ADVANCED BIOLOGY
www.advanced-bio.com

# C-Terminal Domain Controls Protein Quality and Secretion of Spider Silk in Tobacco Cells

*Chonprakun Thagun, Tomohiro Suzuki, Yutaka Kodama, and Keiji Numata**

**The remarkable mechanical strength and extensibility of spider dragline silk spidroins are attributed to the major ampullate silk proteins (MaSp). Although fragmented MaSp molecules have been extensively produced in various heterologous expression platforms for biotechnological applications, complete MaSp molecules are required to achieve instinctive spinning of spidroin fibers from aqueous solutions. Here, a plant cell-based expression platform for extracellular production of the entire MaSp2 protein is developed, which exhibits remarkable self-assembly properties to form spider silk nanofibrils. The engineered transgenic Bright-yellow 2 (BY-2) cell lines overexpressing recombinant secretory MaSp2 proteins yield 0.6–1.3 μg $L^{-1}$ at 22 days post-inoculation, which is four times higher than those of cytosolic expressions. However, only 10–15% of these secretory MaSp2 proteins are discharged into the culture media. Surprisingly, expression of functional domain-truncated MaSp2 proteins lacking the C-terminal domain in transgenic BY-2 cells increases recombinant protein secretion incredibly, from 0.9 to 2.8 mg $L^{-1}$ per day within 7 days. These findings demonstrate significant improvement in the extracellular production of recombinant biopolymers such as spider silk spidroins using plant cells. In addition, the results reveal the regulatory roles of the C-terminal domain of MaSp2 proteins in controlling their protein quality and secretion.**

## 1. Introduction

Spider silk has been used as a biologically-derived substance long before it has become one of the major biomaterials in research. Due to the excellent properties such as high mechanical strength and extendibility, and more especially, biocompatibility and biodegradability compared to other polymer-based biomaterials, spider silk has exceptionally been utilized in various biotechnological and biomedical applications.[1] Spidroins are generally large structural protein (>300 kDa) composites in web- and dragline silk fibers.[2,3] In the case of dragline silk, major ampullate spidroin proteins, MaSp1 and MaSp2, are produced and secreted from the epidermal cells of major ampullate gland and majorly contribute to the extraordinary mechanical properties of silk fibers.[4]

A MaSp monomer usually comprises three imperative domains.[1,2,5] These functionally- and structurally-important domains include an N-terminal domain (NTD; N) that functions in fabrication of MaSp, the glycine/alanine (Gly/Ala)-rich repetitive region (R) that plays dynamic roles in $\beta$-sheet formation, and a cryptic C-terminal domain (CTD; C), for which its correlative function is still unclear.[2,4,6] Intermolecular interaction of NTD and CTD, and supercontraction of repetitive region of MaSp stringently control molecular assembly to form droplet-like structures and, finally, spider silk fibers in liquid-phase in natural conditions and in vitro.[3,6] However, the self-assembling nature and formation of silk fiber are concentration-dependent and flexibly organized by environmental ionic strengths.[6]

Heterologous production of MaSp and other spidroins using recombinant hosts and synthetic biology approach enables an economical-scale manufacturing of silk-based biomaterials. Usually, artificial spidroins including copious repetitive regions of MaSp and others have been effectively produced in various expression platforms.[7] Prokaryotic cells especially *Escherichia coli* established tremendous performance in production of larger size (≈100–556 kDa) and greater yields (0.5–2.7 g protein $L^{-1}$) of Gly/Ala-rich repeats of MaSp1.[8,9] Contrary, limited lengths (22–300 kDa) and lower production yield of artificial MaSp1 and MaSp2 could also be strictly made in higher eukaryotic expression platforms including yeasts,[10,11] transgenic plants,[12,13]

C. Thagun, K. Numata
Department of Material Chemistry
Graduate School of Engineering
Kyoto University
Kyoto-Daigaku-Katsura
Nishikyo-ku, Kyoto 615–8510, Japan
E-mail: Keiji.Numata@riken.jp
C. Thagun, T. Suzuki, Y. Kodama
Center for Bioscience Research and Education
Utsunomiya University
Tochigi 321–8505, Japan
Y. Kodama, K. Numata
Biomacromolecules Research Team
RIKEN Center for Sustainable Resource Science
2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan
The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adbi.202300011



**DOI: 10.1002/adbi.202300011**

animal cell cultures,[14,15] and transgenic animals.[16–19] Moreover, production of native spidroins in heterologous expression systems is still problematic due to the presence of highly tandem nucleotide- and amino acid repeats of spidroins, and the difference of codon usage in various hosts. In addition, post-translational modification in different eukaryotic hosts can unpredictably influence self-assembly and mechanical characteristics of recombinant spidroins.[20,21]

Plant expression system provides greater advantages over other transgenic hosts in recombinant protein production. Integration of transgene expression cassettes in plant genomes offers stable expression of larger proteins in a cost-effective cultivation environment.[22,23] Moreover, recombinant proteins can be designated to specific plant organelles such as plastids, endoplasmic reticulum (ER), vacuoles, and lipid bodies to improve productivity and simplify downstream purification procedures.[24] Alternatively, secretory production in the up-scalable biosecurity fermenters allows improved accumulation and stability of extracellular proteins that can be successively isolated from liquid culture of plant cells.[25,26] For instance, fusion of recombinant protein to secretory peptide and numerous repeats of Serine-Proline (in hydroxyproline form) glycopeptide motif impressively boost the accumulation and secretion of GFP and human interferon in tobacco cell suspension cultures.[26–29]

Recent progress in exogenous expression of recombinant spidroins in plant cells is still in substantial progress. The first successful development enabled extensive accumulation of ER-targeted chimeric MaSp1 (>100 kDa) up to 2% of total soluble proteins extracted from *Nicotiana tabacum* and *Solanum tuberosum*.[5,12] The largest molecule of spidroins that has even been produced in transgenic tobacco seeds is 250–450 kDa of flagelliform silk multimers.[30,31] Using similar seed-specific expression technology, more than 18% of total proteins of transgenic *Arabidopsis thaliana* seeds could be remarkably produced via ER-targeted accumulation of synthetic analogue of spidroins.[13] However, production yield of plant-derived spidroins remains considerably lower than that achieved from *E. coli*.[8] Certainly, the whole MaSp2 molecule must be entirely produced as secretory proteins to enable self-assembly and artificial spinning of silk fiber from extracellular media of plant cells. Moreover, optimal purification and functionalization are other challenging steps in commercial-scale manufacturing of plant-derived recombinant spidroins.

Inspired by a natural secretion of spider silk proteins, we primarily created a tobacco BY-2 culture-based secretory production system to produce extracellular recombinant MaSp2 proteins, another MaSp that is remarkably capable to form nanofibrils in liquid media.[6,32] Our results demonstrated that recombinant MaSp2 proteins did not substantially accumulate in cytosol of transgenic BY-2 cells. Targeting recombinant MaSp2 proteins to plant secretory pathways impressively increased the accumulation of target proteins in BY2 cell suspension cultures. However, the secretory MaSp2 proteins were hardly released by BY-2 cells to culture media. Our further investigation revealed an interesting fact that the C-terminal domain of MaSp2 not only regulates intermolecular interactions to form homodimers but also strictly controls its protein quality in the cytosol and the secretion of recombinant MaSp2 protein through plant secretory pathways.

## 2. Results and Discussion

### 2.1. Construction of Expression Vectors

Plant expression vectors were constructed for production of recombinant MaSp2 proteins comprising the N-terminal domain of *Trichonephila clavipes* MaSp2 (NTD, N), 6 and 12 repeats of repetitive regions (R6 and R12, respectively), and C-terminal domain derived from *Lathodectus hesperus* MaSp2 protein (CTD, C) in tobacco cells (**Figure 1**A).[6] To facilitate the subsequent affinity-based detection and purification, and recovery of recombinant MaSp2 proteins, the coding sequences of GFP(S65T), the Strep-Tac II (WSHPQFEK), and the TEV-protease cleavage peptide (ENLYFQS) were fused to nucleotide sequences encoding MaSp2 NR6C and NR12C proteins at their N-termini. This resulted in the coding sequences for GFP-NR6C and GFP-NR12C for cytosolic expression. (Figure 1A; Figure S1, Supporting Information). GFP-NR6C and GFP-NR12C fragments were then ligated to nucleotides encoding a signal peptide of the tobacco extensin ($S_{tab}$; MGKMASLFASLLVVLVSLSLA) and 32 repeats of Serine–Proline dipeptides ($(SP)_{32}$)[26,29,33] at their N-termini, giving in the SS-GFP-NR6C and SS-GFP-NR12C constructs for secretory production of recombinant MaSp2 proteins (Figure 1B; Figure S1, Supporting Information). Both cytosolic- and secretory expression cassettes of recombinant MaSp2 NR6C and NR12C proteins were under transcriptional regulation by cauliflower mosaic virus *35S* promoter (CaMV 35S pro) and nopaline synthase terminator (NOS ter) sequences in plant expression vector (Figure 1A,B; Figure S1, Supporting Information). Table S1, Supporting Information, displays the polypeptide compositions and respective theoretical molecular weights of the resulting chimeric MaSp2 proteins.

