclosed 3rd alpha helix sequences were derived and the conservation of the homeobox domain across those species. **Figure 11A** is a schematic of the taxonomic classification of the plant species from which the disclosed 3rd alpha helix sequences were derived. This figure illustrates the breadth of the species that were investigated ranging from flowering plants (Magnoliopsida) to algae (Chlorophyta). **Figure 11B** is a multiple sequence alignment of the third alpha helix of Arabidopsis WUSCHEL (top sequence) with other plant homeodomain proteins tested in Figure 12A. Residues at a given position with high similarity are marked in red and boxed in blue. Conservation is seen across different classes of homeodomain (HD) proteins and across different species. Alignment was done with Clustal Omega and visualized with ESPript 3.0.

[0036] **Figure 12** demonstrating the cell-penetrating capability of a diverse set of sequences derived from plant homeodomain proteins of various protein classes and plant species. **Figure 12A** is a bar-dot plot showing quantification of GFP11 delivery to *N. benthamiana* leaves with 50 μ M of different CPP-GFP11 constructs using CPPs derived from sequences of 30 different homeodomain proteins benchmarked against R9, reportedly the most efficient CPP for protein delivery in plants. Five CPPs, CIWOX, AtNDX, OtWOX, NtWOX, and AtPHD were more efficient than R9, 20 CPPs (right of the 6th, R9 bar) were less efficient than R9 but more efficient than the mock treatment (last bar), while five CPPs (extremely right of R9) were indistinguishable from the mock treatment. **Figure 12B** is a bar-

dot plot showing the concentration dependence of the delivery efficiency of the CIWOX CPP and GgBEL CPP. The CIWOX CPP outperforms the R9 CPP across multiple concentrations up to 200 μ M beyond which delivery efficiency saturates.

[0037] **Figure 13** provides sequence alignments of the top (**Figure 13A**) and lowest (**Figure 13B**) performing sequences from which a design pattern for high efficiency sequences derived from plant homeodomain proteins is deduced. **Figure 13C** is a partial formula according to one embodiment of the present disclosure for a highly efficient sequence deduced from the multiple sequence alignments. X indicates any amino acid may take this position, PN indicates a polar neutral amino acid takes this position, nP indicates a non-polar amino acid takes this position, PP indicates a polar positive amino acid takes this position.

DETAILED DESCRIPTION

[0038] As described herein, the present disclosure presents a method by which to discover new plant-derived cell penetrating peptides that enable delivery of biomolecules to plants. The present disclosure experimentally demonstrates that the 3rd alpha helix of the homeobox domain of WUSCHEL (WUS), a homeodomain plant transcription factor, exhibits plant cell penetrating capabilities. Three unique WUS fragments were tested, and the results showed they can enter plant cells through diffusion, without the use of biolistic force. As provided herein, based on sequence alignment of plant homeodomain proteins across various plant species, there are hundreds of thousands of plant transcription factors (or fragments thereof) across multiple plant species that could exhibit cell penetrating activity. As such, the present disclosure contemplates cell penetrating capabilities of peptides derived from plant transcription factors and contemplates a broad class of sequences for which cell penetrating capabilities are exhibited. Notably, while cell penetrating peptides themselves are not new, the cell penetrating peptides that are reported and used in the literature are derived from and optimized for use in mammalian cells, not plant cells (Xie et al. (2020) *Front Pharmacol.* 11; Copolovci et al. (2014) *ACS Nano* 8(3):1972-1994). As such, the present disclosure presents both a new method by which to discover plant-specific and plant-optimized cell penetrating peptides for biomolecule delivery in plants, and as an example also demonstrates with DCIP that three unique AtWUS fragments enable delivery in plant cells.

[0039] As used herein, the term "plant" refers to whole plants, plant organs (for example, leaves, stems, roots, meristem, spermatogonial, embryonic, pollen, egg, vasculature etc.), seeds, and plant cells and progeny of same. Plant cell, as used herein includes, without limitation, seed suspension cultures, embryos, meristematic regions, callus tissue, leaves,

roots, shoots, gametophytes, sporophytes, pollen, and microspores. Moreover, the CPP delivery disclosed herein is applicable to different plant subcellular locations including without limitation the nucleus, cytosol, mitochondria, chloroplast, etc.

[0040] Plants have a cell wall in addition to their cell membrane, such that most existing cell penetrating peptides work poorly in plants (if at all) as shown in Figure 5A-C. The present disclosure discloses *plant-derived* cell penetrating peptides that are optimized for use in plants and exhibit high (up to 80%) delivery efficiencies in plant tissues. Such plant-derived cell penetrating peptides can be tagged (e.g., fused) to any biomolecule of interest (DNA, RNA, protein) to enable delivery of the biological cargo into plant cells. Currently, DNA, RNA, and protein delivery in plants requires laborious delivery efforts (Agrobacterium or viral delivery which causes transgenesis, or low-efficiency biolistic delivery). By covalently tagging any cargo of interest with the cell penetrating peptides provided herein, those cargoes can now freely diffuse into plant cells to exhibit the biological activity desired by the end user.

[0041] “Biomolecules,” as used herein, may include one or more of a chemical compound, a protein or fragment thereof, a recombinant protein or fragment thereof, a glycoprotein or fragment thereof, a peptide, an antibody or fragment thereof, an enzyme or fragment thereof, a nuclease or fragment thereof, a hormone or fragment thereof, a cytokine or fragment thereof, a transcription factor or fragment thereof, a toxin or fragment thereof, a nucleic acid, a carbohydrate, a lipid, a glycolipid, a drug, a fluorophore, a nutrient, a fluorescent protein or fragment thereof, an antibiotic, a recombinase (e.g., Cre) and/or a plant hormone (e.g., an auxin or a cytokinin).

[0042] In some embodiments, a protein or a polypeptide or fragment thereof is delivered by a CPP. As will be appreciated by those of skill in the art, a CPP or a biomolecule such as a polypeptide may comprise changes (e.g., mutations) that do not negatively impact function. As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See Golub, E. S., and Green, D. R. 1991. *Immunology: A Synthesis*, 2d ed. Sinauer Associates, Sunderland, Massachusetts, which is incorporated herein by reference. Conventional notation is used herein to portray polypeptide sequences: the left-hand end of a polypeptide sequence is the amino-terminus; the right-hand end of a polypeptide sequence is the carboxyl-terminus.

[0043] A “conservative amino acid substitution” is one in which an amino acid residue is substituted by another amino acid residue having a side chain R group with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or

more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. See, e.g., Pearson WR. (1994) *Methods Mol. Biol.* 24:307-31.

[0044] Examples of groups of amino acids that have side chains with similar chemical properties include 1) aliphatic side chains: glycine, alanine, valine, leucine, and isoleucine; 2) aliphatic-hydroxyl side chains: serine and threonine; 3) amide-containing side chains: asparagine and glutamine; 4) aromatic side chains: phenylalanine, tyrosine, and tryptophan; 5) basic side chains: lysine, arginine, and histidine; 6) acidic side chains: aspartic acid and glutamic acid; and 7) sulfur-containing side chains: cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, glutamate-aspartate, and asparagine-glutamine.

[0045] Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet et al., *Science* 256:1443-45 (1992), incorporated herein by reference. A “moderately conservative” replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

[0046] Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, and (4) confer or modify other physicochemical or functional properties of such analogs. Analogs comprising substitutions, deletions, and/or insertions can include various mutants of a sequence other than the naturally-occurring peptide sequence. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts). A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (e.g., a replacement amino acid should not tend to break a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in *Proteins, Structures and Molecular Principles* (Creighton, Ed., W. H. Freeman and Company, New York (1984)); *Introduction to Protein Structure* (C. Branden and J. Tooze, eds., Garland Publishing, New York, N.Y. (1991)); and Thornton et al., *Nature* 354:105 (1991), which are each incorporated herein by reference.

[0047] Sequence similarity for polypeptides, and similarly sequence identity for polypeptides, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various

substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as “Gap” and “Bestfit” which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. See, e.g., GCG Version 6.1. Polypeptide sequences also can be compared using FASTA using default or recommended parameters, a program in GCG Version 6.1. FASTA (e.g., FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183:63-98 (1990); Pearson, *Methods Mol. Biol.* 132:185-219 (2000)). Another preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially *blastp* or *tblastn*, using default parameters. See, e.g., Altschul et al., *J. Mol. Biol.* 215:403-410 (1990); Altschul et al., *Nucleic Acids Res.* 25:3389-402 (1997); incorporated herein by reference.

[0048] An “analog,” such as a “variant” or a “derivative,” is a compound (e.g., a biomolecule) substantially similar in structure and having the same biological activity, albeit in certain instances to a differing degree, to a naturally-occurring molecule. For example, a polypeptide variant, including a CPP or a biomolecule described herein, refers to a polypeptide sharing substantially similar structure and having the same biological activity as a reference polypeptide. Variants or analogs differ in the composition of their amino acid sequences compared to the naturally-occurring polypeptide from which the analog is derived, based on one or more mutations involving (i) deletion of one or more amino acid residues at one or more termini of the polypeptide and/or one or more internal regions of the naturally-occurring polypeptide sequence (e.g., fragments), (ii) insertion or addition of one or more amino acids at one or more termini (typically an “addition” or “fusion”) of the polypeptide and/or one or more internal regions (typically an “insertion”) of the naturally-occurring polypeptide sequence or (iii) substitution of one or more amino acids for other amino acids in the naturally-occurring polypeptide sequence. By way of example, a “derivative” is a type of analog and refers to a polypeptide sharing the same or substantially similar structure as a reference polypeptide, including a CPP or a biomolecule described herein, that has been modified, e.g., chemically.

[0049] A variant polypeptide is a type of analog polypeptide, including a CPP or a biomolecule described herein, and includes insertion variants, wherein one or more amino acid residues are added to a biomolecule amino acid sequence of the invention. Insertions may be located at either or both termini of the protein, and/or may be positioned within internal regions of the therapeutic protein amino acid sequence. Insertion variants, with

additional residues at either or both termini, include for example, fusion proteins and proteins including amino acid tags or other amino acid labels. In one aspect, the biomolecule optionally contains an N-terminal Met, especially when the molecule is expressed recombinantly in a bacterial cell such as *E. coli*. In another aspect, the biomolecule includes histidine tag (His-tag).

[0050] In deletion variants, one or more amino acid residues in a polypeptide, including a CPP or a biomolecule described herein, are removed. Deletions can be effected at one or both termini of the therapeutic protein polypeptide, and/or with removal of one or more residues within the therapeutic protein amino acid sequence. Deletion variants, therefore, include fragments of a therapeutic protein polypeptide sequence.

[0051] In substitution variants, one or more amino acid residues of a polypeptide, including a CPP or a biomolecule described herein, are removed and replaced with alternative residues. In one aspect, the substitutions are conservative in nature and conservative substitutions of this type are well known in the art. Alternatively, the invention embraces substitutions that are also non-conservative.

[0052] As used herein, the terms “nuclease” and “endonuclease” are used interchangeably herein to mean an enzyme which possesses endonucleolytic catalytic activity for polynucleotide cleavage. The term includes site-specific endonucleases such as site-specific endonucleases of clustered, regularly interspaced, short palindromic repeat (CRISPR) systems such as, e.g., Cas polypeptides. In some embodiments, CRISPR associated protein 9 (CAS9), CAS12, CAS13, CAS14, CAS variants, recombinases (such as Cre), CxxC-finger protein-1 (Cfpl), zinc-finger nucleases (ZEN) and/or a transcription activator-like effector nuclease (TALEN) is delivered by a CPP.

[0053] In still other embodiments, a nucleic acid such as a DNA, RNA, antisense oligonucleotide (ASO), microRNA (miRNA), small interfering RNA (siRNA), aptamer, locked nucleic acid (LNA), peptide nucleic acid (PNA), and/or a morpholino is delivered to a plant cell by a CPP. The terms “nucleic acid molecule” and “polynucleotide” are used interchangeably herein, and refer to both RNA and DNA molecules, including nucleic acid molecules comprising cDNA, genomic DNA, synthetic DNA, and DNA or RNA molecules containing nucleic acid analogs. A nucleic acid molecule can be double-stranded or single-stranded (e.g., a sense strand or an antisense strand). A nucleic acid molecule may contain unconventional or modified nucleotides. The terms “polynucleotide sequence” and “nucleic acid sequence” as used herein interchangeably refer to the sequence of a polynucleotide molecule.

[0054] A DNA sequence that "encodes" a particular RNA is a DNA nucleic acid sequence that is transcribed into RNA. A DNA polynucleotide can encode an RNA (mRNA) that is translated into protein, or a DNA polynucleotide can encode an RNA that is not translated into protein (e.g. tRNA, rRNA, or a guide RNA; also called "non-coding" RNA or "ncRNA"). A "protein coding sequence" or a sequence that encodes a particular protein or polypeptide, is a nucleic acid sequence that is transcribed into mRNA (in the case of DNA) and is translated (in the case of mRNA) into a polypeptide in vitro or in vivo when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' terminus (N-terminus) and a translation stop nonsense codon at the 3' terminus (C-terminus). A coding sequence can include, but is not limited to, cDNA from prokaryotic or eukaryotic mRNA, genomic DNA sequences from prokaryotic or eukaryotic DNA, and synthetic nucleic acids. A transcription termination sequence will usually be located 3' to the coding sequence.

[0055] The term "recombinant" nucleic acid molecule as used herein, refers to a nucleic acid molecule that has been altered through human intervention. As non-limiting examples, a cDNA is a recombinant DNA molecule, as is any nucleic acid molecule that has been generated by in vitro polymerase reaction(s), or to which linkers have been attached, or that has been integrated into a vector, such as a cloning vector or expression vector (e.g., an AAV). As non-limiting examples, a recombinant nucleic acid molecule: 1) has been synthesized or modified in vitro, for example, using chemical or enzymatic techniques (for example, by use of chemical nucleic acid synthesis, or by use of enzymes for the replication, polymerization, exonucleolytic digestion, endonucleolytic digestion, ligation, reverse transcription, transcription, base modification (including, e.g., methylation), or recombination (including homologous and site-specific recombination)) of nucleic acid molecules; 2) includes conjoined nucleotide sequences that are not conjoined in nature, 3) has been engineered using molecular cloning techniques such that it lacks one or more nucleotides with respect to the naturally occurring nucleic acid molecule sequence, and/or 4) has been manipulated using molecular cloning techniques such that it has one or more sequence changes or rearrangements with respect to the naturally occurring nucleic acid sequence.

[0056] Homeodomain proteins are well-known in the art (Burglin et al. *Chromosoma*. 125:497-521 (2016); Mukherjee et al. *Mol. Biol. Evol.* 26(12):2775-2794. (2009)). A homeodomain protein contains a homeobox domain of 60-65 amino acids in length which is DNA-binding domain consisting of three consecutive alpha helices. In mammals, it is known that homeodomain proteins contain a cell penetrating peptide domain on the 3rd alpha helix of the homeobox domain which is 5-25 amino acids in length. However, in plants, the analogous cell penetrating domain was previously unknown until the present disclosure.

[0057] “Peptides” including “cell penetrating peptides” (e.g., CPP) as used herein refer to relatively short peptides, 4–40 aa, with the ability to gain access to the cell interior by means of different mechanisms, mainly including endo- cytosis, and/or with the capacity to promote the intracellular effects by these peptides themselves, or by the delivered covalently or noncovalently conjugated bioactive cargoes (biomolecules).

[0058] Homeodomain proteins (e.g., plant homeodomain proteins) may include, without limitation, a protein or fragment thereof as provided in UniProt under domain keyword “Homeobox” or “Homeodomain” (UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic acids research*, 51(D1): D523–D531). By way of example, the present disclosure provides that a peptide derived from any one or more of the following proteins may be used in the methods described herein: WUSCHEL, CLSY2, PINTOX, LUMI, NDX, PHD, RLT2, BEL1, KNAT1, ZHD3, ATB16, HDG7, HAT1, REV, WOX3B, HOX12, Q8LLD8, WOX, Q40238, Q69G85 and M1UW87.

[0059] Specific CPPs contemplated herein are provided in Table 1 below:

[0060] TABLE 1: CPPs contemplated in the present disclosure.

Name	Species	HD Class	Sequence
AtWUS_1	Arabidopsis	WOX	KNVFYWFQNHKARERQKKRFN (SEQ ID NO: 1)
AtWUS_2	Arabidopsis	WOX	KNVFYWFQNHKARERQ (SEQ ID NO: 2)
AtWUS_3	Arabidopsis	WOX	HKARERQ (SEQ ID NO: 3)
At_SAW	Arabidopsis	SAWADEE	KQIWNWFQNNRYALRAR (SEQ ID NO: 4)
At_PINTOX	Arabidopsis	PINTOX	KRVLKWFEDKRAEDGVP (SEQ ID NO: 5)
At_LD	Arabidopsis	LD	AQVREFFVTQKTRVRKQ (SEQ ID NO: 6)
At_NDX	Arabidopsis	NDX	SQLKNWLNNRKAKLARA (SEQ ID NO: 7)
At_PHD	Arabidopsis	PHD	EKVNKWFKNTRYMALRN (SEQ ID NO: 8)
At_DDT	Arabidopsis	DDT	RQLQMWFCHRRLLKERKS (SEQ ID NO: 9)
At_BEL	Arabidopsis	BEL	SQVSNWFINARVRLWKP (SEQ ID NO: 10)
At_KNOX	Arabidopsis	KNOX	KQINNWFINQRKRHWKP (SEQ ID NO: 11)
At_PLINC	Arabidopsis	PLINC	RVLKVWMHNNKQNLSSK (SEQ ID NO: 12)
At_ZIPI	Arabidopsis	HD-ZIP I	RQVAVWFQNNRRARWGTK (SEQ ID NO: 13)
At_ZIPIV	Arabidopsis	HD-ZIP IV	KQIKFWFQNNRTQMKTQ (SEQ ID NO: 14)

At_ZIP II	Arabidopsis	HD-ZIP II	RQVEVWFQNRRTARTKLK (SEQ ID NO: 15)
At_ZIP III	Arabidopsis	HD-ZIP III	KQIKVWFQNRRCRDQQR (SEQ ID NO: 16)
Zm_WOX	Zea mays	WOX	KNVFYWFQNHKARERQR (SEQ ID NO: 17)
Os_HOX12	Oryza sativa	HD-ZIP II	KQVAVWFQNRRTARHKS (SEQ ID NO: 18)
St_PHD	Solanum tuberosum	PHD	YQVSKWFENARHCHRS (SEQ ID NO: 19)
Nt_WOX	Nicotiana tabacum	WOX	KNVFYWFQNHKARQRQK (SEQ ID NO: 20)
Sp_DDT	Solanum peruvianum	DDT	RQLQMWFCHRRLLDKNT (SEQ ID NO: 21)
Lj_NDX	Lotus japonicus	NDX	SQLKNWLNNRKARLART (SEQ ID NO: 22)
Cm_BEL	Cyanidioschyzon merolae	BEL	KQVSNWFINFRKRSWHG (SEQ ID NO: 23)
Pp_ZIP III	Physcomitrella patens	HD-ZIP III	KQIKVWFQNRRCREKQR (SEQ ID NO: 24)
Cr_KNOX	Chlamydomonas reinhardtii	KNOX	KQVTDWFTNWRARHWRP (SEQ ID NO: 25)
Ot_WOX	Ostreococcus tauri	WOX	SNVYNWFQNRRSREKKL (SEQ ID NO: 26)
Sm_ZIP III	Selaginella moellendorffii	HD-ZIP III	RQIKVWFQNRRCREKQR (SEQ ID NO: 27)
Pa_KNOX	Picea abies	KNOX	KQINNWFINQRKRNWHS (SEQ ID NO: 28)
Cl_WOX	Cuninghamia lanceolata	WOX	TNVYNWFQNRRTARTKRK (SEQ ID NO: 29)
Gg_BeL	Gnetum gnemon	BEL	SQVSNWFINARVGLWKP (SEQ ID NO: 30)
Pt_WOX	Populus trichocarpa	WOX	TNVYNWFQNRRTARSKRK (SEQ ID NO: 31)
Cp_ZIP III	Ceratopteris pteridoides	HD-ZIP III	KQLKVWFQNRRCREKQR (SEQ ID NO: 32)

[0061] As described herein, CPPs may be fused to one or more biomolecules according to numerous embodiments of the present disclosure. “Fusions” may be engineered recombinantly as is known in the art. Fusion proteins are often referred to as having been “tagged” (e.g., with a fluorescent tag or other biomolecule of interest). Delivery of nucleic

acids, for example, may in some embodiments involve non-covalent (electrostatic, pi stacking, Van der Waals) interactions with potentially any part or amino acid of the CPP. In some embodiments, linkers may also be used. The role of the linker, such as a GSGS linker described herein, is to provide some spacing and flexibility (with, for example, G residues) and hydrophilicity (with, for example, S residues) with this linker. Thus, a "protein linker" such as (GS)_n, where n is an integer, is provided herein.