Successful expression of recombinant spidroins in plant cells could be prior examined by transient expression.[34] As a confirmatory step, we transiently expressed GFP-fused MaSp2 NR6C and NR12C proteins in tobacco (*N. tabacum*) leaf cells via *Agrobacterium*-mediated leaf infiltration (agroinfiltration) (**Figure 2**A). Confocal laser scanning microscopy (CLSM) imaging and immunoblotting results showed that cytosolic GFP-NR6C and GFP-NR12C proteins did not adequately accumulate in tobacco leaf cells at 3–4 days post infiltration (DAI) (Figure 2B,C; Figure S2, Supporting Information). In contrast, transient expression of secretory SS-GFP-NR6C and SS-GFP-NR12C constructs exhibited an improved accumulation of recombinant proteins in tobacco leaves, greatly higher than the cytosolic expressions (Figure 2B–D; Figure S2, Supporting Information). We also noted that the accumulation of SS-GFP-NR12C was significantly lower than SS-GFP-NR6C in agroinfiltrated tobacco leaves, suggesting that longer recombinant MaSp2 proteins are more strenuous to synthesize in plant cells (Figure 2C,D; Figure S3, Supporting Information). The presence of lengthier nucleotide sequence of repetitive regions in *MaSp2 NR12C* transcripts may cause premature termination of translation instigated by tRNA starvation.[35] The other possibility is an anonymous gene silencing mediated by highly-tandem mRNA sequences for the repetitive regions of MaSp2.[36]

Various affinity tags and constrictive purification domains, such as polymeric histidine (His-tag), hemagglutinin (HA), FLAG-octapeptide, and intein-splicing domain, have been

 

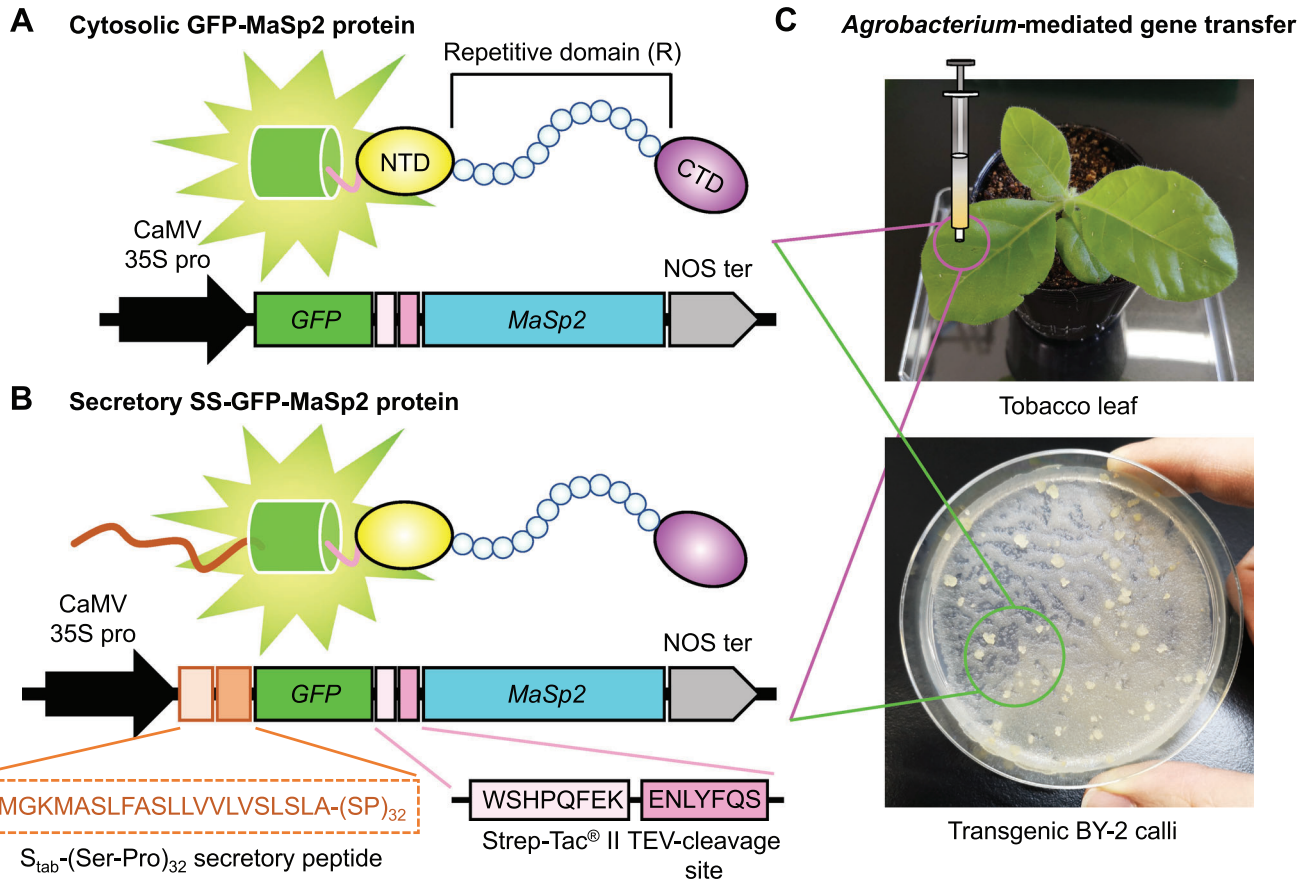


**Figure 1.** Construction of gene expression cassettes for production of plant-derived recombinant MaSp2 spidroin. Recombinant MaSp2 protein expression cassettes were designed for A) expression in the cytosol or B) secretion of tobacco cells. These cassettes were constructed synthetically by assembling nucleotide sequences coding for the N-terminal domain (NTD) of MaSp2 from *Trichonephila clavipes*, the C-terminal domain (CTD) of *Lathodectus herperus* MaSp2, and 6 or 12 repeats of Gly/Ala-rich repetitive domains. The resulting genes, *MaSp2 NR6C* and *NR12C*, were fused to a *GFP(S65T)* gene and a coding region of the Strep-Tac II/TEV-protease cleavage site at their N-termini. To target recombinant proteins to the plant secretory pathways, a $S_{tab}$-(Ser-Pro)$_{32}$ secretory peptide-coding sequence was integrated into the GFP-MaSp2 gene fragments at the N-termini (B). The gene expression cassettes were regulated by the cauliflower mosaic virus 35S gene promoter (CaMV 35S pro) and nopaline synthase gene terminator (NOS ter) for both cytosolic (A) and secretory constructs (B). C) These cassettes were transformed into *N. tabacum* (tobacco) leaves (upper) and Bright-Yellow 2 (BY-2) cells (lower) using *A. tumefaciens*-mediated gene transfer techniques to produce plant-derived recombinant MaSp2 proteins. Figure 1 illustrates the gene constructs for overexpression of cytosolic GFP-NR12C and secretory SS-GFP-NR12C.

utilized to aid in the isolation of recombinant spidroins from plants.[12,30,31,34,37] Nevertheless, an alternative Strep–Tag system has gained more interest conferring its economically-feasible one-step isolation of plant-based recombinant proteins and recycling column for maximizing protein purification capacity.[38–40] We rationally included Strep–Tac II peptides and TEV protease cleavage site to recombinant proteins to expedite downstream isolation of plant-derived MaSp2 via the commercially-available Strep–Tactin column (Figures 1 and 2A,E). Due to their higher abundances in tobacco leaves, we subsequently purified recombinant GFP-NR6C and SS-GFP-NR6C proteins from the tobacco leaf proteins by Strep–Tactin resin and performed on column TEV-protease digestion to validate this strategical protein purification prospect (Figure 2E). Immunoblotting result showed that both of the recombinant proteins could be enriched from tobacco leaf proteins, even the cytosolic GFP-NR6C which was lowly expressed (Figure 2F; Figure S4, Supporting Information). In addition, recombinant proteins bound to the Strep–Tactin matrix could be successively cleaved by TEV-protease to discharge native MaSp2 NR6C proteins from the secretory $S_{tab}$-(SP)$_{32}$ peptide and fusion tags (Figure 2G,H; Figure S5A,B, Supporting Information). Furthermore, the existence of peptide fragments corresponding to the TEV-protease cleavage products of recombinant GFP-NR6C and SS-GFP-NR6C proteins was elucidated by mass spectrometry, indicating that these recombinant proteins were correctly expressed in plant cells (Figure S5, Supporting Information). Our results suggest that the recombinant MaSp2 proteins comprising chimeric domains can be expressed and subsequently purified from transformed tobacco leaves.

### 2.2. Expression of Recombinant MaSp2 Proteins in Tobacco BY-2 Cell Lines

Plant cell cultures enable cost-effective biotechnological production of biomolecules such as bioactive compounds and

 

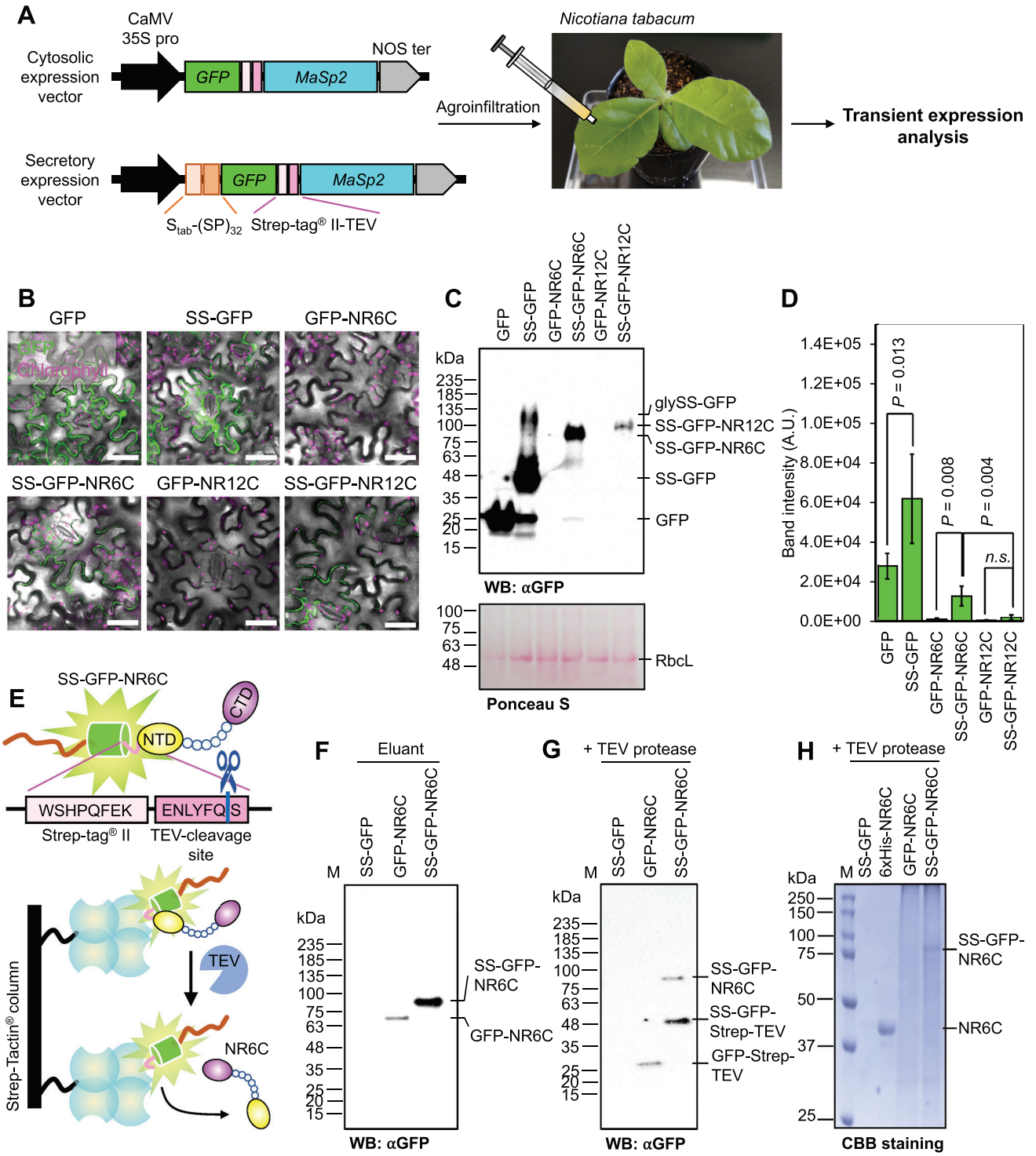


**Figure 2.** Transient expression of recombinant MaSp2 proteins in agroinfiltrated tobacco leaves. A) The expression of recombinant MaSp2 proteins in agroinfiltrated tobacco leaf cells was performed by infiltrating tobacco leaves with *A. tumefaciens* cells carrying cytosolic and secretory expression vectors of recombinant MaSp2 proteins. B) Localization of GFP-fused MaSp2 proteins in plant cells examined by CLSM at 3 days post-infiltration (DAI). Scale bars = 50 µm. C) Immunoblotting results of various GFP-fused recombinant proteins in total leaf proteins extracted from agroinfiltrated tobacco leaves. The upper panel displays the immunoblot results against the anti-GFP (αGFP) antibody, where glySS-GFP refers to the chemiluminescent band of post-translationally glycosylated SS-GFP protein. The membrane was first stained with Ponceau-S solution to determine the relevant loading of various protein samples on the membrane, as indicated in the lower panel. RbcL indicates bands corresponding to rubisco large subunit (RbcL) protein on the membrane. D) Quantitative analysis of immunoblotting results of GFP-fused recombinant MaSp2 proteins. The band intensity in the immunoblot images of membranes against anti-GFP antibody was quantified using Fiji ImageJ. The error bars represent the standard deviation of means from three biologically independent experiments ($N = 3$). The statistical difference in the mean was analyzed using the paired-sample *t*-test, where n.s. indicates no

 

recombinant proteins in the industrial-scale bioreactors.[25,41] In addition, plant cells implement several consequential post-translational modifications of recombinant proteins which are hard to accomplish in prokaryotic- and some eukaryotic cell expression systems.[42] We subsequently transformed tobacco Bright-yellow 2 (BY-2) cells, which originated from *N. tabacum*, with the cytosolic- and secretory expression cassettes to establish suspension cultures of transgenic BY-2 cell lines for production of the recombinant MaSp2 proteins in biological containments (**Figure 3**A). After *Agrobacterium*-mediated transformation and ensuing screening of transgenic BY-2 cell lines by immunodetection techniques, we isolated at least five independent transgenic BY-2 calli per construct that produced considerately high level of recombinant MaSp2 proteins in the cells (Figure 3B,C; Figure S6, Supporting Information). Consistent with transient expression studies, secretory SS-GFP-MaSp2 showed higher abundance in transgenic BY-2 calli than cytosolic GFP-MaSp2 (Figure 3B,C; Figure S6B, Supporting Information). Moreover, SS-GFP-NR6C exhibited greater expression in transgenic BY-2 calli than in SS-GFP-NR12C (Figure 3B,C; Figure S6B, Supporting Information). However, we could not detect the expected post-translationally glycosylated products of secretory SS-GFP-NR6C and SS-GFP-NR12C proteins in transgenic calli (Figure 3B,C; Figure S6B, Supporting Information). Together with our agroinfiltration studies, these results imply that targeting the SS-GFP-MaSp2 proteins to secretory systems substantially increases the accumulation of recombinant MaSp2 proteins in BY-2 cells. Lower accumulation of cytosolic GFP-MaSp2 indicates that these recombinant proteins are being stringently regulated and may undergo posttranslational degradation to preside over the protein quality in cytosol.[43,44] Subsequently, we chose an independent transgenic callus from each construct that displayed prohibitive accumulation of recombinant MaSp2 proteins to establish suspension cultures of transgenic BY-2 cells.