[0062] The CPPs described herein can be used, in some embodiments, to quantitate protein delivery. Numerous assays can be used in conjunction with the CPPs provided herein including, for example, a delivered complementation in planta (DCIP) assay (Wang et al. (2022b). Quantification of cell penetrating peptide mediated delivery of proteins in plant leaves *bioRxiv*. doi: 10.1101/2022.05.03.490515; Wang et al. 2023. Fluorescence complementation enables quantitative imaging of cell penetrating peptide-mediated protein delivery in plants including WUSCHEL transcription factor. *bioRxiv*. doi: 10.1101/2022.05.03.490515). Briefly, the DCIP assay includes the following steps in some embodiments: (a) fusing the 11th beta strand of green fluorescent protein to a delivery tool (such as a CPP); (b) expressing a reporter system consisting of mCherry fluorescent protein fused to a modified green fluorescent protein missing the 11th beta strand; and (c) introducing the fusion product of (a) into the plant tissue of (b) and performing microscopy to quantify the number of plant cells expressing both red fluorescence and green fluorescence.

[0063] In this disclosure, "comprises", "comprising", "containing", "having", and the like have the meaning ascribed to them in U.S. patent law and mean "includes", "including", and the like; the terms "consisting essentially of" or "consists essentially" likewise have the meaning ascribed in U.S. patent law and these terms are open-ended, allowing for the presence of more than that which is recited so long as basic or novel characteristics of that which is recited are not changed by the presence of more than that which is recited, but excluding prior art embodiments.

[0064] As used herein, the term "equal" generally means the same value +/- 10%. In some embodiments, a measurement, such as number of cells, etc., can be +/- 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10%. Similarly, as used herein and as related to amino acid position or nucleotide position, the term "approximately" refers to within 1, 2, 3, 4, or 5 such residues. With respect to the number of cells, the term "approximately" refers to +/- 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10%.

[0065] Before the present invention is further described, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing

particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[0066] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[0067] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present invention, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[0068] It must be noted that as used herein and in the appended claims, the singular forms "a," "and," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a conformation switching probe" includes a plurality of such conformation switching probes and reference to "the microfluidic device" includes reference to one or more microfluidic devices and equivalents thereof known to those skilled in the art, and so forth. It is further noted that the claims may be drafted to exclude any element, e.g., any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely," "only" and the like in connection with the recitation of claim elements, or use of a "negative" limitation.

[0069] As will be apparent to those of skill in the art upon reading this disclosure, each of the individual embodiments described and illustrated herein has discrete components and features which may be readily separated from or combined with the features of any of the other several embodiments without departing from the scope or spirit of the present invention. Any recited method can be carried out in the order of events recited or in any other order which is logically possible. This is intended to provide support for all such combinations.

EXAMPLES

[0070] General methods

[0071] Reagents and Antibodies

[0072] Reagents, buffers, and media components were procured through Sigma-Aldrich unless otherwise noted. Solid-phase chemical peptide synthesis of GFP11 and CPP fusions was performed by a third-party manufacturer (GenScript and LifeTein). Enzymes used for cloning reactions were procured through New England Biolabs. Anti-GFP11 antibody was purchased through Thermo-Fisher and the anti-mCherry and anti-rabbit IgG-HRP secondary antibody through Cell Signaling Technologies. All oligonucleotides and DNA sequences were purchased from Integrated DNA Technologies (IDT).

Plant growth conditions and agroinfiltration

[0073] *N. benthamiana* were grown in a growth chamber kept at 24 °C and a light intensity of 100-150 $\mu\text{mol m}^{-2} \text{s}^{-1}$. The photoperiod was kept at 16H light/8H dark. Seeds were sown in inundated soil (Sunshine Mix #4) and left to germinate for 7-10 days at 24°C before being transferred to 10cm pots for growth. Fertilization was done on a weekly basis with 75 ppm N 20-20-20 general-purpose fertilizer and 90 ppm N calcium nitrate fertilizer reconstituted in water. Infiltrations were performed on 4–5-week-old plants on the third and fourth expanded leaves. Agroinfiltrations were performed via needleless syringe using overnight cultures of *A. tumefaciens* bearing the DCIP constructs. The day of infiltration, the overnight 30°C cultures were pelleted at 3200xg, rinsed with infiltration buffer (10mM MES pH 5.7, 10mM MgCl₂) and then resuspended in infiltration buffer containing 200 μM acetosyringone to an OD₆₀₀ of 0.5-1.0. The cultures were then left shaking at ambient conditions for 2-4 hours before final adjustment of the OD₆₀₀ to 0.5 with infiltration buffer. During infiltration, care was taken to minimize the number of damaging infiltration spots to completely saturate the leaf. The plants were then left at ambient conditions overnight to dry before being transferred to the growth chamber for the total incubation time of 3 days.

Plasmid Construction and Bacterial Strains

[0074] A list of parent plasmids and newly constructed plasmids is provided in Table 2 below:

[0075] TABLE 2: Plasmids utilized in the described studies.

Plasmid Name	Notes	Source	Protein Product Sequence	Protein Length (AA)	Protein MW (kDa)

DCIP	35s::NLS:mCherry:sfGFP1-10::tNOS fusion for nuclear localized expression using <i>Agrobacterium</i> in pDGB1-a1R vector backbone	this work	MPKKKRKVGVSKEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYKGGGSGGSLQMIDSKGEE LFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTCLKICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKRHDFKSA MPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNI LGHKLEYNFNHNVYITADKQKNGI KANFTVRHNVEDGSVQLADHYQQ NTPIGDGPVLLPDNHYLSTQTVLSKDPNEK* (SEQ ID NO: 33)	471	52.7
cytoCDC IP	35s::mCherry:sfGFP1-10::tNOS fusion for cytosolic expression using <i>Agrobacterium</i> in pDGB1-a1R vector backbone	this work	MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYKGGGSGGSLQMIDSKGEE LFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTCLKICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKRHDFKSA MPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNI LGHKLEYNFNHNVYITADKQKNGI KANFTVRHNVEDGSVQLADHYQQ NTPIGDGPVLLPDNHYLSTQTVLSKDPNEK* (SEQ ID NO: 34)	463	51.8
1B-GFP1-10	sfGFP1-10 inserted into 1B for <i>E. coli</i> expression and Hisx6 tag	this work	MGSSHHHHHHENLYFQSNATMIDSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTCLKICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKRHDFKSA MPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNI LGHKLEYNFNHNVYITADKQKNGI KANFTVRHNVEDGSVQLADHYQQ NTPIGDGPVLLPD	236	26.6

			NHYLSTQTVLSKDPNEK* (SEQ ID NO: 35)		
1BR9-mCherry	mCherry insertion of 1BR9 for <i>E. coli</i> expression and Hisx6 tag	this work	MGSSHHHHHHENLYFQSNARDH MVLHEYVNAAGITGGGSGGGGS YFQSNVSKGEEDNMAIIEFMRF KVHMEGSVNGHEFEIEGEGEGRPY EGTQTAKLKVTGGPLPFAWDILSP QFMYGSKAYVKHPADIPDYLKLSFP EGFKWERVMNFEDGGVVTVTQDS SLQDGEFIYKVKLRGTNFPDGPV MQKKTMGWEASSERMYPEDGAL KGEIKQRLKLDGGHYDAEVKTTYK AKKPVQLPGAYNVNIKLDITSHNED YTIVEQYERAEGRHSTGGMDELYK GIGSGSNGSSGSVSRRRRRRRR* (SEQ ID NO: 36)	309	34.5
1BR9-mCherry STOP	mCherry insertion of 1BR9 <i>E. coli</i> expression and Hisx6 tag with stop codon to prevent R9 tagging	this work	MGSSHHHHHHENLYFQSNARDH MVLHEYVNAAGITGGGSGGGGS YFQSNVSKGEEDNMAIIEFMRF KVHMEGSVNGHEFEIEGEGEGRPY EGTQTAKLKVTGGPLPFAWDILSP QFMYGSKAYVKHPADIPDYLKLSFP EGFKWERVMNFEDGGVVTVTQDS SLQDGEFIYKVKLRGTNFPDGPV MQKKTMGWEASSERMYPEDGAL KGEIKQRLKLDGGHYDAEVKTTYK AKKPVQLPGAYNVNIKLDITSHNED YTIVEQYERAEGRHSTGGMDELYK (SEQ ID NO: 37)	286	32.0
1BR9-Lifeact	Lifeact peptide insertion of 1BR9 <i>E. coli</i> expression and Hisx6 tag	this work	MGSSHHHHHHENLYFQSNARDH MVLHEYVNAAGITGGGSGGGGS YFQSNAMGVADLIKKFESISKEEGI GSGSNGNSGSVSRRRRRRRR* (SEQ ID NO: 38)	91	9.89
PEP101E-NLS-mCherry	UBQ10 driven expression of NLS-mCherry-sfGFP1-10 and Hisx6 tag	this work	MPKKRKRKVGSKGEEDNMAIIEF MRFKVHMEGSVNGHEFEIEGE GRPYEGTQTAKLKVTGGPLPFAW DILSPQFMYGSKAYVKHPADIPDYL KLSFPEGFKWERVMNFEDGGVVT VTQDSSLQDGEFIYKVKLRGTNFP DGPVMQKKTMGWEASSERMYPE DGALKGEIKQRLKLDGGHYDAEV KTTYKAKKPVQLPGAYNVNIKLDIT SHNEDYTIVEQYERAEGRHSTGGM DELYKGGGSGGGSLQMIDSKGEE LFTGVVPILVELDGDVNGHKFSVRG EGEGDATIGKLTLCFICTTGKLPVP WPTLVTTLTGYGVQCFSRYPDHMKR	471	52.7

			HDDFFKSAMPEGYVQERTISFKDDG KYKTRAVVKFEGDTLVNRIELKGT FKEDGNILGHKLEYNFNSHNYYITA DKQKNGIKANFTVRHNVEDGGSVQ LADHYQQNTPIGDGPVLLPDNHYL STQTVLSKDPNEK* (SEQ ID NO: 39)		
1BR9	LIC backbone derivatized from 1B for expression of N-terminal GFP11 and C- terminal R9 tagged proteins in <i>E. coli</i>	this work	MGSSHHHHHHENLYFQSNARDH MVLHEYVNAAGITGGGGSGGGGS YFQSNIGSGSNGNSGSVSRRRRRR RRR* (SEQ ID NO: 40)	72	7.86
PEP101E	Cytosolic UBQ10 driven expression of sfGFP1-10 Plasmid from Prof. Dinesh Kumar (Park et al Plant Cell. 2017 Jul;29(7):1571- 1584.)	Addgene #97387			
1B	LIC backbone for hisx6 tagged expression of proteins in <i>E. coli</i> from Scott Gradia at UC Berkeley MacroLab	Addgene #29653			
H6- mCherry	mCherry coding sequence from Scott Gradia at UC Berkeley MacroLab	Addgene #29722			
1BR9- AtWUS	<i>E. coli</i> codon optimized AtWUS insertion of 1BR9 for <i>E. coli</i> expression and Hisx6 tag	this work	MGSSHHHHHHENLYFQSNARDH MVLHEYVNAAGITGGGGSGGGGS YFQSNAEPPQHQQHHHQADQES GNNNNNKSGSGGYTCRQTSTRWT PTTEQIKILKELYNNAIRSPTADQI QKITARLRQFGKIEGKNVFWFQN HKARERQKKRFNGTNMTTPSSSPN SVMMAANDHYHPLLHHHHGVPM QRPANSVNVKLNQDHHLYHHNKP YPSFNNGNLNHASSGTECGVNAS NGYMSSHVYGSMEQDCSMNYNN VGGGWANMDHHYSSAPYNFFDR	365	41

			AKPLFGLEGHQEEEECGGDAYLEH RRTLPLFPMHGEDHINGGSGAIWK YGQSEVRPCASLELRLNGIGSGSNG SSGSVSRRRRRRRRR (SEQ ID NO: 41)		
1BR9- AtWUSS TOP	<i>E. coli</i> codon optimized AtWUS insertion of 1BR9 for <i>E. coli</i> expression and Hisx6 tag with stop codon to prevent R9 tagging	this work	MGSSHHHHHHENLYFQSNARDH MVLHEYVNAAGITGGGGSGGGGS YFQSNAEPPQHQH HH HQADQES GNNNNNKSGSGGYTCRQTSTRWT PTTEQIKILKELYNNAIRSPTADQI QKITARLRQFGKIEGKNVFWFQN HKARERQKKRFNGTNMTTPSSSPN SVMMAANDHYHPLLHHHHGVPM QRPANSVNVKLNQDHHLYHHNKP YPSFNNGNLNHASSGTECGVNAS NGYMSSHVYGSMEQDCSMNYNN VGGGWANMDHHYSSAPYNFFDR AKPLFGLEGHQEEEECGGDAYLEH RRTLPLFPMHGEDHINGGSGAIWK YGQSEVRPCASLELRLN (SEQ ID NO: 42)	342	38.5
1BR9- BFP	BFP insertion of 1BR9 for <i>E. coli</i> expression and Hisx6 tag	this work	MGSSHHHHHHENLYFQSNARDH MVLHEYVNAAGITGGGGSGGGGS YFQSNAASSSELIKENMHMKLYME GTVDNHHFKCTSEGEKPYEGTQT MRIKVVEGGPLPFAFDILATSFYGS KTFINHTQGIPDFFKQSFPEGFTWE RVTTYEDGGVLTATQDTSLQDGCLI YNVKIRGVNFTSNGPVMQKKT LG WEAFTETLYPADGGLEGRNDMAL KLVGGSHLIANIKTTYRSKKPAKNLK MPGVYYVDYRLRIKEANNETYVE QHEVAVARYCDLPSKLGHKLNGIG SGSNGSSGSVSRRRRRRRRR (SEQ ID NO: 43)	305	34.3
1BR39- BFPSTOP	BFP insertion of 1BR9 for <i>E. coli</i> expression and Hisx6 tag with stop codon to prevent R9 tagging	this work	MGSSHHHHHHENLYFQSNARDH MVLHEYVNAAGITGGGGSGGGGS YFQSNAASSSELIKENMHMKLYME GTVDNHHFKCTSEGEKPYEGTQT MRIKVVEGGPLPFAFDILATSFYGS KTFINHTQGIPDFFKQSFPEGFTWE RVTTYEDGGVLTATQDTSLQDGCLI YNVKIRGVNFTSNGPVMQKKT LG WEAFTETLYPADGGLEGRNDMAL KLVGGSHLIANIKTTYRSKKPAKNLK MPGVYYVDYRLRIKEANNETYVE	282	31.8

			QHEVAVARYCDLPSKLGHKLN (SEQ ID NO: 44)		
1BR9- AtWUSS TOP- $\Delta\alpha 3$	GFP11 tagged AtWUS with the third homeodomain helix deleted (AA82-101)	this work	MGSSHHHHHHENLYFQSNARDH MVLHEYVNAAGITGGGSGGGGS YFQSNAEPPQHQQHHHQADQES GNNNNNKSGSGGYTCRQTSTRWT PTTEQIKILKELYNNAIRSPTADQI QKITARLRQFGKIEGKNVFWFQN HKARERQKKRFNGTNMTTPSSSPN SVMMAANDHYHPLLHHHHGVPM QRPANSVNVKLNQDHHLYHHNKP YPSFNNGNLNHASSGTECGVVNAS NGYMSSHVYGSMEQDCSMNYNN VGGGWANMDHHYSSAPYNFFDR AKPLFGLEGHQEEEECGGDAYLEH RRTLPLFPMHGEDHINGGSGAIWK YGQSEVRPCASLELRN (SEQ ID NO: 45)	321	35.6

[0076] Additionally, the predicted protein products and organization of all constructed plasmids are provided. Primers and synthetic DNA used for cloning are included in Table 3. For all cloning steps, plasmids were transformed into XL1-blue *E. coli*. The coding sequence of DCIP was constructed by ligating mCherry sequence into a PstI 5' of the sfGFP1-10 coding sequence in pPEP101 (Park, E et al. (2017) *Plant Cell*. 29(7):1571-1584.). In the nuclear localized variant of DCIP, NLS was attached during PCR of the mCherry sequence. The NLS is omitted in cytoDCIP. Next, the coding sequences of DCIP and cytoDCIP were amplified by PCR for domestication into pUDP2 before final Golden Braid (GB2.0) assembly following the standard GB2.0 protocol using Esp3I (Sarrion-Perdigones, A et al. (2013) *Plant Physiol*. 162(3):1618-1631). The DCIP and cytoDCIP transcriptional units were assembled using pUDP2-35S-oTMV for the promoter and pUDP2-tNOS for the terminator in a BsaI restriction-ligation GB2.0 reaction. cytoDCIP and DCIP were then transformed into GV3101 *Agrobacterium tumefaciens* bearing pSOUP (Hellens, RP et al. (2000) *Plant Mol Biol*. 42(6):819-832) and plated onto LB agar containing rifampicin (50µg/mL), gentamicin (25µg/mL), and kanamycin (50 µg/mL).