### 2.3. Growth of BY-2 Cells and Expression Profiles of Recombinant MaSp2 Proteins in Suspension Cultures

Liquid cultures of the selected transgenic lines were consequently generated for large-scale production of recombinant MaSp2 proteins. We prior confirmed the subcellular compartmentalization and accumulation of GFP-fused MaSp2 proteins in different BY-2 cell suspension cultures by CLSM. To determine the sorting of fluorescent proteins to plant secretory components, BY-2 cells were prior stained with glibenclamide-conjugated ER Tracker fluorescent dye that selectively binds to the sulphonylurea receptor of ATP-sensitive $K^+$ channel on ER membrane.[45,46] Fluorescence imaging demonstrated that secretory SS-GFP-MaSp2 had higher abundance in BY-2 cells in suspension cultures than cytosolic GFP-MaSp2 (Figure 3D; Figure S7, Supporting Information). In addition, SS-GFP-MaSp2 apparently co-localized with ER Tracker in ER, similar to SS-GFP protein (Figure 3D; Figure S7, Supporting Information). Our fluorescence imaging data suggests that targeting recombinant MaSp2 proteins to secretory pathways, prominently in ER, remarkably increases their abundances in BY-2 cells.

Ectopic expression of chimeric proteins that compose the exceeding hydrophobic repeats such as leucine-rich repeats in plant cells occasionally adjusts cellular mechanisms and ultimately retards growth profile of plant cells.[47–49] Certainly, the spider MaSp2 protein embeds a highly Gly/Ala-rich repetitive domain between its NTD and CTD.[6] Overexpression of this repetitive domain as well as the NTD and CTD of MaSp2 protein may alter physiological characteristics of transgenic BY-2 cells in liquid cultures. We rationally studied the growth profiles of various suspension cultures expressing cytosolic- and secretory recombinant MaSp2 proteins. Pre-cultures of transgenic BY-2 cells were initially grown in mLS media before transferring to SHM media for studying the cell growths and recombinant protein expression in liquid culture system. Fresh cell weights (FCW) of the sedimented BY-2 cells were determined every 2-days interval of a cultivation period (days 0–22) (**Figure 4**A). With the initial inoculation of 0.025 gFCW $mL^{-1}$ (0.025% inoculum), the biomass of transgenic BY-2 cells in 100 mL cultures comparably reached 0.19 ± 0.07 gFCW $mL^{-1}$ within 16 days post inoculation (Figure 4A). Besides, there was no significant difference of specific growth rates among the transgenic BY-2 cell suspension cultures (Figure 4B). This result indicates that accumulation of recombinant MaSp2 proteins either in cytosol or ER does not change growth profiles of transgenic BY-2 cells in liquid cultures.

Overexpression of extracellular recombinant proteins guided to plant secretory pathway by $S_{tab}$-$(SP)_{32}$ designer peptide enabled exceptional yields over 50–250 μg $mL^{-1}$ of BY-2 cell cultures, which were excessively higher than the commercialization aspect.[27,28] The secretion of recombinant MaSp2 protein was primarily studied by measuring fluorescence and immunoblotting of GFP-fused MaSp2 proteins in both plant cell cultures and extracellular media without plant cells. Our fluorescence measurement showed that the secretory SS-GFP line exceedingly secreted 104 ± 34 μg protein $mL^{-1}$ of recombinant proteins (≈79% secretion), enabling apparent observation of fluorescent protein under LED-light exposure at 22 days after inoculation (DAI) (Figure 4C–E; Figure S8, Supporting Information). In contrast, transgenic SS-GFP-NR6C and -NR12C lines obsequiously expressed 1.3 ± 0.3 and 0.6 ± 0.1 μg protein $mL^{-1}$, respectively, and secreted only 0.2 ± 0.1 μg $mL^{-1}$ of secretory SS-GFP-MaSp2 proteins to media (Figure 4C–E; Figure S8,

statistically significant difference in means. E) Affinity-based isolation of recombinant MaSp2 proteins via Strep-Tac II/TEV-mediated purification strategy. An example of SS-GFP-NR6C was shown in this figure. F) Purification of recombinant MaSp2 proteins from plant lysates. The cell lysate extracted from Agroinfiltrated-tobacco leaves was passed through the Strep–Tactin spin column. The recombinant MaSp2 proteins were subsequently eluted from the resin by elution buffer containing biotin. The accumulation of GFP-fused NR6C proteins was subsequently analyzed by immunoblot analysis against the anti-GFP antibody. G) On-column cleavage of the purification tag from recombinant MaSp2 NR6C proteins. The NR6C proteins were cleaved from chimeric MaSp2 NR6C proteins bound to the Strep–Tactin resin by TEV-protease. The successful cleavage of recombinant proteins by TEV-protease was confirmed by Western blotting of the eluants containing SS-GFP-Strep-TEV or GFP-Strep-TEV fragments retrieved from Strep–Tactin resins against the anti-GFP antibody. H) The protein fractions from TEV-protease-digested purified recombinant protein were stained with Coomassie Brilliant Blue (CBB) solution. 6×His-NR6C represents purified protein derived from *E. coli* expression system.

 

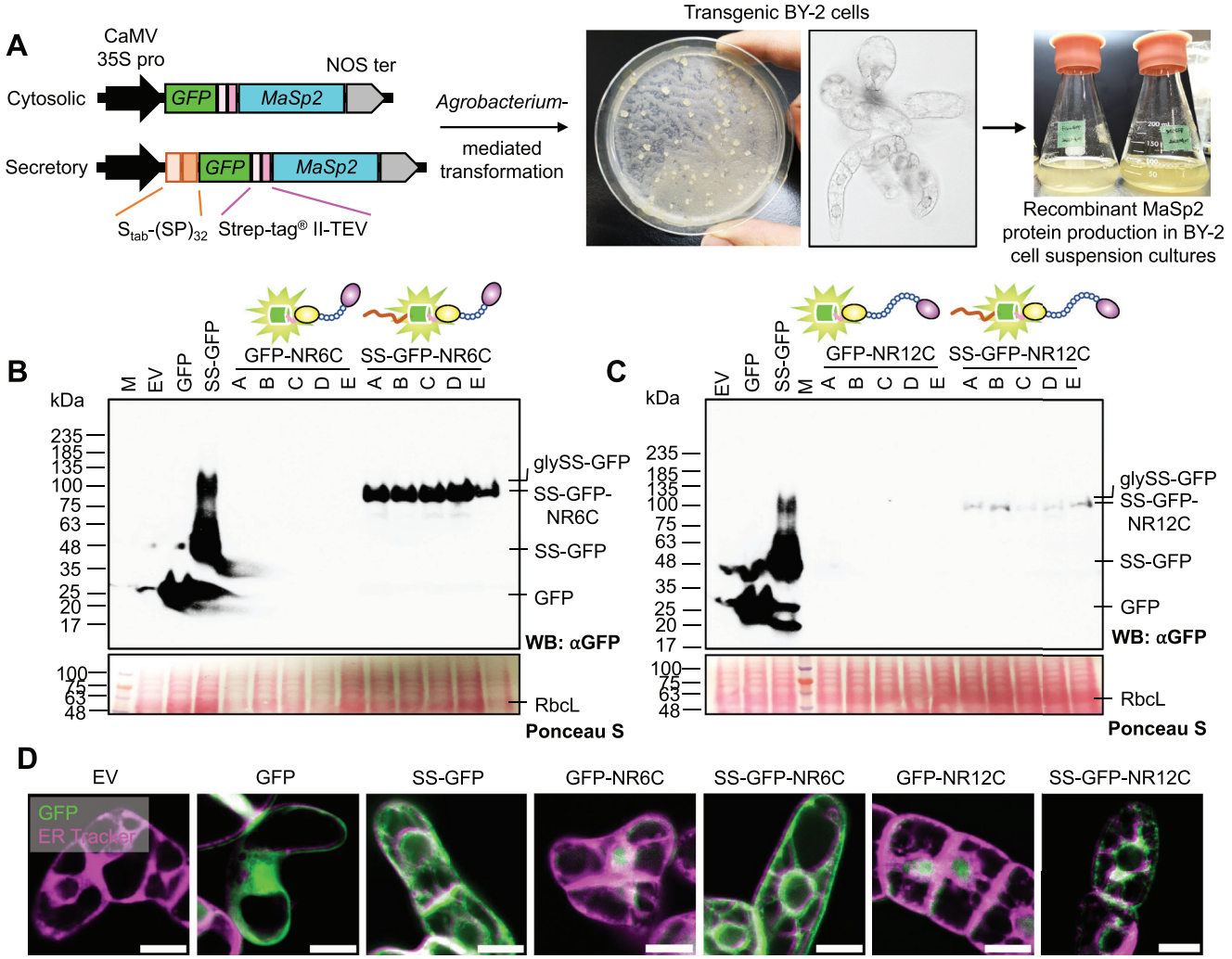


**Figure 3.** Expression of recombinant MaSp2 proteins in stable transgenic BY-2 cell lines. A) Transgenic BY-2 cell lines were generated for exogenous production of recombinant MaSp2 proteins in BY-2 cells. Expression constructs were introduced into the cells through *Agrobacterium*-mediated transformation to enable cytosolic and secretory expression of chimeric proteins. Putative transgenic calli were screened for antibiotic resistance and confirmed through immunoblot analysis. B,C) Expression of recombinant MaSp2 proteins in independent transgenic BY-2 cell lines. The accumulation of GFP-fused MaSp2 proteins in the cytosol (GFP-NR6C and GFP-NR12C) and secretory components (SS-GFP-NR6C and SS-GFP-NR12C) of the transgenic plant cell lines was analyzed through immunoblotting against anti-GFP polyclonal antibody (αGFP). Prior to immunoblotting, the membranes were stained with Ponceau-S solution to determine protein loading. The glySS-GFP band indicates the post-translational glycosylated SS-GFP protein on the membrane. RbcL shows protein bands corresponding to the large subunits of RubisCo in total callus proteins. D) Localization of GFP-fused MaSp2 proteins in plant cells. Transgenic BY-2 cells in suspension culture from the selected transgenic line were labeled with ER Tracker fluorescent dye at 3 days after inoculation (DAI) and observed under CLSM. The scale bars are 20 μm.

Supporting Information). However, these levels were higher than the unscrupulous cytosolic GFP-MaSp2 proteins that were infrequently discharged from the deteriorated cells (Figure 4C,D; Figure S8, Supporting Information). The fluorescence measurement indicates that periodic production of cytosolic- and secretory GFP-MaSp2 proteins in cell suspension cultures is growth-dependent (Figure 4A,C,D; Figure S8, Supporting Information). Quantitative immunoblotting results showed that secretory SS-GFP-MaSp2 proteins were favorably accumulated in the cultured BY-2 cells and difficultly released to culture media (Figure 4E–H; Figure S9, Supporting Information). However, we could not detect substantial profusion of cytosolic GFP-MaSp2 proteins in suspension cultures of transgenic BY-2 cells (Figure 4G,H; Figure S9, Supporting Information). Similar to the differential expressions of chimeric MaSp2 proteins in transient expression studies in tobacco leaves, we observed lower accumulation of SS-GFP-NR12C compared to SS-GFP-NR6C proteins in transgenic BY-2 cell cultures (Figures 2C,D and 4F–H). These results strongly suggest that longer repetitive protein sequences are more difficult for plant cells to synthesize. Moreover, we did not observe the expected glycosylation pattern of secretory SS-GFP-NR6C and -NR12C proteins in culture media, which was curiously distinctive to that of glycosylated SS-GFP proteins (Figure 4F; Figure S9A,B, Supporting Information).

In general, heterologous production of recombinant proteins in plant cell suspension system is consecutively associated to the

 

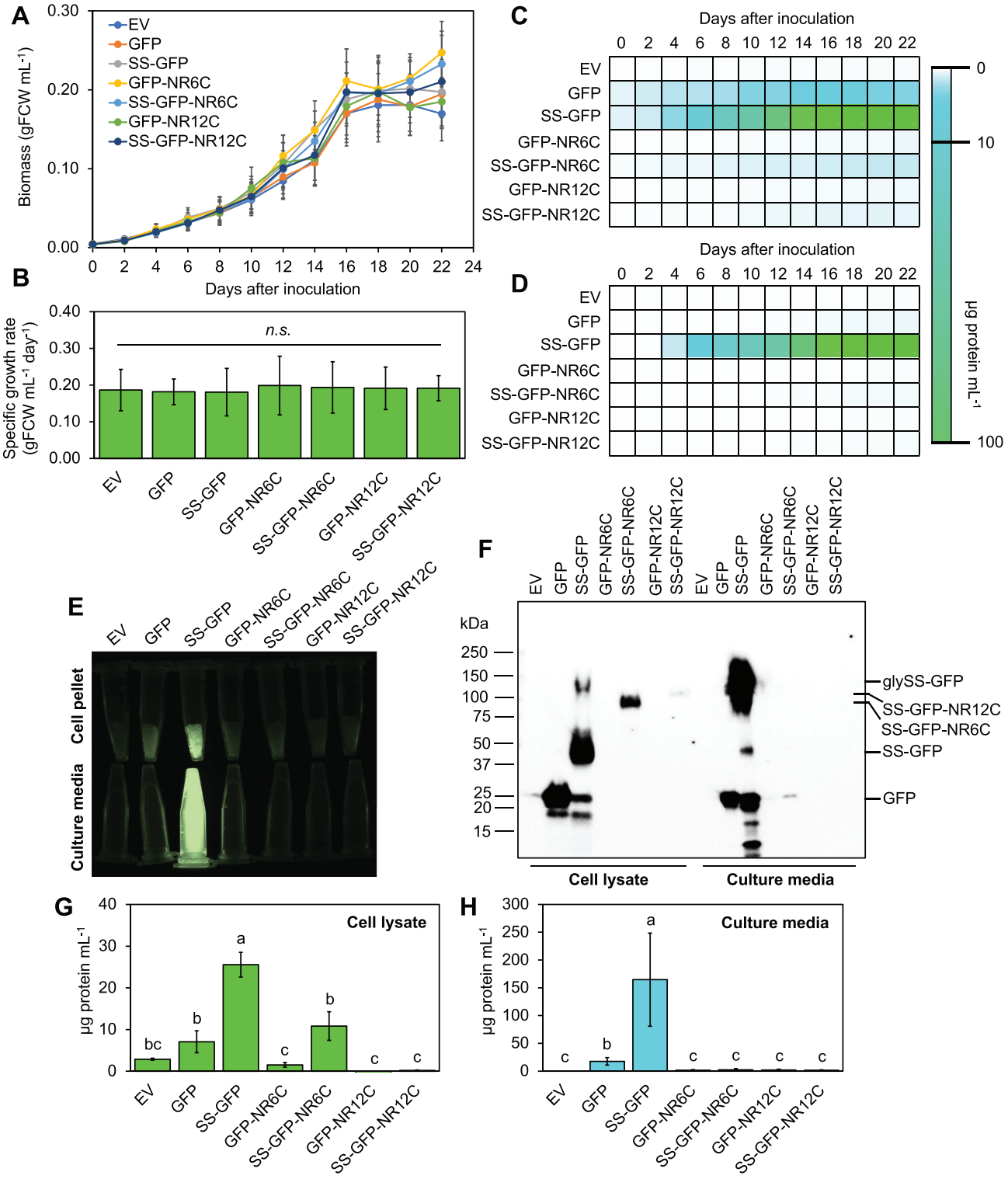


**Figure 4.** Growth and expression profiles of transgenic BY-2 cells in cell suspension systems. A) Production of biomass in shake flask cultivation system of transgenic BY-2 cells. Fresh cell weight (FCW) was measured every 2-day interval for 22 days post-inoculation. Error bars represent the standard deviation of means from four independent experiments ($N = 4$). B) Specific growth rates of transgenic plant cells in suspension cultures were determined and error bars represent the standard deviation of means from four independent experiments ($N = 4$). n.s. indicates no statistical difference of means among transgenic lines analyzed by one-way ANOVA with Tukey's HSD method at $p = 0.05$. C) Accumulation of GFP-fused MaSp2 proteins in transgenic BY-2 cell cultures. D) Cumulative abundance of recombinant fluorescent proteins in extracellular media collected from liquid cultures of individual transgenic cell lines. Production yields of GFP-fused MaSp2 proteins in suspension cultures (C) and culture media (D) over a period of