[0077] The recombinant GFP1-10 expression vector, 1B-GFP1-10, was constructed by PCR amplifying the sfGFP1-10 gene from a pPEP101 with the ligation-independent cloning tags for plasmid 1B. The full protocol for LIC cloning used was provided by the UC Berkeley Macro Lab: <http://colon-forward-slash-forward-slash-qb3.berkeley.edu/facility/qb3->

macrolab/projects/lic-cloning-protocol-forward slash. The resulting amplicon was then inserted by LIC into 1B and transformed into *E. coli* for expansion, purification, sequencing, and transformation into an expression *E. coli* strain. The 1BR9 plasmid was constructed by inserting a short, chemically synthesized DNA sequence containing N-terminal GFP11 and C-terminal R9 tag into plasmid 1B via LIC. Between the tags, a new LIC site was regenerated such that future LIC reactions would insert the protein of interest between the N- and C- tags. The LIC approach allowed insertion of PCR amplified mCherry and BFP sequence into 1BR9 to generate 1BR9-mCherry and 1BR9-BFP. Incorporation of a TAA stop codon into the reverse primer generated 1BR9-mCherrySTOP and 1BR9-BFPSTOP which excludes the c-terminal R9 motif. 1BR9-Lifeact was produced by inserting a chemically synthesized DNA sequence for Lifeact into 1BR9. 1BR9-AtWUS and 1BR9-AtWUSSTOP was constructed by LIC insertion of an *E. coli* codon optimized (IDT) synthesized DNA AtWUS (TAIR: At2g17950.1) DNA sequence with or without a TAA stop codon into the 1BR9 vector. 1BR9-AtWUSSTOP- $\Delta\alpha 3$, for the expression of GFP11-AtWUS- $\Delta\alpha 3$, was created using around-the-horn cloning with 5' phosphorylated primers to exclude the third alpha helix of the homeodomain via PCR using 1BR9-WUSSTOP as the template.

Recombinant Protein Expression and Purification

[0078] For all recombinant protein expression, Rosetta 2 (DE3) pLysS *E. coli* were transformed with recombinant protein expression vectors and plated on to selective chloramphenicol (25 $\mu\text{g/mL}$) and kanamycin (50 $\mu\text{g/mL}$) LB agar plates for overnight growth at 37°C, 250rpm. Single colonies were used to inoculate 10mL seed cultures in LB. After overnight starter culture growth at 37°C and 250rpm, 1L LB with selective antibiotics were set to grow at 37°C, 250rpm in 2L baffled flasks. Induction was performed with 0.5 mM IPTG at 37°C when the culture reached 0.8 OD600. After 4 hours of induction, cultures were pelleted for 20 minutes at 3200xg, and flash frozen in liquid nitrogen. Cell lysis was conducted using thawed pellets in lysis buffer (50mM Tris-HCL, 10mM imidazole, 500mM NaCl, pH 8.0) with 1x protease inhibitor cocktail (Sigma-Aldrich: S8830) using probe tip sonication. The resulting lysate was clarified by centrifugation at 40,000xg for 30 minutes. For GFP11-10 and mCherry fusion proteins, the soluble fraction was incubated with 1mL Ni-NTA (Thermo Scientific: 88221) slurry for an hour. After incubation and washing, the proteins were eluted using elution buffer (500mM imidazole, 150mM Tris-HCL, pH 8.0). The resulting eluate was then concentrated and buffer exchanged into 10mM Tris-HCL, 100mM NaCl pH 7.4 via ultrafiltration in a 3500Da cutoff filter (Emdmillipore: C7715). The GFP11-mCherry and GFP11-mCherry-R9 were then further polished using SEC (Cytiva: HiLoad 16/600 Superdex 200pg) and exchanged into storage buffer (10mM Tris, 10mM NaCl pH 7.4) before ultrafiltration concentration and flash freezing for storage. BFP purification proceeded

similarly except for the R9 construct where all steps are done at pH 10.8, 20mM CAPS buffer instead of Tris. For GFP11-Lifeact-R9, the insoluble pellet from clarification was solubilized in 8M Urea, 50mM Tris, pH 8.0 before incubation with Ni-NTA for an hour. In addition to standard wash steps, a high pH wash (20mM CAPS pH 10.8, 1M NaCl) was required to remove residual nucleic acids from the protein. Protein was then eluted with elution buffer before spin concentration and exchange into storage buffer and flash freezing. Aliquots of each recombinant protein were run on SDS-PAGE for confirmation (Figure 7).

Recombinant GFP11-AtWUS Purification

[0079] GFP11-AtWUS-R9, GFP11-AtWUS, and GFP11-AtWUS- $\Delta\alpha 3$ was expressed as mentioned previously. However, purification proceeded by sonication lysis in 6M Urea, 50mM Tris, 0.5mM TCEP, and 2mM MgCl₂ pH 7.5 in the presence of 25 U/mL of benzonase and protease inhibitor cocktail. After lysis, the lysate was then incubated at 37C for 30 minutes with occasional mixing. After incubation, an additional 25 U/mL of benzonase was added and the lysate was clarified by centrifugation at room temperature, 40,000xg for 30 minutes. To the clarified supernatant, 2mL of Ni-NTA slurry per liter of starting culture was added and incubated at room temperature for 2 hours. The Ni-NTA was then washed sequentially with at least 15 bed volumes each of wash A (6M Urea, 50mM Tris pH 7.5), then wash B (6M Urea, 50mM Tris, pH 7.5, 500mM NaCl), then wash C (6M Urea, 50mM Tris, pH 7.5, 50mM Imidazole). The protein was eluted twice with the addition of 1 bed volume of 1M Imidazole, 20mM MES, pH 6.9, 200mM NaCl. The fractions were then combined and buffer exchanged into buffer P (30mM Sodium Acetate, 5mM NaCl, pH 4.0, 0.5mM TCEP) by dialysis or by desalting in a G25 sephadex pre-packed PD-10 column (Cytiva) and then analyzed with SDS-PAGE (Figures 7A-D) and *in vitro* complementation.

In vitro GFP complementation assay

[0080] In a 96-well qPCR plate (Biorad), 5 μ L of 10 μ M total protein of sfGFP1-10 Ni-NTA eluate in storage buffer (10mM Tris, 10mM NaCl pH7.4) was combined with 5 μ L buffer or 20 μ M of each tested peptide or protein in storage buffer. Each well was mixed by pipetting and was tested in triplicate. GFP complementation of AtWUS proteins was accomplished similarly except 10 μ L of 15 μ M either GFP11-AtWUS-R9 or R9-GFP11 in buffer P was added to 10 μ L of 10 μ M GFP1-10. An additional 5 μ L 1M Tris pH 8.0 was also added to overcome the acidity of buffer P. GFP complementation was quantified using a Biorad CFX96 qPCR machine by measuring the green fluorescence at 1-minute intervals over the course of 6 hours at either 22°C or 4°C.

SDS-PAGE and Western Blot

[0081] For Western Blot of DCIP and cytoDCIP, agroinfiltrated leaves were harvest 3 d.p.i, flash frozen, ground, and lysed with RIPA buffer (Abcam: ab156034) for 20 minutes. Lysates were then clarified by centrifugation and boiled in 1X Laemmli buffer (Biorad) and 10 (V/V) % beta-mercaptoethanol for 5 minutes before being loaded into a 4-20% gradient SDS-PAGE gel (Biorad: 4561096) and run according the manufacturer's instructions. Membrane transfer was done according to manufacturer's instructions onto Immobilon PVDF membrane (Millipore). Blocking was performed using 5% milk in PBS with 0.1% Tween (PBST). The incubation with primary anti-mCherry antibody (CST: E5D8F) was performed overnight at 1:1000 dilution in 3% BSA in PBST at 4°C with orbital shaking at 60 rpm. Imaging was performed after probing with anti-Rabbit IGG-HRP secondary antibody (CST: 7074) at 1:10,000 dilution in 5% milk PBST and ECL prime chemiluminescent reagent (Amersham: RPN2236) on a ChemiDoc gel imager (Biorad).

R9-GFP11 Dot Blot Assay

[0082] The third or fourth leaf of 4–5-week-old wild-type *N. benthamiana* were infiltrated with 500 μ M R9-GFP11 in water using a needleless syringe. Infiltrations were staggered such that all treatments could be harvested simultaneously for 0, 4, 8, and 24H time points. Each treatment was performed on a separate plant and the experiment was repeated thrice. A 12mm leaf disc was excised for each treatment using a leaf punch and flash frozen in liquid nitrogen before grinding and lysis in 20 μ L RIPA buffer with 1x plant protease inhibitor cocktail (Sigma-Aldrich: P9599). The lysates were then clarified by centrifugation at 21000xg for 30 minutes. Immediately after, 2 μ L lysates were spotted onto a nitrocellulose membrane (Amersham: GE10600002) and allowed to dry. The membranes were then blocked in 5% milk in PBS-T before being washed and probed with an anti-GFP11 antibody (Invitrogen: PA5-109258) (1:500 dilution in PBST with 3% BSA). Secondary antibody probing and imaging was performed similarly to Western Blot.

Delivered complementation *in planta* infiltration

[0083] Three days after agroinfiltration, DCIP expressing leaves were infiltrated with the treatment solutions. Unless otherwise stated, an 8mm punch was then excised from the infiltrated area and plated, abaxial side up, onto ½ MS pH 5.7 agar plates. The plates were then left to incubate under ambient conditions for 4-5 H before imaging. For cold temperature treatment, after infiltration with ice-cold solutions of R9-GFP11 and disc excision, the leaf discs were plated onto ice-cold agar and immediately transferred to a 4°C refrigerator for incubation. After incubation, the agar plates with leaf discs were kept on ice until the moment of imaging. For *in situ* incubation, leaves were simply infiltrated and the plant was returned to the growing chamber for incubation. All peptides were dissolved in

sterile water and all recombinant proteins were dissolved in 10mM Tris pH 7.4, 10mM NaCl for infiltration. GFP11-AtWUS-R9 was found to possess low solubility at neutral pH and was exchanged into 10mM MES, pH 5.5, immediately before infiltration.

Confocal Imaging and Image Analysis

[0084] Excised leaf discs were imaged on a Zeiss LSM880 laser scanning confocal microscope. Images for semi-quantitative analysis were collected using a 20x/1.0NA Plan-apochromat water immersion objective and larger field-of-views were collected using a 5x objective. Leaf discs were mounted by sandwiching a droplet of water between the leaf disc and a #1.5 cover glass. BFP, sfGFP, mCherry, and chloroplast autofluorescence images were acquired by excitation with a 405, 488, 561, and 635 nm laser respectively. The emission bands collected for BFP, sfGFP, mCherry, and autofluorescence were 410-529, 493-550 nm, 578-645 nm and 652-728 nm respectively. All images were collected such that the aperture was set to 1 Airy-unit in the mCherry channel. Images and profile plots were prepared for publication using Zen Blue software. Profile plots were smoothed by taking a moving window of three measurements and normalized to the maximum smoothed intensity for each color. For quantification experiments, z-stacks were acquired with the imaging depth set to capture the epidermal layer down to the point where mCherry nuclei could no longer be detected. Z-stacks from four field of views were acquired for every treatment condition. Quantitative image analysis and downstream processing was performed using Cell Profiler 3.0. On a slice-by-slice basis, mCherry fluorescent nuclei were segmented using Otsu's method (Otsu, N. (1979) "A Threshold Selection Method from Gray-Level Histograms," in IEEE Transactions on Systems, Man, and Cybernetics, 9(1): 62-66) and chloroplasts were segmented in the autofluorescence channel. The segmented chloroplasts were applied as a sfGFP channel mask over the image to exclude plastid autofluorescence in downstream image analysis. After masking, maximum intensity projections of identified nuclei were generated in the sfGFP, autofluorescence, and mCherry channels. Cell Profiler was then used to quantify the number, mCherry intensity, GFP intensity, and red (mCherry)/green (sfGFP) ratio of the projected nuclei.

Arabidopsis AtWUS Seedling Treatment

[0085] *Arabidopsis* (Col-0) seedlings were grown in 12-well culture plates with 8 to 10 seedlings per well in 1 mL of 1x MS media supplemented with 0.5% sucrose and 2.5 mM MES, pH 5.7. Seeds were sterilized by washing in 70% ethanol for 30 seconds followed by a 15-minute incubation in 50% bleach supplemented with 0.5% Tween-20 and rinsed 5x with DI water. Sterilized seeds were stratified in plates at 4°C for 3 days after plating. Seedlings were grown at 22°C under 16-hour photoperiods for 12 days. To treat seedlings, the liquid

media in each well was replaced with control and WUSCHEL treatments. Control wells were refreshed with 1 mL of MS growth media. WUSCHEL proteins were prepared for use by dialysis into 10mM MES pH 5.7 for two hours before dilution to their final concentration. Treatment wells were refreshed with 1 mL of 1 or 3 μ M of protein (R9-GFP11, GFP11-AtWUS-R9, GFP11-AtWUS) dissolved in MS growth media. After 24 hours of treatment, seedlings were frozen in liquid nitrogen and physically disrupted with chrome steel bearing balls.

RT-qPCR analysis

[0086] Total RNA was extracted from Arabidopsis seedlings using a RNeasy Plant Mini Kit (Qiagen). Extracted RNA quality was confirmed with a NanoDrop UV-Vis Spectrometer. Complementary DNA was synthesized from RNA using an iScript cDNA Synthesis Kit (Bio-Rad). qPCR was run with PowerUP SYBR Green Master Mix (Applied Biosystems), each reaction was run in triplicate according to manufacturer's recommendations. Melt-curve analysis was run after qPCR cycling to confirm primer specificity. Relative gene expression was determined using the ddCt method (Livak KJ, Schmittgen TD. (2001) *Methods*. 25(4):402-408.) using SAND1 as the reference gene. Relative gene expression was determined from 4 biological pools each containing 8 to 10 seedlings. A list of utilized primers and sequence accession numbers is available in Table 3 below. Statistical comparisons of ddCT values were conducted as done previously (Yuan, JS et al. (2006) BMC Bioinformatics. 7:85) using a T-test with Holm-Šídák correction in GraphPad Prism 9.

[0087] TABLE 3: Primer sequences and sequence accession numbers

DNA Set	FW 5->3' Sequence	REV 5->3' Sequence	Use
1	AAACTGCAGATGCCAAAG AAAAAGCGGAAAGTCGGA GTGAGCAAGGGCGAGGAG (SEQ ID NO: 46)	AAACTGCAGGCTGCCACCGCC GCTACCGCCACCGCCCTTGTA CAGCTCGTCCATGC (SEQ ID NO: 47)	Amplification of mCherry from h6 mCherry Vector with PSTI overhangs for 5' insertion into Dinesh-kumar sfGFP1-10 vector with an NLS sequence and disordered linker
2	TACTTCCAATCCAATGCAG TGAGCAAGGGCGAGGAG (SEQ ID NO: 48)	CTCCCACTACCAATGCCTTAC TTGTACAGCTCGTCCATG (SEQ ID NO: 49)	Amplification of mCherry from h6 mCherry Vector for 1BR9 cloning (with stop codon)
3	TACTTCCAATCCAATGCAG TGAGCAAGGGCGAGGAG (SEQ ID NO: 50)	CTCCCACTACCAATGCCCTTG TACAGCTCGTCCATG (SEQ ID NO: 51)	Amplification of mCherry from h6 mCherry Vector for 1BR9 cloning (no stop codon)
4	TACTTCCAATCCAATGCCA CCATGATCGATAGCAAAG GAGAAG (SEQ ID NO: 52)	TTATCCACTTCCAATGTTATTA TCACTTTTCGTTGGGATCTTTC (SEQ ID NO: 53)	Amplification of sfGFP1-10 from plasmid from Dinesh-kumar for cloning into 1B

5	GCGCCGTCTCGCTCAAAGCTCACTTTT (SEQ ID NO: 54)	GCGCCGTCTCGCTCGAATGCCAAAGA (SEQ ID NO: 55)	Amplification of NLS:mCherry:GFP1-10 for goldenbraid cloning into pUPD2
6	GCGCCGTCTCGCTCGAATGGTGAGCA (SEQ ID NO: 56)	GCGCCGTCTCGCTCGAATGCCAAAGA (SEQ ID NO: 57)	Amplification of mCherry:GFP1-10 for goldenbraid cloning into pUPD2
7	TACTTCCAATCCAATGCACGCGATCACATGGTCCTGCACGAGTACGTGAACGCCGCCGGGATCACTGGTGGCGGAGGTTCTGGAGGCGGTGGATCGTACTTCCAATCCAAATTATTGGTAGTGGGAGCAACGGCAGCAGCGGATCCGTGAGCCGCCGTGCGCGTCGCCGTGCGTCCGTCGCTAATAACATTGGAAGTGGATAA (SEQ ID NO: 58)	TTATCCACTTCCAATGTTATTAGCGACGGCGACGGCGACGGCGACGGCGGCTCACGGATCCGCTGCTGCCGTTGCTCCCACTACCAATATTGGATTGGAAGTACGATCCACCGCCTCCAGAACCTCCGCCACCAGTGATCCCGGCGGCGTTACGTAAGTCTGTCAGGACCATGTGATCGCGTGCATTGGATTGGAAGTA (SEQ ID NO: 59)	LIC Insert for conversion of 1B to 1BR9
8	TACTTCCAATCCAATGCAATGGGCGTGGCCGACCTGATCAAGAAGTTCGAGAGCATCAGCAAGGAAGAGGGCATTGGTAGTGGGAG (SEQ ID NO: 60)	CTCCCACTACCAATGCCCTCTTCCTTGCTGATGCTCTCGAATTCTTGATCAGGTCGGCCACGCCATTGCATTGGATTGGAAGTA (SEQ ID NO: 61)	Lifeact Insert for LIC into 1BR9
9	ATGGAACCTCCCCAGCACGAGCACCATCACCACCAGGCTGATCAAGAGTCGGGAATAACAACAACAATAAAAGCGGGAGCGGGGGTTACACCTGTGCGTCAAACATCCACGCGCTGGACACCAACCACGTAACAATCAAAATTTGAAAGAATTATACTATAACAATGCCATTGTTACCTACCGCAGATCAAATCCAAAAATTACAGCACGCTTGCGCCAGTTCGGAAAGATTGAGGGAAAAACGTGTTCTACTGGTTTCAGAATCATAAGGCACGTGAACGTCAAAAAAACGTTTTAACGGGACTAATATGACAACACCGAGTTCGATGCCCAACTCCGTTATGATGGCAGCGAACGATCACTATCATCCATTACTGCACCACATCATGGAGTCCCAATGCAGCGTCCAGCCAATTCCGTGAATGTCAAGCTGAACAAGACCACCATCTTTATCACACAATAAGCCTTACCCATCTTTTAATAACGGCAACCTGAATCACGCCAGCAGT	CTAATTCAAACGCAGCTCAAGACTTGCGCAAGGGCGTACCTCCGACTGTCCGTAATCCAAATGGCGCCAGACCCACCATTGATATGGTCCTCTCCATGCATCGGGAAAAGCGGTAAAGTACGACGATGTTCTAAGTAAGCGTCACCCCCGCATTCTCCTCCTCCTGATGTCCCTCTAAGCCGAATAACGGTTTAGCACGGTCGAAGAAATTGTACGGCGCAGAGCTGTGTGATGGTCCATATTGGCCCAAGCCCCCTCCTACGTTGTTGTATTTCATTGAACAGTCTTGTTCCATAGAACCGTAGACGTGCGACGACATATATCCATTGACGCAATTGACGACCCCGCATTCCGGTCCACTGCTGGCGTGATTGAGTTGCCGTTATTAAAAGATGGGTAAGGCTTATTGTGGTGATAAATGATGGTGGTCTTGTTTCAGCTTGACATTCACGGAATTGGCTGGACGCTGCATTGGGACTCCATGATGGTGGTGCAGTAATGGATGATAGTGATCGTTCGCTGCCATCATAACGGAGTTGGGACTCGAACTCGGTGTTGTCATATTAGTCCCGTTAAAACGTTTTTTTTGA	<i>E. coli</i> codon optimized DNA sequence for AtWUS