 

growth profiles of transgenic cell lines in liquid media.[23,50,51] Our results showed that production of recombinant MaSp2 proteins in BY-2 cells followed by protein secretion directed by the designer glycosylated secretory peptide; $S_{tab}$-$(SP)_{32}$, started earlier in the late of lag phase and steadily maximized at days 16–22 of cultivation. Practical adjustments of initial inoculum size, medium compositions, and culture conditions would further boost the biomass productivity of transgenic BY-2 cells and enable prosperous increase of recombinant MaSp2 production yield in suspension cultures.[50,51] Nevertheless, one of the imperative features to complete secretion of recombinant proteins mediated by $S_{tab}$-$(SP)_{32}$ signal peptide is the posttranslational O-linked glycosylation of $(SP)_{32}$ sequence with arabinogalactans.[26–28,42] Our immunoblot result suggests that the secretory SS-GFP-MaSp2 limitedly accomplished posttranslational glycosylation and was insufficiently discharged from BY-2 cells. The minimum excretion of SS-GFP-MaSp2 indicates that there are other mechanisms governing the release of SS-GFP-MaSp2 from BY-2 cells to extracellular media, which are seemingly constricted to the multifarious functional domains of MaSp2.

### 2.4. C-Terminal Domain Controls Posttranslational Appearance of Recombinant MaSp2 Proteins in BY-2 Cells

Recombinant proteins fused to secretory peptide; $S_{tab}$-$(SP)_{32}$ are ordinarily glycosylated at hydroxyprolines in $(SP)_{32}$ repeats, resulting in increased secretion and solubility of extracellular proteins.[26–28] However, a failure in adding proper polysaccharides to other possible glycosylation sites on secretory SS-GFP-MaSp2 may cause a compression of protein sorting mediated by ER and secretory vesicles.[29,52,53] In addition, our expression studies showed that cytosolic GFP-fused MaSp2 NR6C and NR12C proteins were inadequately accrued in BY-2 cells. According to these data, we hypothesized that there are other progressive molecular mechanisms directing the efficacious accumulation and secretion of recombinant MaSp2 proteins in BY-2 cells.

Glycoproteins have been previously revealed for silk spidroins.[54,55] Spider dragline silk mostly contains glycocoat layer composing glycoproteins with molecular weights above 200 kDa similar to most characterized spidroins.[20] This finding suggests that the dragline silk proteins have potentially been posttranslational-modified by the addition of polysaccharide moieties and others before releasing from silk gland cells to secretory duct. We primarily used computational analysis to predict any conceivable posttranslational glycosylation sites in recombinant SS-GFP-MaSp2 proteins rather than the distinguished hydroxyprolines of $S_{tab}$-$(SP)_{32}$ secretory peptide. Our computational prediction of O-linked glycosylation on SS-GFP-MaSp2 revealed the existences of orderly glycosylation sites in $(SP)_{32}$ repeats and several in R6 and R12 sequences, which were mostly serine residues (**Figure 5**A). However, there were also unwarranted predictable serines that frequently have prolines as the adjacent residues (S–P or P–S) in the MaSp2 CTDs (Figure 5A). Posttranslational modifications of serines generally ensue by distinct mechanisms including phosphorylation, dihydroxylation, and methylation that create a variety of structural individualities, biological appearances, and functionalities of MaSp1 and MaSp2 proteins.[21,32,56] Moreover, hydroxylation of prolines in GPGGX motifs in the repetitive regions of *T. clavipes* flagelliform deserves structural configuration and stability of this spidroin.[32,57,58] O-linked glycosylation of serines and hydroxyprolines in the repetitive region and CTD may influence the quality and secretion of recombinant MaSp2 proteins in BY-2 cells.

To investigate whether the C-terminal domains (CTDs) play a role in the cytosolic accumulation and ER-mediated extracellular sorting of MaSp2 in plant cells, we expressed GFP-fused, truncated MaSp2 constructs with different lengths of repetitive domain but lacking the CTD (NR2-, NR4-, and NR6ΔC) as well as GFP-CTD constructs in transgenic BY-2 cells for experimental verification (Figure 5B,C; Figure S10A,B, Supporting Information). However, we omitted the longer MaSp2 construct containing 12 copies of the repetitive domain (NR12ΔC) due to its lower expression level in BY-2 cells (Figure 4F). CLSM imaging confirmed subcellular localization of the respected proteins in cytosol and ER of BY-2 cells in suspension cultures (Figure 5D,E; Figure S10C,D, Supporting Information). Besides the co-localization of GFP-fused proteins with the ER arrays, we also observed the formation of fluorescent protein granules in ER Tracker-labeled ER and secretory vesicles in BY-2 cells overexpressing SS-GFP-NRΔ*C* and SS-GFP-CTD (Figure 5D,E; Figure S10C,D, Supporting Information). Compatibly, the formation of protein granules (i.e., protein bodies) in the ER and other secretion organelles is a prevalent feature when recombinant proteins are highly accumulated in specific plant compartments.[29,52,53]

We further examined the production yield of recombinant proteins in suspension cultures and in culture media of transgenic cell lines at 7 DAI. Quantitative analysis of GFP-fused recombinant proteins in cultures revealed that all cytosolic GFP-NRΔ*C* had relevant abundance in BY-2 cells ($0.64 \pm 0.12$ μg protein $mL^{-1}$ in average) (Figure 5F). Surprisingly, cytosolic GFP-CTD had statistically lower accumulation levels ($0.31 \pm 0.07$ μg protein $mL^{-1}$) in BY-2 cells when compared to GFP-NRΔ*C* proteins (Figure 5F). Secretory SS-GFP-NRΔ*C* and SS-GFP-CTD proteins showed significantly higher abundance ($16.4 \pm 9.3$ μg protein $mL^{-1}$ in average) in 7-day-old BY-2 cell cultures than their cytosolic counterparts (Figure 5F). Remarkably, 69–80% of recombinant

cultivation time were presented as heatmap expression diagrams, where the color bar represents the measurable concentration of recombinant proteins in solution. E) Fluorescent proteins in cell pellets and culture media examined under LED-light exposure. Cell pellet and extracellular media were collected from suspension cultures of transgenic BY-2 cells at 22 DAI. F) Immunoblot analysis of GFP-fused MaSp2 proteins in suspension cultures (cell lysate) and culture media of transgenic cell lines was conducted. Recombinant GFP-fused MaSp2 proteins in samples were blotted onto PVDF membrane and probed against anti-GFP antibody. EV (empty vector) designates a sample from transgenic plant cells transformed by the pBI121 vector. G,H) Quantitative immunoblot analysis was conducted to determine the amount of recombinant proteins in cell lysates (G) and culture media (H) collected from liquid cultures of transgenic BY-2 cell lines at 16 DAI. The accumulation of GFP-fused recombinant proteins was determined in three biologically-independent samples ($N = 3$) using Western blotting, as shown in Figure S9, Supporting Information. The error bars indicate the standard deviation of the means. Statistical significance differences among the samples were analyzed using one-way ANOVA with Tukey's HSD test at $p = 0.001$.

 

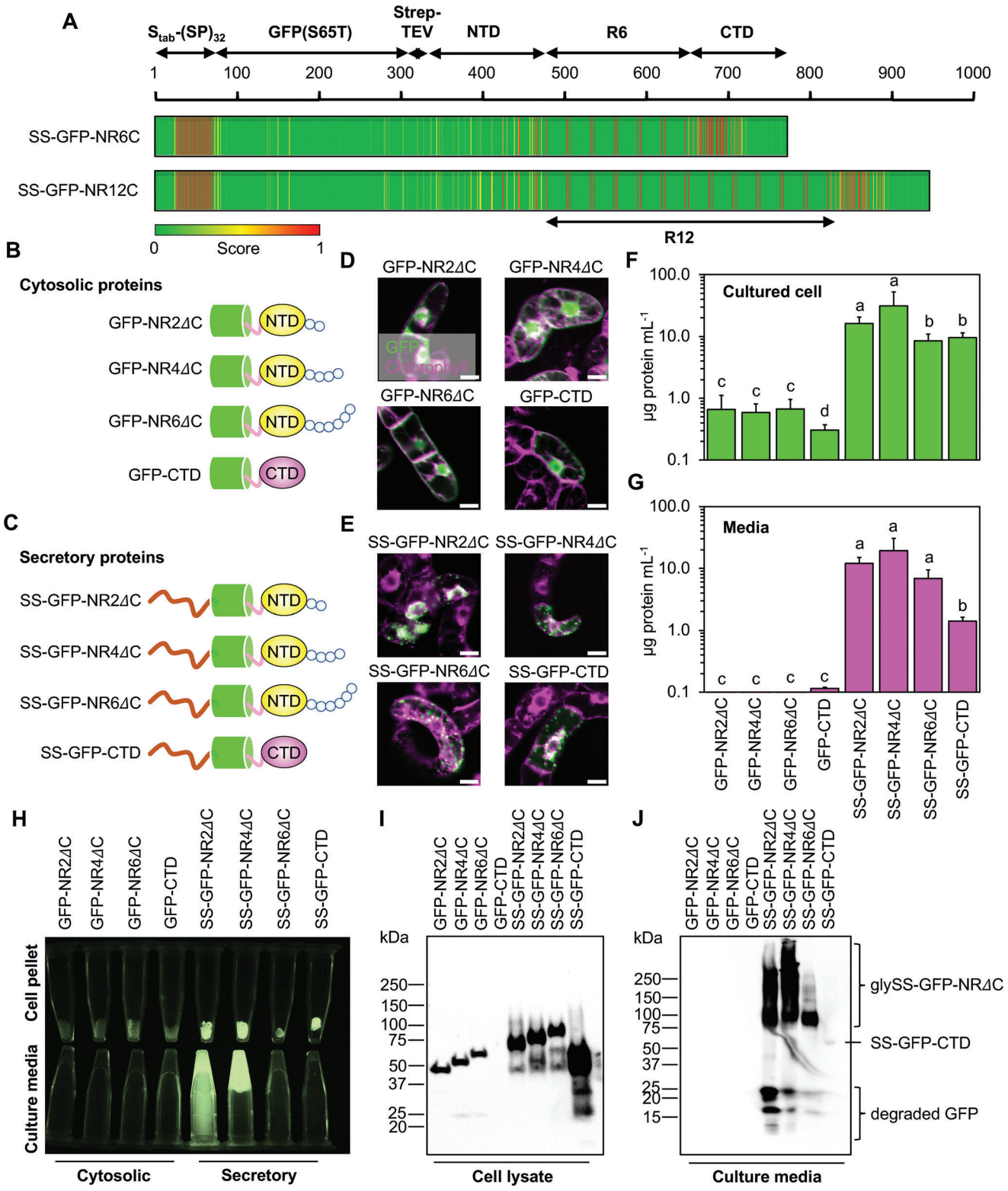


**Figure 5.** CTD controls protein quality and secretion of MaSp2 in BY-2 cells. A) Computational prediction of possible O-linked glycosylation sites in SS-GFP-MaSp2 proteins. The confident scores of each predicted glycosylation sites were presented as colormap expression diagram. Color bar represents the predicted scores range from 0.0 to 1.0. The ruler shows the length of amino acids of each protein. B) Cytosolic forms of recombinant truncated MaSp2 proteins. C) Secretory proteins of deleted recombinant MaSp2. D) Cytosolic localization of various GFP-fused MaSp2 NRΔC and GFP-CTD proteins. E) ER-targeting and granule formation of secretory SS-GFP-NRΔC and SS-GFP-CTD proteins in transgenic BY-2 cells. BY-2 cells in suspension cultures were prior stained by ER Tracker, and subcellular localization of GFP-fused recombinant proteins in BY-2 cells was observed by CLSM at 3 days after

 

SS-GFP-NRΔ*C* proteins (12.7 ± 4.9 μg protein $mL^{-1}$) was successfully secreted to media and could be merely observed under LED-light exposure, especially SS-GFP-NR2- and SS-GFP-NR4Δ*C* proteins (Figure 5F–H). Moreover, although it was incredibly amassed in ER (9.5 ± 1.9 μg protein $mL^{-1}$), only 15% of SS-GFP-CTD proteins (1.4 ± 0.2 μg protein $mL^{-1}$) was detected in culture media (Figure 5F,G). In addition, secretory expression of recombinant SS-GFP-NR6Δ*C* proteins exhibited superior levels of accumulation than the SS-GFP-NR6C proteins in transgenic BY-2 cells (Figures 4C,D and 5F,G).

Although perceptive GFP-fused recombinant proteins could be quantitatively determined by the fluorescence-based techniques, immunoblot analysis showed relevant expression and posttranslationally-modified patterns of different proteins in BY-2 cells and culture media (Figure 5I,J). Our immunoblot result revealed higher accumulation of SS-GFP-fused secretory proteins than their cytosolic forms in BY-2 cell lysates (Figure 5I). While targeting SS-GFP-CTD protein to secretory compartments incredibly amassed the accretion of recombinant protein, its cytosolic counterpart (GFP-CTD) could not be sensitively detected in the cell lysates (Figure 5I). Successful secretion of $S_{tab}$-$(SP)_{32}$-fused recombinant proteins from BY-2 cells to extracellular media could be examined by the mobility shift of glycosylated proteins in immunoblotting.[26–29,51] We observed pertinent band-shifts of all recombinant SS-GFP-NRΔ*C* proteins in culture media collected from suspension cultures of 7-day-old BY-2 cells overexpressing secretory proteins, indicating that these secreted SS-GFP-NRΔ*C* proteins were effectively glycosylated (Figure 5J). On the contrary, we could not detect substantial glycosylation pattern of SS-GFP-CTD proteins in culture media even though this recombinant protein was greatly accumulated in BY-2 cells (Figure 5I,J). Further investigation of the O-linked glycosylation of the chimeric secretory proteins, using an antibody that specifically recognizes arabinogalactans, suggested that the recombinant SS-GFP-NRΔ*C* proteins in the culture media of transgenic BY-2 cells were appropriately glycosylated with the targeted glycosides (Figure S11A,B, Supporting Information). In contrast, the recombinant SS-GFP-CTD did not acquire sufficient glycosylation and conspicuously remained inside the BY-2 cells (Figure S11B, Supporting Information). However, there was no detectable level of the cytosolic GFP-fused recombinant proteins in culture media (Figure 5J). Together with fluorescence detections, our results indicate that CTD is responsible for quality control of MaSp2 in cytosol and regulates the secretion of recombinant proteins through ER-mediated protein sorting pathways.

Nevertheless, protein sorting and secretory mechanism of spider silk spidroins are still understudied. Bioinformatic analysis suggested that MaSp and other spidroins are extracellular proteins secreted from cells in spider silk glands with a mechanism organized by consensus signal peptides located at the NTD of each spidroins.[59] This mechanism facilitated by signal peptides of spider spidroins is presumably similar to the secretion of silkworm (*Bombyx mori*) fibroin in posterior silk gland cells.[60] Interestingly, extracellular organization of fibroin from the posterior silk gland cells requires ER/Golgi-mediated secretory pathway in association with the microtubule-microfilament polarization.[61] Moreover, protein truncation and expression analysis in transgenic silkworms show that CTD of H-chain fibroin is a key domain controlling its secretion from posterior silk gland cells.[62] Importantly, effectual glycosylation on the CTD of H-chain fibroin is necessary for protein secretion, maintenance of H–L heterodimer, and formation of the elementary unit of silkworm fibroin.[62,63] Our extended mobility shift assay showed that cytosolic GFP-CTD and secretory SS-GFP-CTD could form homodimers under non-denatured conditions (Figure S12B,D, Supporting Information). However, these proteins were partially degraded in the native conditions of BY-2 cell extracts (Figure S12B,D, Supporting Information). These findings suggest that the post-translational degradation-prone CTD plays a role in facilitating the self-assembly and fiber formation of spidroin silk by mediating the formation of MaSp2 dimers through intermolecular interactions. Based on our current experimental data, which consider the functional similarity between silkworm fibroin and spider silk spidroin, we propose that the CTD of MaSp2 controls protein quality and secretion in spider silk gland cells.