	GGCACCGAATGCGGGGTC GTCAATGCGTCGAATGGA TATATGTCGTCGCACGTCT ACGGTTCTATGGAACAAGA CTGTTCAATGAATTACAAC AACGTAGGAGGGGGCTGG GCCAATATGGACCATCACT ACAGCTCTGCGCCGTACA ATTTCTTCGACCGTGCTAA ACCGTTATTCGGCTTAGAG GGACATCAGGAGGAGGAG GAATGCGGGGGTGACGCT TACTTAGAACATCGTCGTA CTTTACCGCTTTTCCCGAT GCATGGAGAGGACCATAT CAATGGTGGGTCTGGCGC CATTTGGAAGTACGGACA GTCGGAGGTACGCCCTTG CGCAAGTCTTGAGCTGCG TTTGAATTAG (SEQ ID NO: 62)	CGTTCACGTGCCTTATGATTCT GAAACCAGTAGAACACGTTTT TTCCCTCAATCTTTCCGAACTG GCGCAAGCGTGCTGTAATTTT TTGGATTTGATCTGCGGTAGG TGAACGAATGGCATTGTTATA GTATAATTCTTTCAAATTTTG ATTTGTTCAGTGGTTGGTGTC CAGCGCGTGGATGTTTGACGA CAGGTGTAACCCCGCTCCCG CTTTATTGTTGTTGTTATTTT CCGACTCTTGATCAGCCTGGT GGTGATGGTGCTGGTGCTGG GGAGGTTCCAT (SEQ ID NO: 63)	
10	TACTTCCAATCCAATGCAG AACCTCCCCAGCACCAG (SEQ ID NO: 64)	CTCCCACTACCAATGCCATTC AAACGCAGCTCAAGAC (SEQ ID NO: 65)	Amplification of AtWUS sequence from synthetic DNA for insertion into 1BR9
11	AACTCTATGCAGCATTTGA TCCAC (SEQ ID NO: 66)	TGATTGCATATCTTTATCGCCA TC (SEQ ID NO: 67)	AtSAND1 qPCR Primer (TAIR: AT2G28390.1)
12	CGGTATGACTGGATATGA C (SEQ ID NO: 68)	ATTTGCATCGGAGAGCTC (SEQ ID NO: 69)	AtARR6 qPCR primer (TAIR: AT5G62920.1)
13	CGAAGGGTTTtaggactac ATGAAG (SEQ ID NO: 70)	GTGGGTTTcacatgatggTGCA A (SEQ ID NO: 71)	AtCLV3 qPCR Primer (TAIR: AT2G27250.1)
14	TACTTCCAATCCAATGCAG AACCTCCCCAGCACCAG (SEQ ID NO: 72)	CTCCCACTACCAATGCCTTAAT TCAAACGCAGCTCAA (SEQ ID NO: 73)	Amplification of AtWUS sequence from synthetic DNA for insertion into 1BR9 (With Stop Codon)
15	TACTTCCAATCCAATGCAG CTTCCTCCAGCGAGCTG (SEQ ID NO: 74)	CTCCCACTACCAATGCCATTA AGCTTGTGCCCCAGTT (SEQ ID NO: 75)	Amplification of BFP for 1BR9 Cloning (no stop codon)
16	TACTTCCAATCCAATGCAG CTTCCTCCAGCGAGCTG (SEQ ID NO: 76)	CTCCCACTACCAATGCCTCAA TTAAGCTTGTGCCCCA (SEQ ID NO: 77)	Amplification of BFP for 1BR9 Cloning (with stop codon)
17	ATCGACGGATCAAAGGGC A (SEQ ID NO: 78)	ATGGCCATGTCTCTCGTGAAG (SEQ ID NO: 79)	AtKANADI1 qPCR primer (TAIR: AT5G16560.1)
18	AGCAGCTTCGTCAGGACA AT (SEQ ID NO: 80)	AACTTCCCATGCAAACGTGC (SEQ ID NO: 81)	AtKANADI2 qPCR primer (TAIR: AT1G32240.1)
19	CACAAAGAACACCATTGAT CCCA (SEQ ID NO: 82)	GGGTTGCATACAAAAGCGTG T (SEQ ID NO: 83)	AtASL2 qPCR primer (TAIR: AT1G65620.1)
20	CACTGTCCTTGCTGTGAGT (SEQ ID NO: 84)	AGCTCTCATGCTCACGGTC (SEQ ID NO: 85)	AtYAB3 qPCR primer (TAIR: AT4G00180.1)
21	GGGACTAATATGACAAC ACCG (SEQ ID NO: 86)	TCCCTCAATCTTTCCGAACT (SEQ ID NO: 87)	Mutagenesis of 1BR9- AtWUSSTOP to delete alpha helix 3 of the WUS homeodomain

Sequence Alignment

[0088] The sequences for *Arabidopsis thaliana* WUSCHEL (UniProt: Q9SB92), Shoot Meristemless (UniProt: Q38874), *Zea mays* WUS2 (UniProt: A0AAS6), *Homo sapiens* VAX1 (UniProt: Q5SQQ9), OCT4 (UniProt: D5K9R8), and *Drosophila melanogaster* ANTENNAPEDIA (UniProt: P02833) were aligned using UniProt Clustal Omega (<https://www.uniprot.org/align>). The aligned sequences were then prepared using the ESPript 3.0 web server (Robert X, Gouet P. (2014) *Nucleic Acids Res.* 42(Web Server issue): W320-W324.).

Data Processing and Statistics

[0089] Data obtained from Cell Profiler was processed using a script written in Python 3.9. Using the Cell Profiler data, GFP positive nuclei were counted using a python script by setting a threshold defined as a one-tailed 99% confidence interval above the mean green/red ratio in the untreated control or water infiltration control of each experiment. Any nuclei with green/red ratio higher than this threshold would be identified as GFP positive. The percentage of sfGFP positive nuclei was calculated by dividing sfGFP positive nuclei by the total number of mCherry nuclei counted. In experiments where delivery efficiency is used, delivery efficiency is defined by normalizing the percentage of sfGFP positive nuclei to the 100µM R9-GFP11 treatment in that experiment. All summary statistics were calculated using Python before export and statistical analysis in GraphPad Prism 9. Kruskal-Wallis non-parametric ANOVA (Kruskal WH, Wallis WA. (1952). *Journal of the American Statistical Association (JASA)*. 47(260), 583–621.) was used for analysis of multiple comparisons followed by uncorrected Dunn's non-parametric T-test unless otherwise noted. Single comparisons were made using a one-sample t-test against the normalized value of 1.0. All presented plots were also generated in GraphPad Prism 9.

EXAMPLE 1: Development of a Delivered Complementation *in planta* (DCIP) sensor system

[0090] To detect protein delivery using confocal microscopy, an *Agrobacterium tumefaciens* expression mediated, GFP-complementation based, red/green ratiometric sensor was developed ("Delivered Complementation *in planta*," (DCIP)). In this technique, a robustly folded version of green fluorescent protein, called 'superfolder' GFP (sfGFP) is separated between the 10th and 11th strands of the beta barrel, splitting it asymmetrically into a large non-fluorescent fragment (sfGFP1-10) and a small peptide strand (GFP11) (Hu CD, Chinenov Y, Kerppola TK. (2002) *Mol Cell*. 2002;9(4):789-798; Kamiyama, D et al. (2016) *Nat Commun*. 7:11046). These two fragments are not individually fluorescent; only upon localization to same compartment do sfGFP1-10 and GFP11 bind one another and

reconstitute the GFP fluorescence of an intact fluorophore (Figure 2A). This method has the critical benefit of only producing signal if the peptide tag remains intact, is successfully delivered to the cytosol, and is not sequestered in lytic organelles or trapped in the apoplast. GFP11 also serves as an excellent reporter tag because its short length (16AA) is accessible to chemical synthesis and because it is readily incorporated into recombinant proteins as a terminal tag. The final design of DCIP was also guided by the desire to perform automated image analysis within complex leaf tissues and thus reduce the risk of bias during analysis.

[0091] While bimolecular fluorescence complementation and nanoluciferase complementation have been previously applied for measuring cell penetrating peptide (CPP) mediated delivery in mammalian cells (Milech, N et al. (2015) *Sci Rep.* 5:18329; Schmidt, S et al. (2015) *Angew Chem Int Ed Engl.* 54(50):15105-15108; Teo, SLY et al. (2021) *Nat Commun.* 12(1):3721), it has not been employed to confirm the delivery of bio-cargoes to plant cells. CPPs, are small cationic or amphipathic peptides that when conjugated to cargoes, enable cytosolic delivery (Langel Ü (ed). (2011) *Cell-penetrating peptides: methods and protocols.* New York, NY: Humana Press). CPPs were utilized to test DCIP due to their synthetic accessibility, previous deployment in plant-tailored delivery schemes (Numata, K et al. (2018) *Sci Rep.* 8(1):10966; Thagun, C et al. (2022) *ACS Nano.* 16(3):3506-3521), and because much of their underlying cell penetrating mechanisms in plants remain unstudied. In general, the sfGFP 1-10 is expressed in the cytoplasm of the cells and localized to the nucleus for ease of imaging. The GFP 11-test CPP peptide construct is introduced to the plant cells with reconstitution occurring only if a CPP successfully internalizes cargo. Reconstituted GFP is detected by live confocal microscopy.

[0092] The present disclosure provides a method, in some embodiments, by which to discover new and plant-specific cell penetrating peptides identified by sequence alignment of plant homeodomain proteins. Protein delivery is validated, in some embodiments, with the novel AtWUS-derived CPPs as described herein and using a DCIP delivery sensor protein comprising of sfGFP1-10 that was C-terminally fused to mCherry and an N-terminal SV40 NLS (Hicks, GR et al. (1995) *Plant Physiol.* 107(4):1055-1058). For DCIP, the mCherry fusion was chosen for three reasons: (1) mCherry is easy to spectrally resolve from plant autofluorescence (2) a constitutive fusion allows identification of positively *A. tumefaciens* transfected cells and (3) mCherry fusion permits ratiometric quantification of GFP bimolecular fluorescence complementation since the relative expression of sfGFP1-10 is tied to the expression of mCherry by direct fusion. Because plant cells are heterogeneous in shape and have many autofluorescent bodies, the sensor was localized to the nucleus to produce a round, uniform object that is amenable to automated image analysis and provides

unambiguous confirmation of successful delivery of GFP11 or GFP11-tagged cargoes. The NLS localization of DCIP should allow the detection of a broad range of sizes of delivered cargoes as the size exclusion limit for efficient transport through the nuclear envelope is presumed to be greater than 60 kDa (Wang R, Brattain MG. (2007) *FEBS Lett.* 581(17):3164-3170).

[0093] The DCIP coding sequence was constructed by traditional restriction ligation cloning and the final transcriptional unit assembly performed using Goldenbrai 2.0 (Sarrion-Perdigones, A et al. (2013) *Plant Physiol.* 162(3):1618-1631). The cytosolically localized version of DCIP, cytoDCIP, which lacks SV40 NLS was developed in tandem. These constructs (Figure 1) were transformed into *A. tumefaciens* and agroinfiltrated in *Nicotiana benthamiana* plants. *N. benthamiana* was chosen as a model plant due to its common use in transient expression experiments as well as in delivery experiments (Martin, K et al. (2009) *Plant J.* 59(1):150-162). Successful expression of intact DCIP three days post agroinfiltration (d.p.i.) was verified by microscopic observation (Figure 2B) and by Western blot using an anti-mCherry antibody (Figure 2D), and prior to attempts at GFP11 or GFP11-tagged cargo delivery.

[0094] For a proof of concept testing the ability of DCIP to detect successful GFP11 delivery, the nona-arginine (R9) cell penetrating peptide that was fused to GFP11 by a (GS)2 linker was used. R9 was chosen due to its known effectiveness in both plants (Numata, K et al. (2018) *Sci Rep.* 8(1):10966) and mammalian (Kosuge, M et al. (2008) *Bioconjug Chem.* 19(3):656-664) systems as well as its relatively well characterized mechanism of action in mammalian cells (Wallbrecher, R et al. (2017) *J Control Release.* 256:68-78). A typical experimental workflow using DCIP is provided in Figure 2A. The DCIP protocol involves transient expression of DCIP in *N. benthamiana*. Three days post infiltration (3 d.p.i.), leaves are infiltrated with an aqueous solution of cargo that contains the GFP11 tag. Immediately after infiltration, the infiltrated leaves are either left intact or a leaf disc is excised from the infiltrated area and plated on pH 5.7 ½ MS. After a predetermined incubation time, the leaves are imaged on a confocal laser scanning microscope. For quantitative imaging, the resulting images are then automatically analyzed using Cell Profiler (Stirling, DR et al. (2021) *BMC Bioinformatics.* 22(1):433) for nuclear sfGFP and mCherry fluorescence (Figure 2A). A detailed methodology is provided in the Methods section.

RESULTS

[0095] Without infiltration or infiltration with water, no sfGFP fluorescence was observed (Figure 2B, left panel) and only mCherry containing nuclei can be seen (Figure 2B, right panel) for the leaves agroinfiltrated with DCIP reporter, respectively. However, upon

infiltration with 100 μ M R9-GFP11, robust sfGFP complementation was observed 4-5 hours post-infiltration that colocalized with mCherry (Figure 2C). These data show that protein cargoes require a CPP to enter plant cells. The timing of 4-5 hours was determined by balancing the reported mammalian uptake kinetics for R9 peptides (<1H) (Kosuge, M et al. (2008) *Bioconjug Chem.* 19(3):656-664; Brock R. (2014) *Bioconjug Chem.* 25(5):863-868) against the relatively slow process of GFP complementation experimentally determined herein, where it was found that it required >5H for total complementation using an *in vitro* system with recombinant sfGFP1-10. For comparison, previous dye labeled cargo CPP-mediated delivery experiments in plants often used time points of about 2H (Numata, K et al. (2018) *Sci Rep.* 8(1):10966). In this case, an 8mm leaf disc was excised from peptide infiltrated tissue and plated onto $\frac{1}{2}$ MS to control possible apoplastic flow and uncontrolled drying of the infiltrated liquid which may change the effective concentration of R9-GFP11 the cells experience. An orthogonal projection (Figure 2B) showed GFP complementation deep ($\sim$ 100 μ m) into the z-axis of the leaf in both pavement cells and mesophyll cells. Imaging at a lower magnification showed efficient delivery throughout the leaf disc using DCIP (Figure 2E).

EXAMPLE 2: Validating the use of DCIP in quantifying protein delivery efficiency using R9-GFP11 peptide

[0096] To assess whether DCIP can be used to quantify the relative effectiveness of peptide delivery in planta, R9-GFP11 was again used for validation experiments, wherein infiltration into DCIP expressing *N. benthamiana* was carried out at concentrations ranging from 0-100 μ M R9-GFP11. The initial concentration range was determined from previously reported effective concentrations for mammalian cells (Pantarotto, D et al. (2004) *Chem Commun (Camb)*. (1):16-17; Kosuge, M et al. (2008) *Bioconjug Chem.* 19(3):656-664). After 4-5 hours incubation, a clear concentration dependent upshift of the green/red ratio was observed in DCIP expressing nuclei (Figure 3A). Seven biological repeats (one plant per repeat) were acquired using this methodology. The resulting mean green/red ratio from each repeat was averaged and showed a statistically significant increase of green/red at concentrations greater than or equal to 50 μ M R9-GFP11 (Figure 3B). Further gating for percent GFP positive nuclei to account for variations in quantitating fluorescence data in microscopy also showed significant ($p < 0.05$) delivery at 20 μ M (Figure 3C left panel). Slightly improved linearity of response was also observed with respect to R9-GFP11 concentration using percent positive instead of averaged green/red ratio. Delivery efficiency was then defined by normalization of the percentage of GFP positive cells with respect to the 100 μ M treatment (Figure 3C right panel). Normalization helped reduce the biological variation in the observed absolute uptake efficiency. The lack of lower-end sensitivity below 20 μ M was

attributed to erroneous detection of autofluorescent bodies in the 0 μM control used for thresholding. From this initial performance study, it was observed that relatively higher concentrations of R9-cargo are required for efficient delivery in plants relative to mammalian cells, which can undergo delivery at low micromolar concentrations (Schmidt, S et al. (2015) *Angew Chem Int Ed Engl.* 54(50):15105-15108). However, relatively high delivery efficiency (>40% positive) could be also achieved in plants when leaves were treated with a concentration of 100 μM R9-GFP11 and higher (Figure 3C).

[0097] After the low concentration-range validation, the point of DCIP signal saturation was investigated, which represents the maximal possible protein delivery efficiency in plant leaves. With the aforementioned workflow, DCIP expressing leaves were infiltrated with 0-1000 μM R9-GFP11 in water and incubated as leaf discs for 4-5 hours. Once again, strong upshifting of the green/red ratio of DCIP nuclei was observed in treated leaves as a function of R9-GFP11 concentration (Figure 3D). After six experimental repeats (one plant per repeat), it was also observed that concentrations greater than 300 μM R9-GFP11 showed a statistically significant increase in delivery compared to 100 μM R9-GFP11 and a relative saturation in green/red ratio at concentrations at 300 μM and higher (Figure 3E). Similarly, the data analyzed using percent GFP positive cells showed a saturation at about 75% positive at 500 μM and above, while the difference was significant at 300 μM when the delivery efficiency was normalized to the 100 μM control to account for inter-plant variation in delivery (Figure 3F). Examples of maximum intensity projections of the GFP channel for a titration DCIP experiment are provided in Figure 3G. Importantly, although quantitation failed to detect statistical significance across samples treated with less than 20 μM R9-GFP11, successful delivery could still be occasionally observed for these low peptide concentrations as sfGFP complemented nuclei (Figure 3G). Multi-channel images showing mCherry expression showed strong colocalization of the sfGFP signal and mCherry nuclear signals at all tested concentrations (Figure 4).

EXAMPLE 3: Use of DCIP as a screening platform for identifying effective plant cell-penetrating peptides (CPPs) and as a tool to determine CPP mechanism of cargo delivery

[0098] After validating DCIP for quantitative peptide delivery, an assessment was conducted to determine whether DCIP could be used to screen for effective cell penetrating peptide sequences in leaves. Three commonly used cell penetrating peptide sequences were assessed, based on CPPs that have been optimized for use and shown to work for biomolecule delivery in mammalian cells (Figure 5A). BP100 is a microbially derived CPP that has been previously reported in numerous studies to deliver cargoes in mammalian cells and has also shown to be effective in plants through a dye conjugation and delivery

experiment with unknown quantitative efficiencies (Numata, K et al. (2018) *Sci Rep.* 8(1):10966). TAT is an arginine-rich HIV-1 derived peptide and one of the first cell penetrating peptides characterized (Ziegler, A et al. (2005) *Biochemistry.* 44(1):138-148). R9 is a derivative of TAT where all amino acids are substituted for arginine (Kosuge, M et al. (2008) *Bioconjug Chem.* 19(3):656-664.). Each of these CPPs were produced through solid-phase synthesis as fusions to GFP11, separated by a short (GS)₂ linker. An *in vitro* bimolecular fluorescence complementation assay was used to ensure that the CPP fusions did not interfere with complementation activity (Figure 9). After infiltrating 100 μ M of each peptide construct into DCIP expressing leaves and incubating for 4-5 hours, confocal image analysis revealed that both TAT and R9 were effective at delivering GFP11 into plant cells and enabled delivery efficiencies ranging from 30-80% (Figure 5B). R9 appeared to be the most effective of the tested CPPs with TAT being 0.87 times as effective as R9, and BP100 or GFP11 alone showing no statistically significant signal. These results indicate that without a CPP, GFP11 is not able to enter the cytosol of plant cells. Surprisingly, BP100 mediated delivery was not statistically significantly higher than either the water infiltration control or GFP11 alone (Figure 5B), however, rare instances of successful delivery for 100 μ M BP100-GFP11 but not for the negative control or GFP11 alone were observed. Closer inspection of the imaged nuclei also revealed strong nucleolar localization of sfGFP in TAT-GFP11 and R9-GFP11 treatments (Figure 5C). These images suggest that R9 and TAT remain intact when bound to sfGFP1-10 in the cell, as poly-arginine motifs are known to localize to the nucleolus (Martin, RM et al. (2015) *Nucleus.* 6(4):314-325). These results also confirm that DCIP can be used to show that CPPs that were previously developed and optimized for use in mammalian cells are largely ineffective for protein delivery in plants. Furthermore, data is provided to show that CPP-based plant cell internalization is likely endocytosis-independent, since the application of endocytosis inhibitors Wortmannin and Ikarugamycin prior to attempted R9-GFP11 delivery did not alter the peptide delivery efficiency (Figures 5D-F). These results are important because they suggest that CPP-based delivery in plants is not subject to endocytic entrapment, which compromises cargo effectiveness, and may explain the high (~80%) protein delivery efficiencies.

[0099] After validation of R9-GFP11 as the best performing CPP for protein delivery in plants, the mechanism by which R9 delivers cargoes to the plant cell was probed. Specifically, experiments were set up to determine whether R9 delivery was endocytosis dependent or independent. The delivery efficiency (based on percentage of positive cells) in leaf discs infiltrated with 100 μ M R9GFP11 and incubated at 4°C or room temperature was not statistically different. However, the normalized delivery intensity, as defined by green/red ratio normalized to the room temperature treatment showed that the 4°C discs possessed

lower sfGFP fluorescence. This aligns with *in vitro* data showing that bimolecular fluorescence complementation is possible at 4°C, although compromised in efficiency (Figure 6). These data suggest that R9 delivery is largely independent of cellular activity such as endocytosis. Co-infiltration of R9-GFP11 and endocytosis inhibitors wortmannin or ikarugamycin (Bandmann V, Homann U. (2012) *Plant J.* 2012;70(4):578-584; Bandmann, V et al. (2012) *FEBS Lett.* 586(20):3626-3632; Elkin, SR et al. (2016) *Traffic.* 17(10):1139-1149) similarly resulted in no statistically significant decrease in delivery efficiency (Figure 5F). Taken together, these data align with previously reported studies in mammalian cells and a singular plant protoplast centered study (Chugh A, Eudes F. (2007) *Biochim Biophys Acta.* 1768(3):419-426) suggest that at concentrations greater than 10 µM, R9 enter cells through a combination of direct membrane permeation and through endocytosis with subsequent endosomal escape (Kosuge, M et al. (2008) *Bioconjug Chem.* 19(3):656-664; Wallbrecher, R et al. (2017) *J Control Release.* 256:68-78).

EXAMPLE 4: CPP mediated delivery of recombinant WUSCHEL

[0100] After establishing the feasibility of using DCIP to assess recombinant protein delivery in intact leaves, the possibility of R9-mediated delivery of the *Arabidopsis* plant morphogenic transcription factor, WUSCHEL (AtWUS), was next examined in *N. benthamiana* leaves. AtWUS was chosen as a candidate cargo due to its applications for somatic embryogenesis in plants and its high degree of molecular characterization (Zuo, J et al. (2002) *Plant J.* 30(3):349-359; Ikeda M, Mitsuda N, Ohme-Takagi M. (2009) *Plant Cell.* 21(11):3493-3505). DCIP expressing leaves infiltrated with 140 µM GFP11-AtWUS-R9 (MW = 41 kDa) showed robust nuclear GFP complementation at 6H that did not colocalize with plastid autofluorescence. Because AtWUS lacks intrinsic fluorescence unlike mCherry, quantitative DCIP analysis could be demonstrated using GFP11-AtWUS-R9 with 50% of DCIP expressing cells being GFP positive, suggesting a 50% AtWUS delivery efficiency with R9 as the carrier, on a per-cell basis (Figure 8A, B). Because AtWUS is a transcription factor, GFP11-AtWUS-R9 had been expected to localize to the nucleus even without the SV40 NLS of DCIP. To test this, a cytoDCIP expressing *N. benthamiana* leaf was infiltrated with 140µM GFP11-AtWUS-R9. If the delivered AtWUS had active NLS activity, green fluorescence was expected to localize only to the nucleus with the excess, uncomplemented mCherry-GFP1-10 remaining in the cytosol. Indeed, at 6H numerous GFP positive nuclei surrounded by cytosolic mCherry fluorescence were observed, thus confirming native NLS targeting of delivered AtWUS. In contrast, R9-GFP11 treated cytoDCIP showed general, cytosolic localization. These data showed that R9 fusion is effective for WUS delivery and that the purified R9-tagged transcription factor is able to enter the nucleus.

[0101] To test whether the delivered AtWUS was transcriptionally active, 12-day-old *Arabidopsis* seedlings were treated for 24H with 1 μ M GFP11-AtWUS-R9. *A. thaliana* was chosen as a model species due to the well characterized AtWUS pathway in *A. thaliana*. Seedlings were subsequently harvested and subjected to RT-qPCR analysis for six known direct targets of AtWUS. GFP11-AtWUS-R9 delivery led to down-regulation of ARR6 (0.42-fold), ASL2 (0.36-fold), KAN1 (0.58-fold), KAN2 (0.44-fold), YABBY3 (0.28-fold), and upregulation of CLV3 (2.06-fold), when compared to the buffer-treated control. Statistical analysis of the measured CT values showed that R9-GFP11 alone did not mediate significant transcriptional changes while GFP11-AtWUS-R9 recapitulated the expected transcriptional response to AtWUS overexpression. Ectopic expression of AtWUS has been shown to downregulate the expression of cell identity markers ARR6, ASL2, KAN1/2 and YAB3 by binding directly to their promoters (Leibfried, A et al. (2005) *Nature*. 438(7071):1172-1175; Yadav, RK et al. (2013) *Mol Syst Biol.* 9:654), and upregulate the expression of its own negative regulator, CLV3 (Schoof, H et al. (2000) *Cell*. 100(6):635-644). A repeat experiment with a separately purified batch of proteins led to similar results with all expected down-regulated targets being downregulated, although CLV3 was not upregulated to a statistically significant degree and ARR6 was only down-regulated relative to the R9-GFP11. In this second experiment, 3 μ M of protein was also used due to apparent batch-to-batch variation of activity and lower purity of the recombinant protein (Figures 7A and B). These data suggest that not only is AtWUS delivery possible, but that this delivered transcription factor can be transcriptionally active in plants. Next, to confirm the utility of delivered WUS transcription factor, rice callus tissue was treated with WUS2-R9 at 0.5 μ M and 1 μ M (Figure 9I). Compared to mock treated callus, tissue treated with 1 μ M of WUS2-R9 displayed signs of somatic embryogenesis with significant greening and rooting. This greening and rooting exhibited in WUS2-R9 treated callus suggests WUS delivered with CPPs is capable of improving plant regeneration by either 1) reducing the time to shooting or 2) increasing the number of shooting events. Furthermore, this data suggests that CPPs can deliver transcription factors to plant somatic embryos. In the present case, we show we can deliver WUS into walled plant cells, show WUS remains transcriptionally active, and show WUS induces the expected phenotype in plant tissues. These results support the use of CPPs in plants to deliver transcription factors and other functional proteins including genome editing nucleases for plant biotechnology.

EXAMPLE 5: Identification of cell penetrating peptides

[0102] The prior example demonstrated that AtWUS could be delivered through its covalent tagging to R9, which was previously shown to be the only mammalian cell-

optimized CPP that could deliver proteins to plants (Figure 5A-C). The R9 tag was then removed and the ability of GFP11-AtWUS to penetrate plant cells on its own was tested using DCIP. In contrast to all other proteins tested for delivery without the R9 tag AtWUS was found to enter plant cells without an R9 fusion (Figure 8A) with similar efficiency as the R9 containing construct at 140 μ M (Figure 8B). The AtWUS protein also possessed nuclear localization activity when infiltrated into a cytoDCIP expressing leaf (Figure 9A). These results suggest that AtWUS itself or derivatives thereof inherently contain cell penetrating abilities in plants. In search for an explanation for the native cell penetrating behavior, the sequence of the third homeodomain helix of AtWUS was aligned with several animal homeodomain proteins (ANTENNAPEDIA, VAX1, OCT4) (Figure 8C) which have been also shown to be cell penetrating in mammalian cells (Perez, F et al. (1992) *J Cell Sci.* 102 (Pt 4):717-722; Balayssac, S et al. (2006) *Biochemistry.* 45(5):1408-1420; Harreither, E et al. (2014) *Cell Regen.* 3(1):2) and two plant homeodomain proteins (WUS2 and STM). Moreover, a global analysis of homeodomain proteins in human cells has shown the majority to be cell penetrating and potential paracrine signaling molecules (Lee, EJ et al. (2019) *Cell Rep.* 28(3):712-722.e3). Additionally, one of the first CPPs, characterized, penetratin (RQIKIWFQNRRMKWKK (SEQ ID NO: 88)), is derived from the 3rd alpha helix of the homeobox domain of *Drosophila* ANTENNAPEDIA homeodomain protein. This homeobox domain is conserved in Maize KNOTTED-LIKE 1, the 3rd alpha helix of which is reported as cell penetrating in mammalian cells (Perez, F et al. (1992) *J Cell Sci.* 102 (Pt 4):717-722; Balayssac, S et al. (2006) *Biochemistry.* 45(5):1408-1420). To further confirm this, a peptide fragment was synthesized consisting of the conserved amino acids 82-102 (KNVIFYWFQNHKARERQKKRFN (SEQ ID NO: 1)) from the 3rd alpha helix of the homeobox domain of AtWUS and GFP11 (WUSP-GFP11) (Figure 8C). Using confocal analysis, it was found that this peptide was indeed cell penetrating (Figure 9B), although not as effective as R9 CPP (Figure 8D). Deletion of amino acids 82-102 from GFP11-AtWUS ($\Delta\alpha 3$) resulted in a protein with compromised cell penetration ability ($\Delta\alpha 3$ = 15.7% vs WT = 51% mean GFP positive) and treatment of DCIP expressing plants with this protein resulted in non-significant ($p>0.05$) delivery regardless of an outlier test-based exclusion of a repeat where high level uptake was measured (Figures 9C & D). The low, basal-level delivery of $\Delta\alpha 3$ might have resulted from the low amounts of non-specific uptake at the investigated concentration. Although these results may seemingly contradict previous research showing plasmodesmata-based trafficking of WUS (Daum, G et al. (2014) *Proc Natl Acad Sci U S A.* 111(40):14619-14624), the direct, exogenous delivery method bypasses the plasmodesmata and relies on diffusion through the cell wall and penetration through the plasma membrane. GFP11-AtWUS protein also showed lower transcriptional activity than the R9 CPP containing construct, regulating only half (3/6) or one sixth (1/6) of the tested genes in the

expected manner across two experimental repeats respectively (Figure 9E-H). Unexpectedly, CLV3 downregulation was also observed in GFP11-AtWUS treated plants. One possible explanation could be an AtWUS concentration-dependent regulation of the WUS-CLV3 axis (Rodriguez, K et al. (2022) *Sci Adv.* 8(32): eabo6157). The observed lower delivery efficacy of just the AtWUS derived CPP may also explain the difference in transcription modulating activity between GFP11-AtWUS-R9 construct and the GFP11-AtWUS at low, micro-molar concentrations.

[0103] Given the AtWUS sequence (SEQ ID NO: 1) derived from the 3rd alpha helix of WUSCHEL is cell-penetrating, additional sequences derived from the WUSCHEL homeodomain were investigated. Two shorter units of AtWUS (SEQ ID NO: 1) were also found to be cell-penetrating, AtWUS_2 (SEQ ID NO: 2) and AtWUS_3 (SEQ ID NO: 3) both successfully delivering GFP11 to *N. benthamiana* leaves expressing the DCIP reporter. AtWUS_3 (SEQ ID NO: 3) achieved the highest delivery efficiency at ~60% which is greater than R9 at equimolar 50 μ M concentrations of both (Figure 10A). Thus, the 3rd alpha helix of WUSCHEL, like most mammalian homeodomain proteins, is cell penetrating. Given mammalian homeodomain proteins are generally cell penetrating, it is likely that plant homeodomain proteins are also generally cell penetrating. Thus, a wide breadth of species was investigated including monocots, dicots, moss, green-algae, and red-algae (Figure 11A). A multiple sequence alignment of a library of 30 homeodomain proteins derived from the 14 classes of homeodomain proteins and a variety of species to the 3rd alpha helix of WUSCHEL was performed (Figure 11B). Sequence conservation was seen across all classes and species investigated. Therefore the 3rd alpha helix of plant homeodomain proteins is likely generally cell-penetrating. Given that about 200,000 plant homeodomain proteins can be found in the UniProt database (The UniProt Consortium. *Nucleic Acids Research.* 51:523-531. (2023)), this could be a rich source of novel, high efficiency CPPs for delivery of biomolecules to plant cells.

[0104] As a preliminary investigation, the cell-penetrating capability of over thirty distinct sequences derived from the 3rd alpha helix of plant homeodomain proteins, some of which are depicted in the sequence alignment in Figure 11B, was tested using DCIP. In general, most sequences were cell-penetrating, although some like AtSAWADEE are not cell-penetrating (Figure 12A). Some the tested sequences outperformed R9 delivery efficiency namely: CIWOX, AtNDX, OtWOX, NtWOX, and AtPHD. For the sequence CIWOX, greater delivery efficiency could be achieved by increasing the concentration plants are treated with (Figure 12B). Notably, across multiple concentrations CIWOX outperforms R9, confirming the superiority of homeodomain-derived cell penetrating peptides over canonical cell penetrating peptides, (Figure 12B). Delivery efficiency varies widely across sequences.

Notably, cross species cell-penetration was observed. For example, the sequence OsHOX12 is derived from rice (*Oryza sativa subsp japonica*) and was found to be cell-penetrating. Similarly, a sequence derived from red algae (*Cyanidioschyzon merolae*) was also found to be cell-penetrating.

[0105] In addition to the specific cell penetrating sequences shown in Figure 12A, the general design rules for peptides derived from the 3rd alpha helix of homeodomain proteins that can penetrate plant cells are also disclosed. The disclosed sequences are derived from all 14 classes of plant homeodomain proteins and from multiple evolutionary diverse species ranging from flowering plants (Magnoliopsida), ferns (Polypodiopsida), moss (Bryopsida), and algae (Chlorophyta), (Figures 11A & B). Thus, in general the 3rd alpha helix of plant homeodomain proteins is cell-penetrating and serves as a rich source of new high efficiency sequences for plant biotechnology applications. In general, it was observed that some 3rd alpha helix sequences were more efficient than others, as a design rule the high efficiency sequences were observed to have a characteristic pattern where certain amino acid positions are conserved, Figure 13A. From the analysis, certain amino acid positions were also found to be conserved for sequences that are not cell penetrating (Figure 13B). The emergent sequence pattern for a cell penetrating peptide derived 3rd alpha helix of homeodomain proteins is provided by the formula depicted in Figure 13C. These observations will help narrow the design space to find highly efficient cell-penetrating peptides from the space of 200,000 catalogued plant homeodomain protein sequences in UniProt.

What is claimed is:

1. A composition comprising a plant-derived cell-penetrating peptide (CPP) comprising an amino acid sequence derived from a plant homeodomain protein.
2. The composition of claim 1, wherein the homeodomain protein is a transcription factor.
3. The composition of claim 1, wherein the homeodomain protein comprises a sequence derived from WUSCHEL, CLSY2, PINTOX, LUMI, NDX, PHD, RLT2, BEL1, KNAT1, ZHD3, ATB16, HDG7, HAT1, REV, WOX3B, HOX12, Q8LLD8, WOX, Q40238, Q69G85 and CMJ244C.
4. The composition of claim 3, wherein the sequence comprises 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 amino acids.
5. The composition of claim 1, wherein the CPP is selected from the group consisting of: AtWUS_1 (SEQ ID NO: 1), AtWUS_2 (SEQ ID NO: 2), AtWUS_3 (SEQ ID NO: 3), At_SAW (SEQ ID NO: 4), At_PINTOX (SEQ ID NO: 5), At_LD (SEQ ID NO: 6), At_NDX (SEQ ID NO: 7), At_PHD (SEQ ID NO: 8), At_DDT (SEQ ID NO: 9), At_BEL (SEQ ID NO: 10), At_KNOX (SEQ ID NO: 11), At_PLINC (SEQ ID NO: 12), At_ZIPI (SEQ ID NO: 13), At_ZIPIV (SEQ ID NO: 14), At_ZIPII (SEQ ID NO: 15), At_ZIPIII (SEQ ID NO: 16), Zm_WOX (SEQ ID NO: 17), Os_HOX12 (SEQ ID NO: 18), St_PHD (SEQ ID NO: 19), Nt_WOX (SEQ ID NO: 20), Sp_DDT (SEQ ID NO: 21), Lj_NDX (SEQ ID NO: 22), Cm_BEL (SEQ ID NO: 23), Pp_ZIPIII (SEQ ID NO: 24), Cr_KNOX (SEQ ID NO: 25), Ot_WOX (SEQ ID NO: 26), Sm_ZIPIII (SEQ ID NO: 27), Pa_KNOX (SEQ ID NO: 28), Cl_WOX (SEQ ID NO: 29), Gg_BEL (SEQ ID NO: 30), Pt_WOX (SEQ ID NO: 31), and Cp_ZIPIII (SEQ ID NO: 32).
6. The composition of any one of claims 1-3, wherein the CPP comprises an amino acid sequence selected from the group consisting of SEQ ID Nos: 1-32 or a fragment thereof, analog or derivative thereof.

7. A composition comprising a CPP with an amino acid sequence selected from the group consisting of: (SEQ ID NO: 1) KNVFWFQNHKARERQKKRFN; (SEQ ID NO: 2) KNVFWFQNHKARERQ; and (SEQ ID NO: 3) HKARERQ.
8. A polynucleotide encoding a CPP, wherein the polynucleotide is selected from the group consisting of:
 - a) a polynucleotide encoding a peptide having at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID Nos: 1-32; and
 - b) a polynucleotide encoding a peptide comprising the amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 2 or SEQ ID NO: 3.
9. A complex for delivery of a biomolecule inside a cell comprising:
 - a) a CPP of any one of claims 1-6; and
 - b) a biomolecule wherein the biomolecule is fused at the N-terminal or C-terminal of the CPP.
10. The complex of claim 9, wherein the biomolecule is selected from the group consisting of a chemical compound, a protein or fragment thereof, a recombinant protein or fragment thereof, a glycoprotein or fragment thereof, a peptide, an antibody or fragment thereof, an enzyme or fragment thereof, a nuclease or fragment thereof, a hormone or fragment thereof, a cytokine or fragment thereof, a transcription factor or fragment thereof, a toxin or fragment thereof, a nucleic acid, a carbohydrate, a lipid, a glycolipid, a drug, a fluorophore, a fluorescent protein or fragment thereof, an antibiotic, a recombinase, and a plant hormone.
11. The complex of claim 9, wherein the biomolecule is fused via a linker.
12. The complex of claim 11, wherein the linker is a GSGS linker.
13. The complex of claim 10, wherein the biomolecule is selected from the group consisting of CRISPR associated protein 9 (CAS9), CAS12, CAS13, CAS14, CAS variants, CxxC-finger protein-1 (Cfp1), zinc-finger nucleases (ZEN) and transcription activator-like effector nuclease (TALEN).
14. The complex of claim 10, wherein the nucleic acid is selected from the group consisting of DNA, RNA, antisense oligonucleotide (ASO), microRNA (miRNA), small interfering RNA (siRNA), aptamer, locked nucleic acid (LNA), peptide nucleic acid (PNA), and morpholino.

15. The complex of claim 10, wherein the recombinant protein is selected from the group consisting of a morphogenic protein, a growth factor, a receptor, a signaling protein, a membrane protein, and a transmembrane protein.
16. The complex of claim 10, wherein the nuclease is an RNA-guided endonuclease, a CRISPR endonuclease, a type I CRISPR endonuclease, a type II CRISPR endonuclease, a type III CRISPR endonuclease, a type IV CRISPR endonuclease, a type V CRISPR endonuclease, a type VI CRISPR endonuclease, CRISPR endonuclease, CRISPR associated protein 9 (Cas9), Cpf1, a zinc-finger nuclease (ZFNs), a Transcription activator-like effector nucleases (TALENs), a homing endonuclease, or a meganuclease.
17. The complex of claim 16, wherein the nuclease is a CRISPR endonuclease further comprising a guide RNA, a crRNA, a tracrRNA, or both a crRNA and a tracrRNA.
18. A method for delivering a biomolecule into a plant cell, the method comprising contacting said plant cell with a composition according to any one of claims 1-6 or a complex according to any one of claims 9-17, under conditions that allow the CPP to penetrate the plant cell.
19. A method of identifying a CPP capable of transporting a biomolecule to a subcellular location in a plant, the method comprising the steps of:
 - a) fusing a biomolecule of interest to a CPP according to any one of claims 1-6;
 - b) contacting a plant cell with a composition comprising the CPP of (a);
 - c) performing an assay to determine the ability of the CPP to translocate the biomolecule to a subcellular location of the cell.
20. The method of claim 19, wherein the biomolecule is a fluorophore and the assay is a delivered complementation *in planta* (DCIP) assay.
21. A method of quantitating protein delivery comprising the steps of:
 - a) fusing a protein of interest to a CPP according to any one of claims 1-6;
 - b) contacting a plant cell with a composition comprising the CPP-protein fusion of (a); and

c) performing an assay to quantitate the amount or number of CPP-protein fusion to translocate a subcellular location of the cell.

22. The method of claim 21, wherein the protein is a fluorophore and the assay is a delivered complementation *in planta* (DCIP) assay.

Figure 1

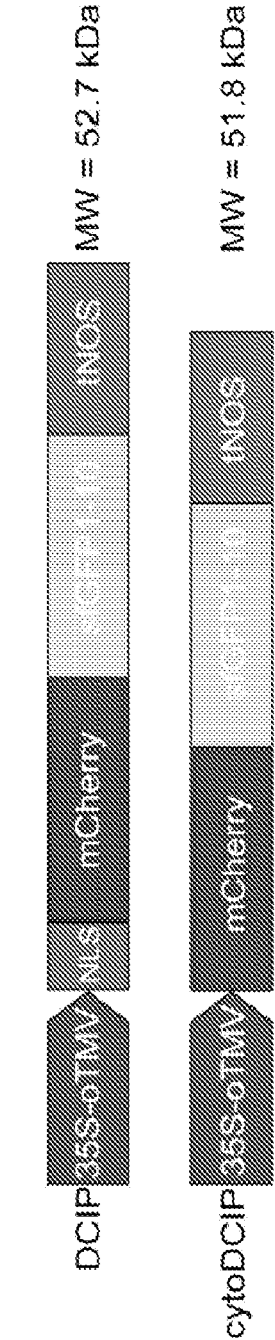
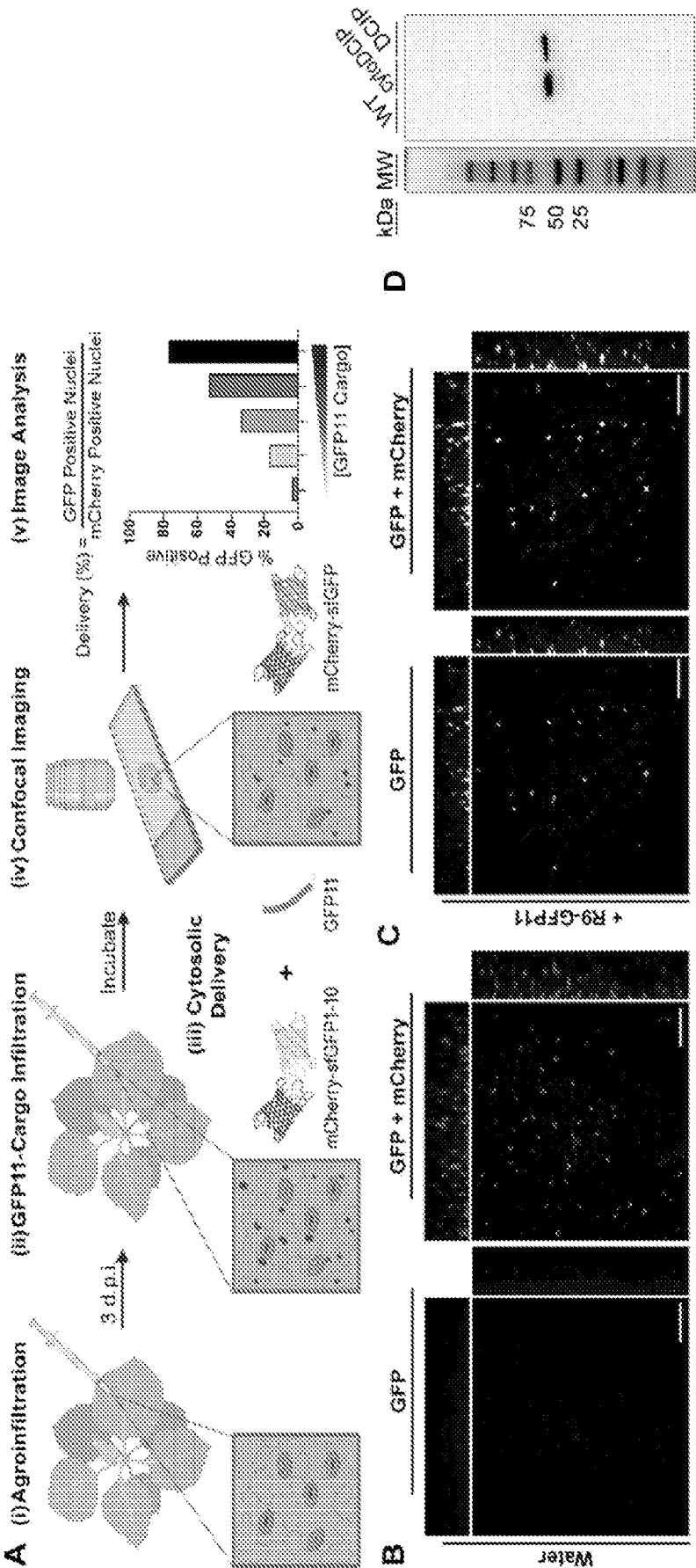


Figure 2



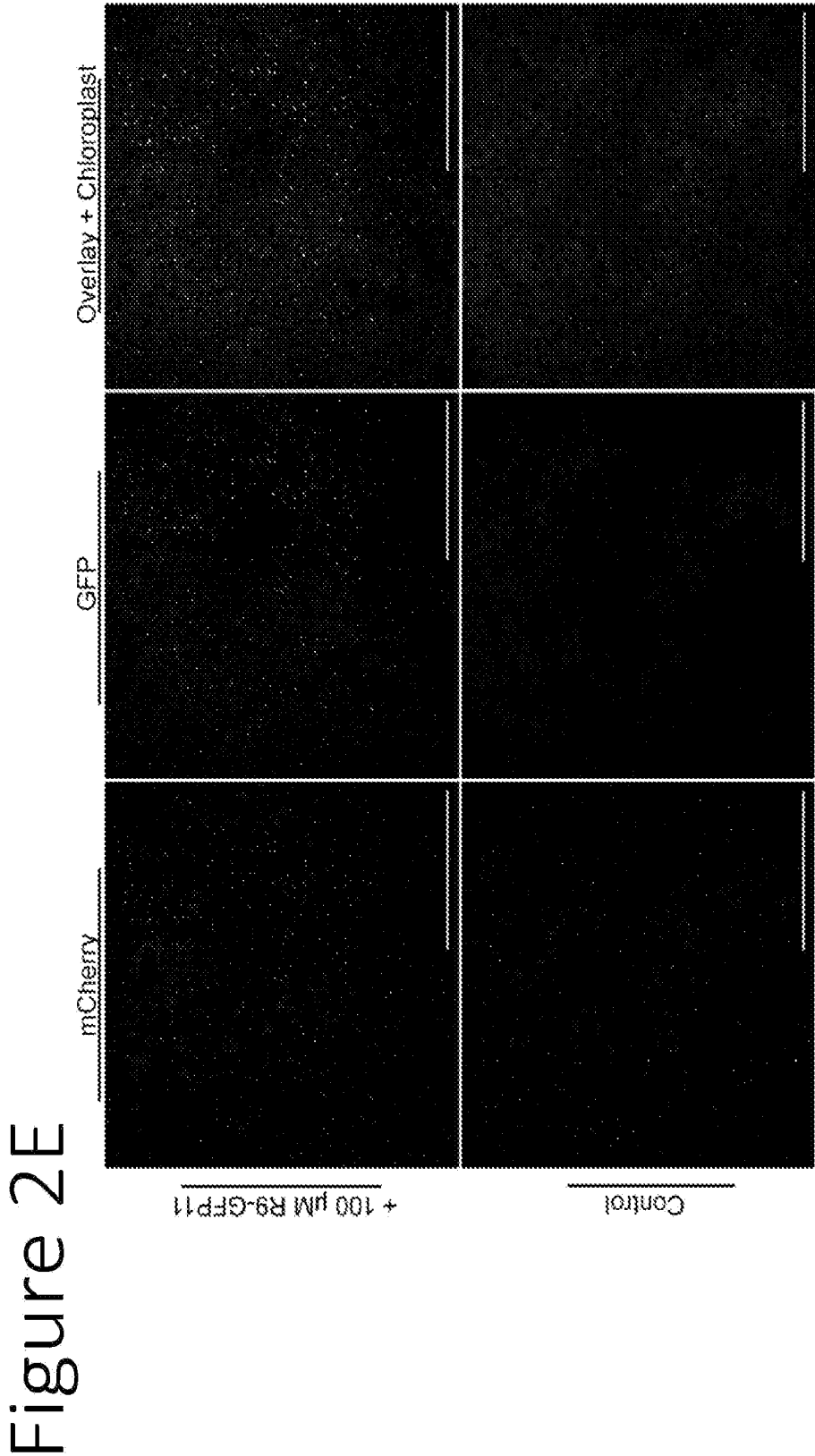


Figure 3 A-F

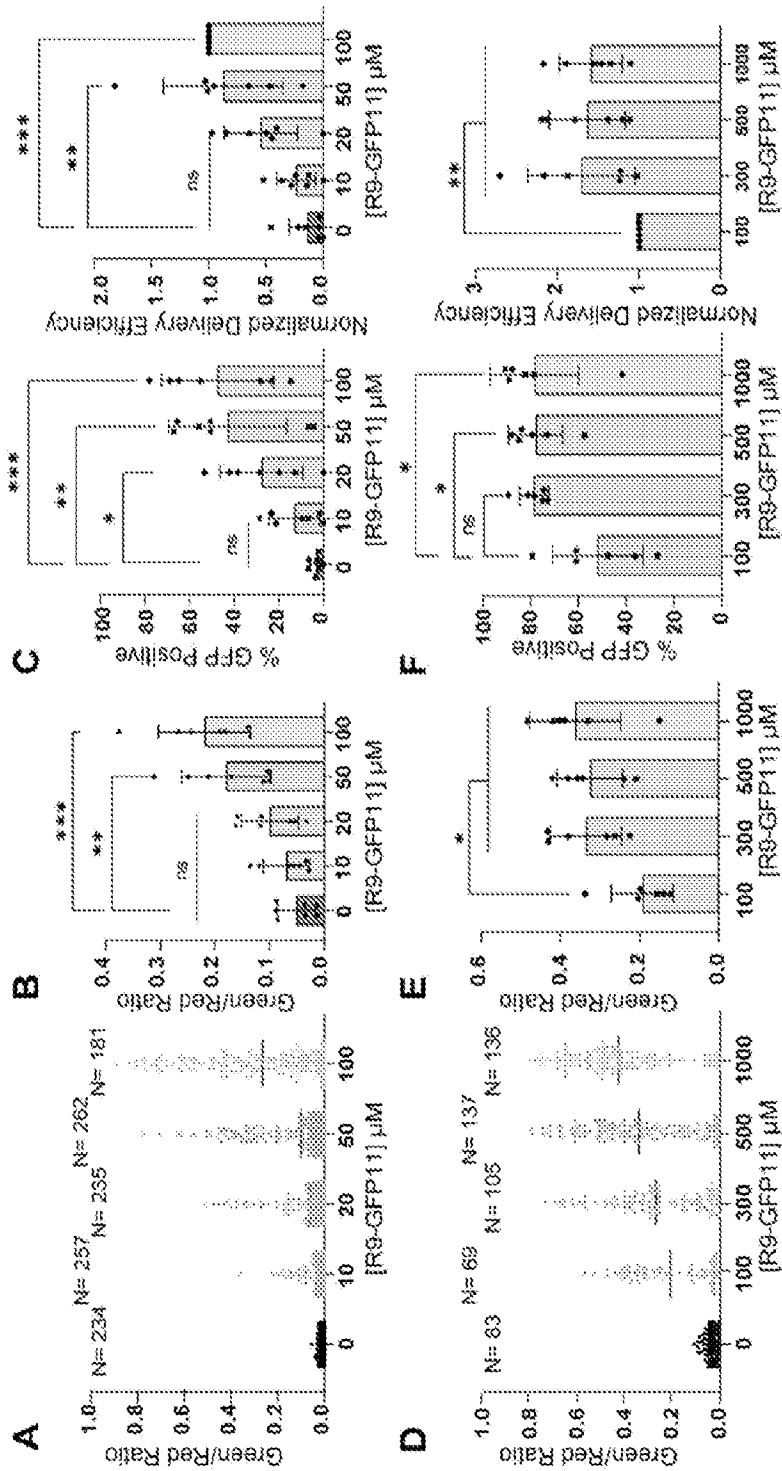


Figure 3G

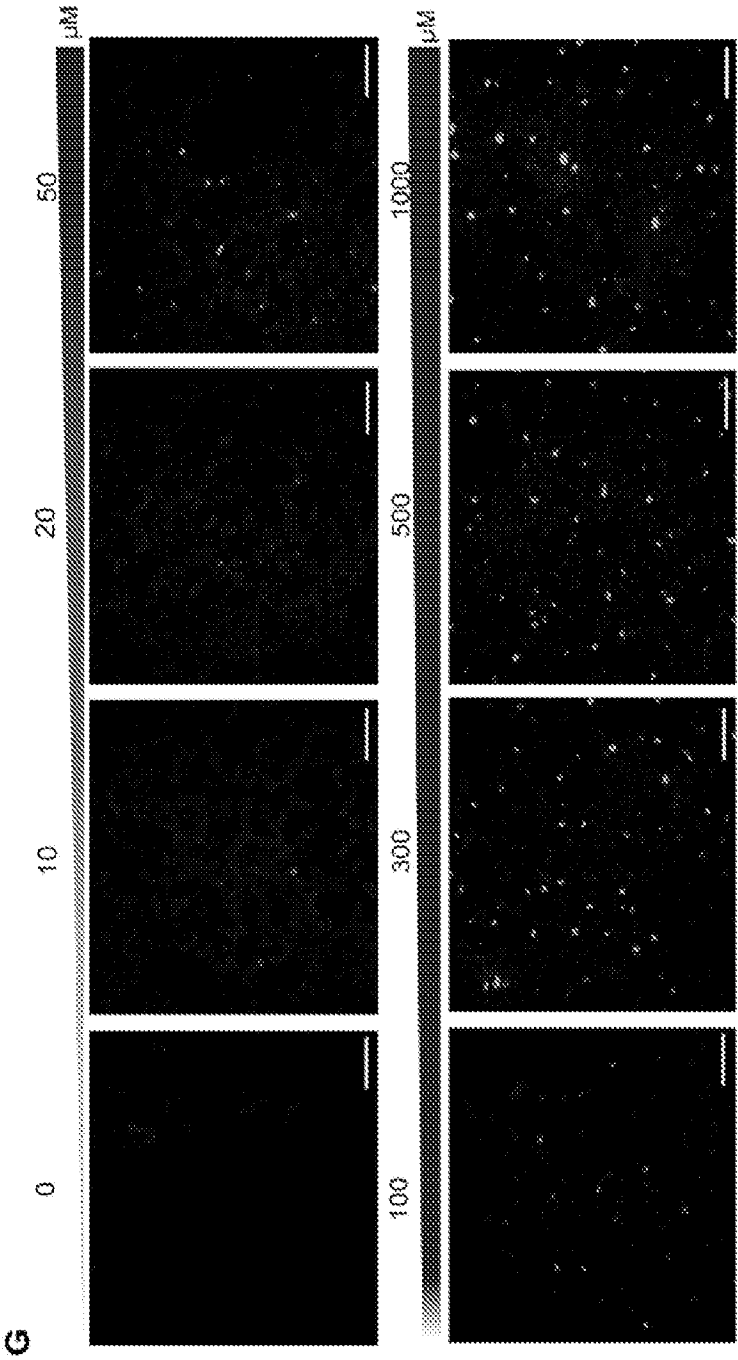


Figure 4

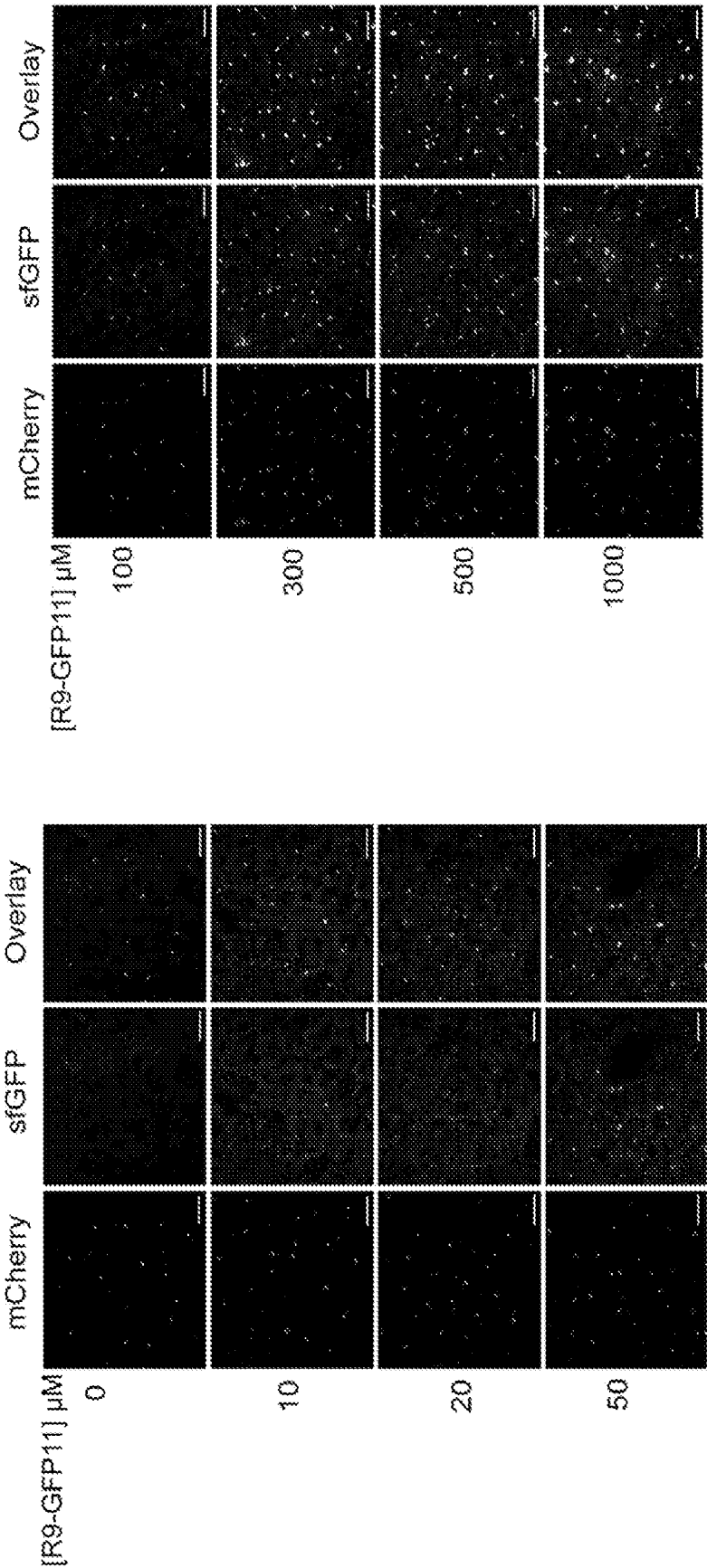


Figure 5 A-C

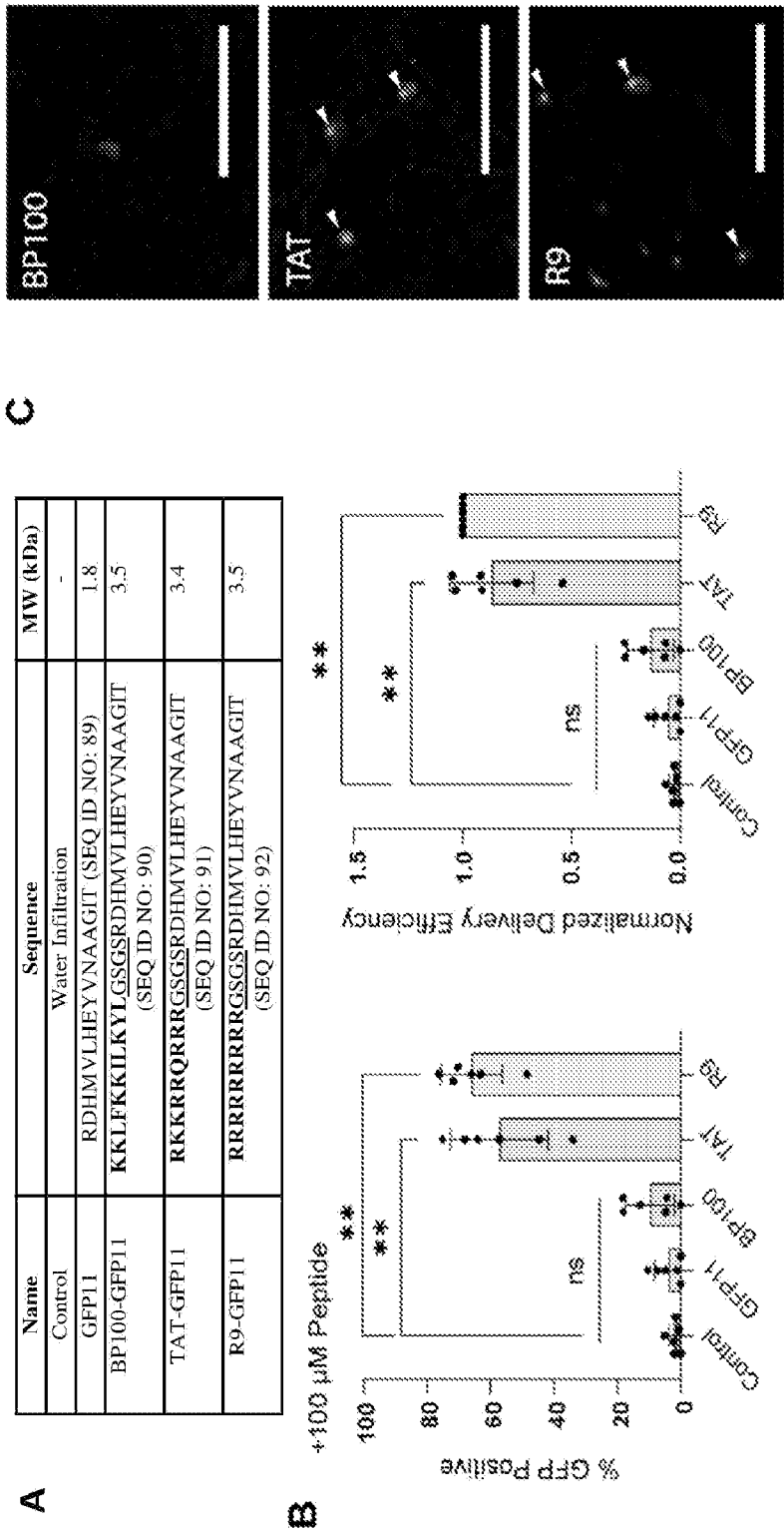


Figure 5 D-F

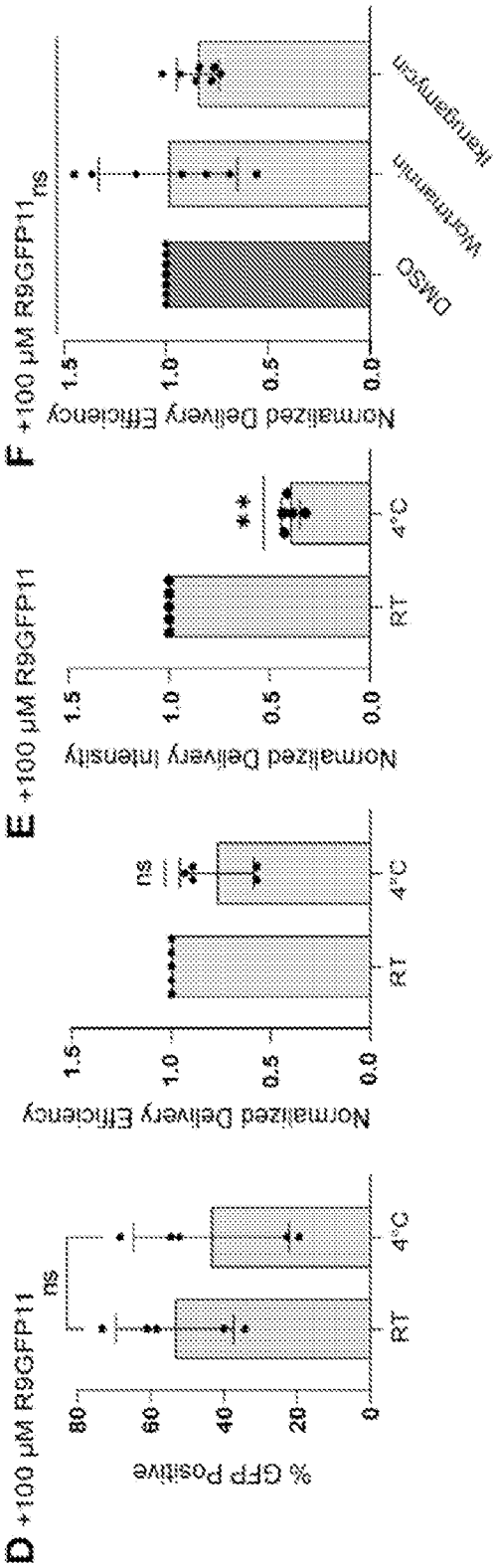


Figure 6

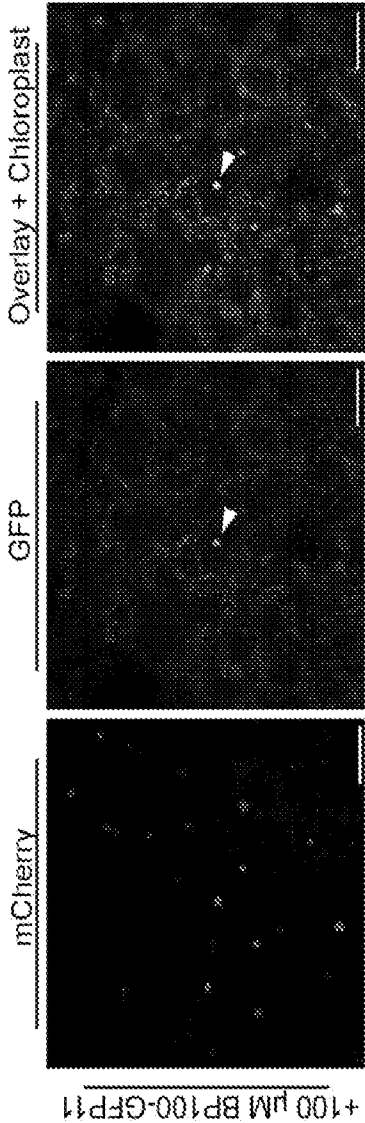


Figure 7

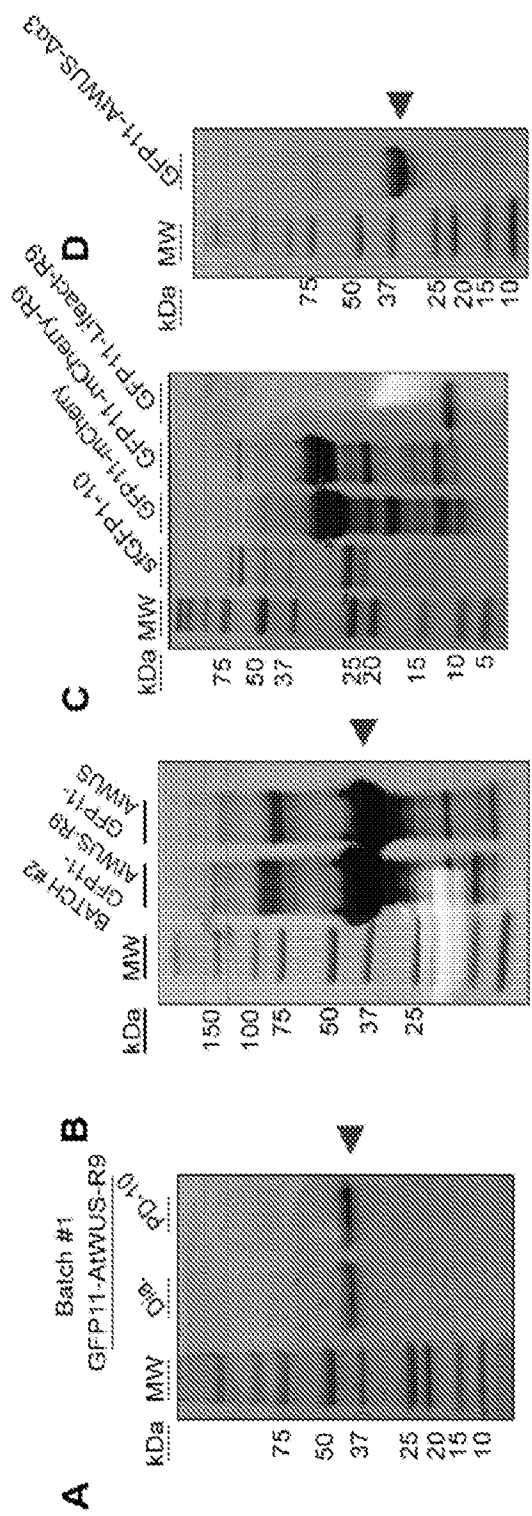


Figure 8

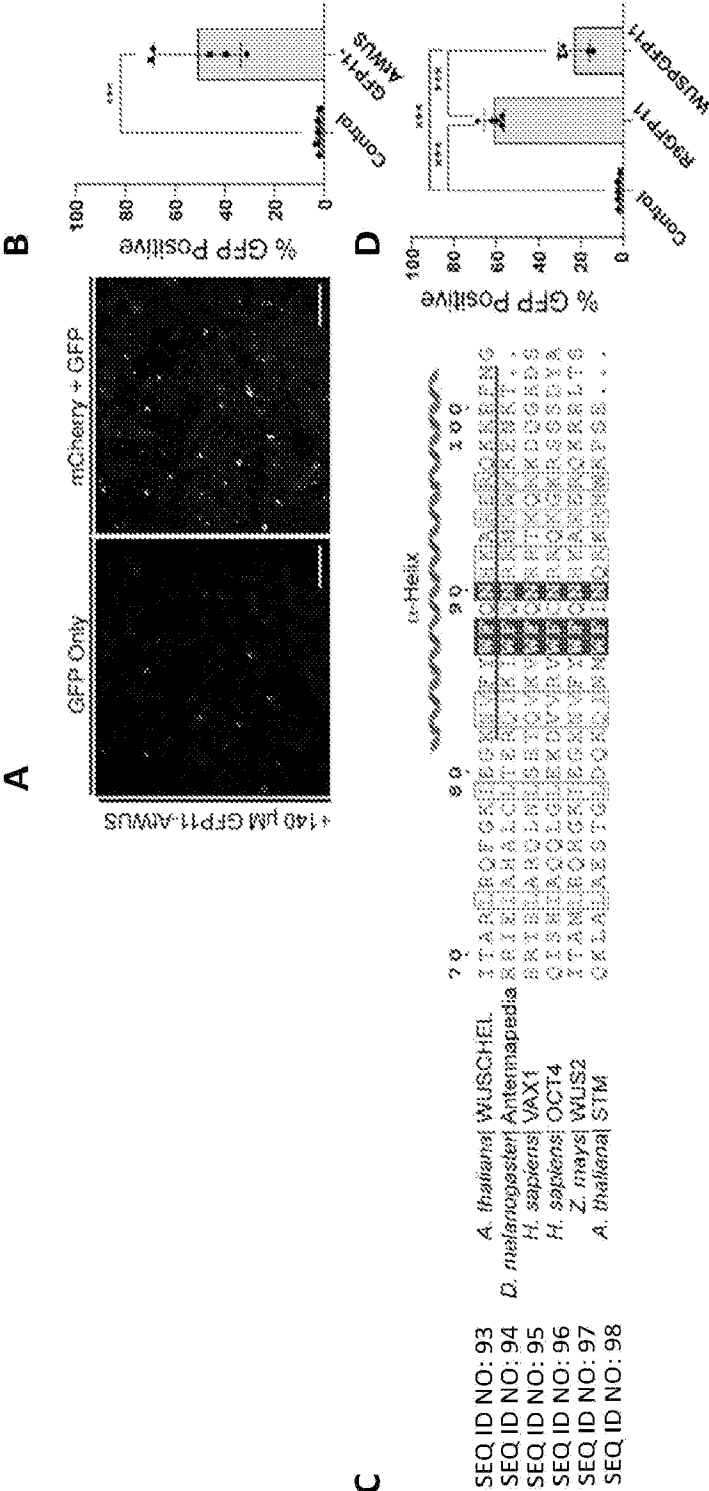


Figure 9A & B

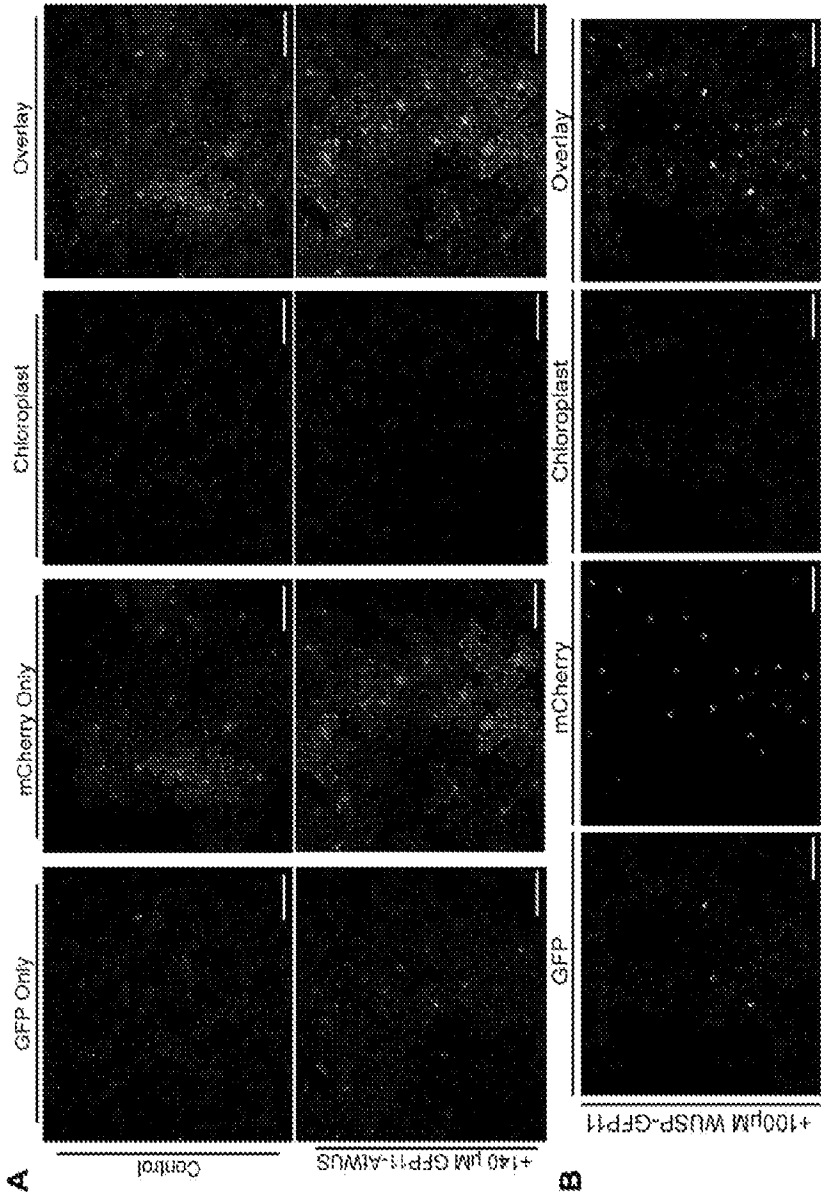
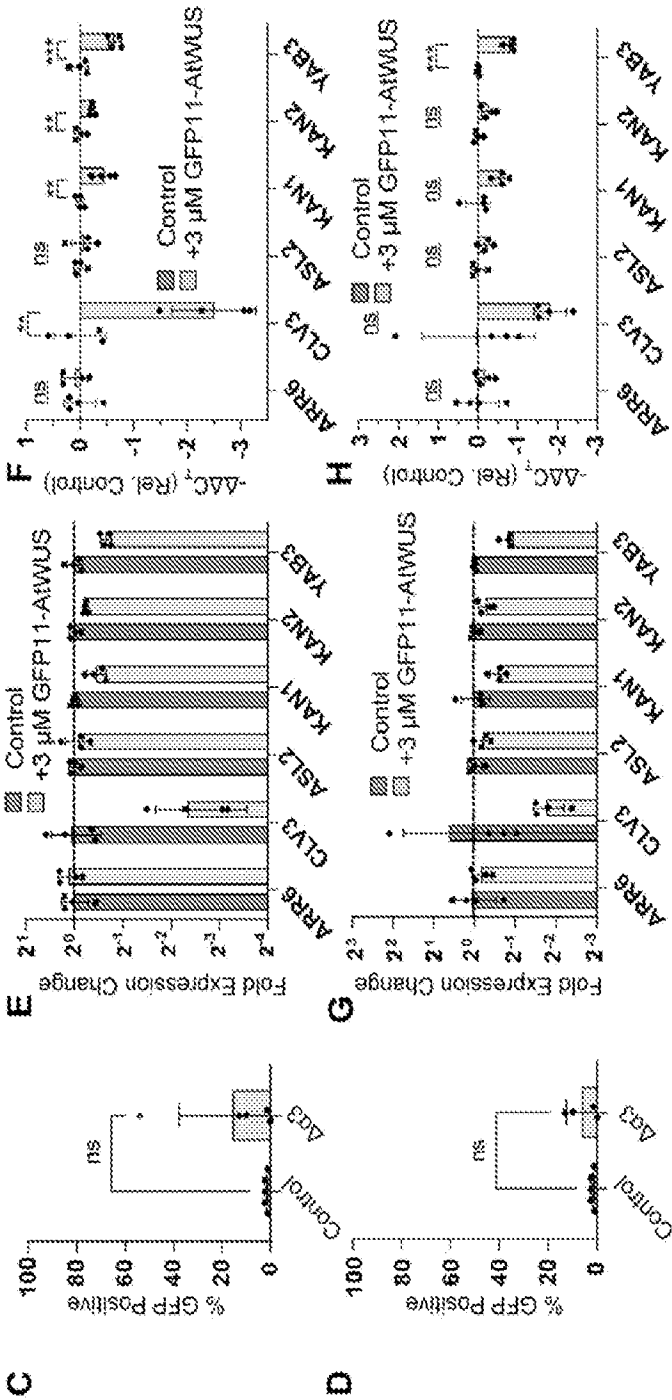


Figure 9C-H



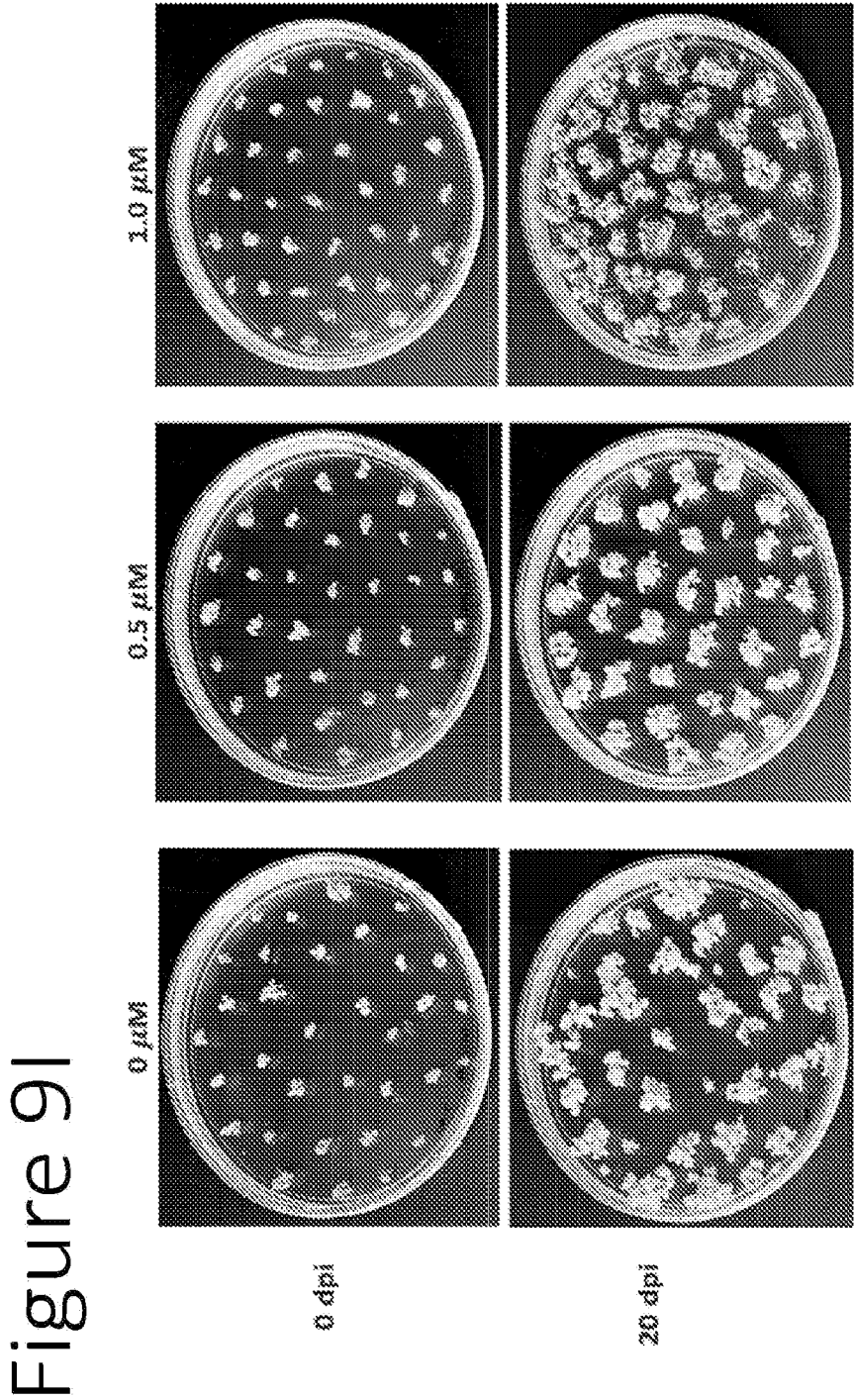
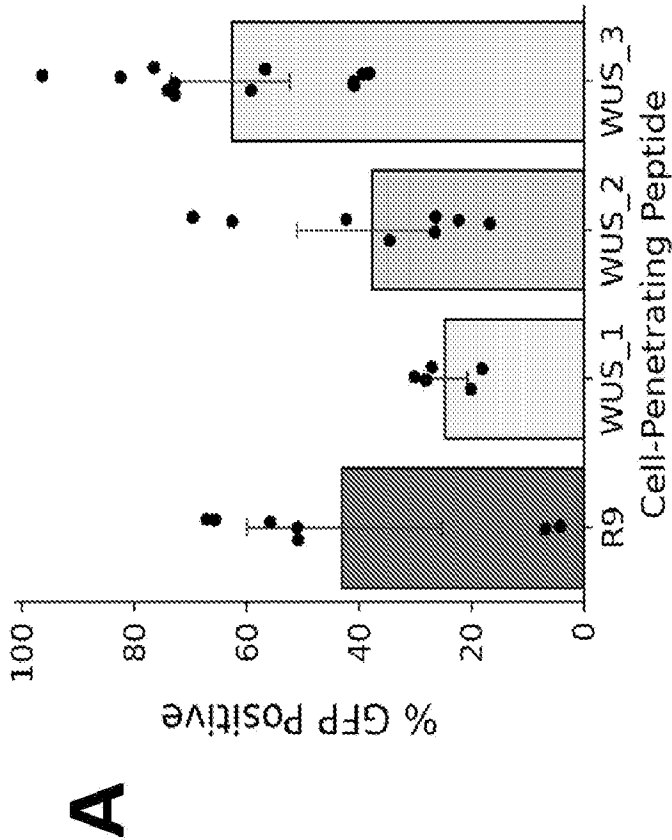


Figure 10



B

Name	Sequence
R9	RRRRRRRR (SEQ ID NO: 99)
WUS_1	KNVFWFQNHKARERQKKRFN (SEQ ID NO: 1)
WUS_2	KNVFWFQNHKARERQ (SEQ ID NO: 2)
WUS_3	HKARERQ (SEQ ID NO: 3)

Figure 11A

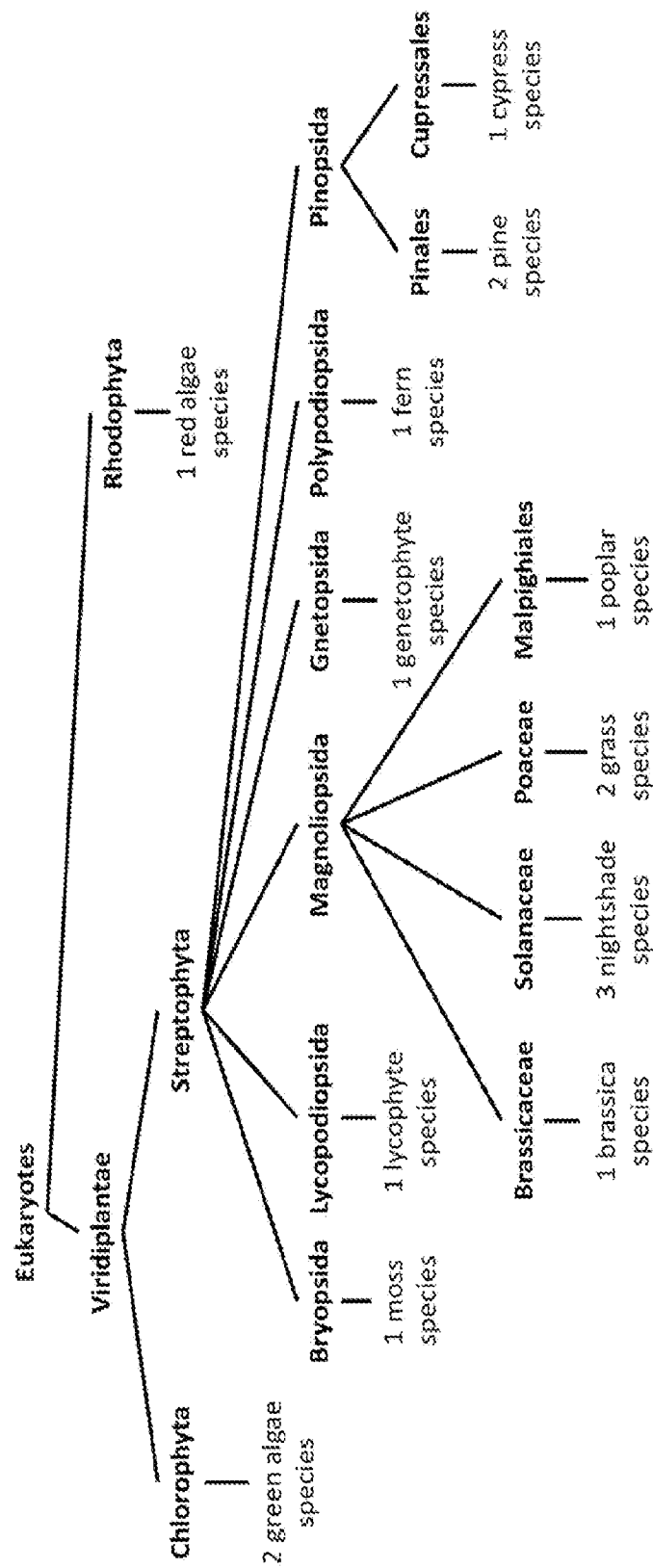


Figure 11B

HD Class Variation		Species		α3	
HD Class		Species		90	
WOX	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
SAWADEE	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
PINTOX	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
LD	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
NDX	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
PHD	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
DDT	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
BEL	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
KNOX	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
PLINC	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
HD-ZIP-I	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
HD-ZIP-IV	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
HD-ZIP-II	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
HD-ZIP-III	Arabidopsis thaliana	KQIVVF	YVVF	QVVF	QVVF
WOX	Zea mays	KQIVVF	YVVF	QVVF	QVVF
HD-ZIP-II	Oryza sativa	KQIVVF	YVVF	QVVF	QVVF
WOX	Nicotiana tabacum	KQIVVF	YVVF	QVVF	QVVF
DDT	Solanum peruvianum	KQIVVF	YVVF	QVVF	QVVF
NDX	Brassica napus	KQIVVF	YVVF	QVVF	QVVF
BEL	Cyanidiaschyzon meralae	KQIVVF	YVVF	QVVF	QVVF
HD-ZIP-III	Physcomitrium patens	KQIVVF	YVVF	QVVF	QVVF
KNOX	Chlamydomonas reinhardtii	KQIVVF	YVVF	QVVF	QVVF
WOX	Ostreococcus tauri	KQIVVF	YVVF	QVVF	QVVF
HD-ZIP-III	Selaginella moellendorffii	KQIVVF	YVVF	QVVF	QVVF
KNOX	Picea abies	KQIVVF	YVVF	QVVF	QVVF
WOX	Cunninghamia lanceolata	KQIVVF	YVVF	QVVF	QVVF
BEL	Gnetum gnemon	KQIVVF	YVVF	QVVF	QVVF
PHD	Solanum tuberosum	KQIVVF	YVVF	QVVF	QVVF
WOX	Populus trichocarpa	KQIVVF	YVVF	QVVF	QVVF
HD-ZIP-III	Ceratopteris pteridoides	KQIVVF	YVVF	QVVF	QVVF

[illegible]

Figure 12B

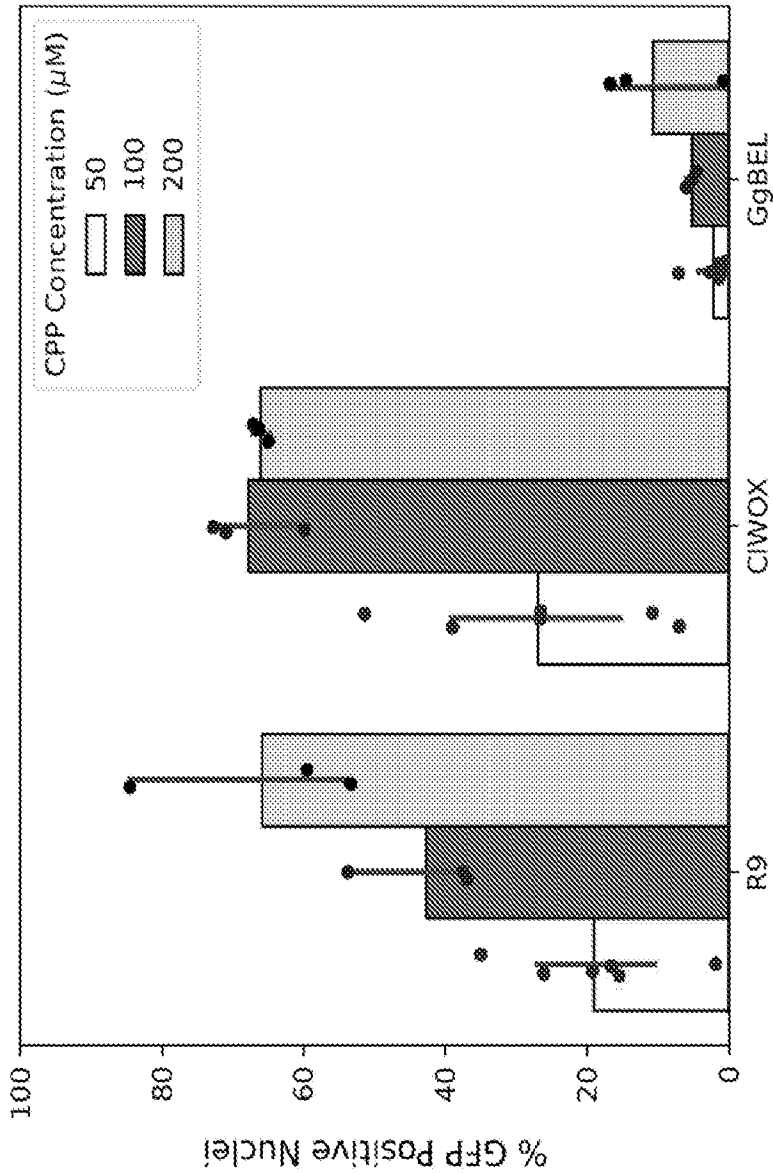


Figure 13