## 3. Conclusion

Exogenous expression of a whole MaSp2 protein in tobacco cells is eminently challenging due to rapid degradation of protein as examined in our research. Interestingly, directing recombinant proteins to plant secretory mechanisms significantly increases the production yields of MaSp2 in cell suspension cultures. Nevertheless, our plant cell-based expression studies disclose that the CTD critically controls the abundance of MaSp2 in cytosol and its discharge from BY-2 cells to extracellular media. Last, removing CTD from plant-derived MaSp2 proteins tremendously improves productivity of the secreted recombinant proteins in the culture media of transgenic BY-2 cells. Our data informs a potential postulation to unveil regulatory function of MaSp CTD and future design of appropriate production platforms of recombinant spidroins. Certainly, intellectual insights from plant experiments would not completely prevail the existing mechanisms in spider cells. Further characterization of mechanism underlying the secretion of MaSp allows a comprehensive engineering strategy to manufacture scaffolds of extracellular proteins such as other spidroins, collagens, and elastin in commercially-feasible plant cell culture systems.

inoculation (DAI) to liquid cultures. Scale bars = 20 μm. F,G) Accumulation of truncated MaSp2 proteins in suspension cultures (F) and culture media without BY-2 cells (G) at 7 DAI. Error bars = standard deviation of means from six independent cultures ($N = 6$). Letters indicate statistically significant differences of mean analyzed by one-way ANOVA with Tukey's HSD test at $p = 0.05$. H) Visualization of GFP-fused truncated MaSp2 proteins in cell pellets and liquid media collected from 7 DAI suspension cultures of transgenic BY-2 cells. I,J) Immunoblot analyses of recombinant proteins in cell lysate and culture media collected from suspension cultures of different transgenic lines at 7 DAI. GFP-fused proteins on immunoblot membrane were probed with anti-GFP antibody. glySS-GFP-NRΔ*C* indicated the mobility shift of SS-GFP-NRΔ*C* proteins as a result of post-translational glycosylation. GFP-specific bands with the molecular weights smaller than 25 kDa were assumed to be fragmented GFP molecules and indicated by degraded GFP.

 

## 4. Experimental Section

*Plant Cultivation*: *N. tabacum* seeds were sowed and germinated on premixed soil (HonenAgri Co., Ltd., Niigata, Japan) at 24 °C under a 12/12 light/dark photoperiod at 100 µmol photons $m^{-2}{\cdot}s^{-1}$ in a growth chamber. Two-week-old seedling was transferred and grown in an individual pot under the same conditions. Fully-expanded leaves from 6-week-old plants were used in agroinfiltration experiments.

*N. tabacum* BY-2 cell line (rpc00001) was provided by the RIKEN BRC, participating in the National BioResource Project of the MEXT/AMED Japan. BY-2 cell line was maintained in a modified Linsmaier and Skoog (mLS) liquid media, pH 5.8 supplemented with 0.2 mg $L^{-1}$ 2,4-dichrorophenoxyacetic acid (2,4-D) and 30 g $L^{-1}$ sucrose. Suspension cultures were incubated in the dark at 26 °C with rotary shaking at 120 rpm. For subculturing, 1 mL of 7-day-old culture was transferred to 100 mL of mLS media and incubated at the same culture conditions.

*Plant and Bacteria Expression Constructs*: Full length sequences of *MaSp2* genes with 6 and 12 repetitive domains (*NR6C* and *NR12C*) were cloned from pET15b-*NR6C* and pET15b-*NR12C* with gene specific primers (Table S2, Supporting Information). *GFP(S65T)* gene was amplified from pBI121-*GFP* expression vector. Coding sequence of a signal peptide of tobacco expansin ($S_{tab}$) was cloned from cDNA sample of tobacco leaf using primers listed in Table S2, Supporting Information. DNA sequence encoding 32 repeats of serine–proline dipeptides ($(SP)_{32}$) was artificially synthesized by PCR extension with nucleotide sequences in Table S2, Supporting Information.[33] Coding sequences of Strep–Tactin purification tag and TEV-protease proteolytic cleavage site were synthesized by PCR extension with primers listed in Table S2, Supporting Information. All PCR fragments were amplified using PrimeSTAR GXL DNA polymerase kit (Takara Bio, Shiga, Japan). DNA fragments were prior cloned into the pTA2-cloning vector with a TArget Clone-Plus cloning kit (Toyobo, Osaka, Japan). Ligation fragments were digested from cloning vectors with appropriate restriction enzymes (Takara) and ligated to linearized pBI121 vector fragments with a Ligation High Ver. 2 DNA ligation kit (Toyobo). Plant expression vectors were transformed and maintained in *E. coli* DH5α cells. Polypeptide sequences and theoretical molecular weights for each domain and the resulting chimeric proteins were given in Table S1, Supporting Information.

*Escherichia coli* expression vector for purification of standard GFP protein was constructed by subcloning *Nde*I/*Xho*I-digested *GFP(S65T)* DNA fragments to linearized pET15b expression vector. The resulting construct was transformed to *E. coli* BL21(DE3) cells and a positive transformant was maintained for protein expression. Recombinant His-GFP was isolated from bacterial cells grown in LB media at 37 °C with shaking at 200 rpm. Bacterial cultures with the optical density (OD) at 600 nm of 0.5 were induced by 400 µM of IPTG at 25 °C with rotary shaking at 100 rpm for 6 h. *Escherichia coli* cells were then lysed in 8 M urea + 10 mM Tris-HCl, pH 7.5 + 150 mM NaCl + one tablet of cOmplete, EDTA-free Protease Inhibitor (Roche Diagnostics GmBH, Manheim, Germany) at ambient temperature for 6 h. Cell lysate was injected to HisTrap Fast Flow column (Merck KGaA, Darmstadt, Germany). The column was then washed with 10 mM Tris-HCl, pH 7.5 + 150 mM NaCl + 20 mM imidazole for ten times. Recombinant His-GFP were eluted from the column by eight fractions of 1 mL of 10 mM Tris-HCl + 150 mM NaCl + 200 mM imidazole. Purified proteins were analyzed by SDS-PAGE (Figure S13A, Supporting Information). Linear regression of recombinant His-GFP diluted in culture media was generated and used to determine the concentration of GFP-fused recombinant protein in plant cell cultures (Figure S13B, Supporting Information).

*Agrobacterium-Mediated Gene Transfer*: Plant expression vectors were transformed to $CaCl_2$-treated *Agrobacterium tumefaciens*-strain LBA4404 competent cells using freeze-thawing protocol.[64] A single colony of *Agrobacterium* transformant was inoculated in YEB broth supplemented with 50 mg $L^{-1}$ kanamycin and 20 µM of acetosyringone and cultured at 30 °C with rotary shaking at 200 rpm for overnight. The liquid culture was used for agroinfiltration to tobacco leaves and cocultivation with BY-2 cells.

For transient expression studies in tobacco leaves, *Agrobacterium* cells were pelleted from overnight cultures by centrifugation at 4000 rpm, 4 °C for 15 min. Cell pellet was resuspended in 10 mM MES, pH 5.7 + 10 mM $MgCl_2$ + 200 µM of acetosyringone (Agroinfiltration solution) and $OD_{600nm}$ of cultures were determined by spectrophotometry. *Agrobacterium* cultures were further diluted to $OD_{600nm}$ of 0.2 with agroinfiltration solution and incubated in the dark at ambient temperature for 2 h. After incubation, infiltration media were introduced to fully-expanded leaves of tobacco on the abaxial side (≈0.2–0.4 mL culture per leaf). The infiltrated plants were incubated at the same growth conditions for 3 days prior to the detections of recombinant protein expression by fluorescence microscopy and immunoblot analysis.

Transgenic BY-2 cell lines were generated by cocultivation of plant cells with *Agrobacterium* cells harboring plant expression vectors.[65] In brief, 4 mL of suspension culture (4-day-old) of BY-2 cells was inoculated with 40 µL of overnight culture of *Agrobacterium* cells and the suspensions were incubated in the dark at 26 °C without shaking for 2 days. After cocultivation, a co-culture was washed five times with 10 mL of mLS media by sedimentation. After washing, BY-2 cells were resuspended with 2 mL of mLS media and 0.5 mL of the culture was sprayed on the surface of mLS agar media (0.7% w/v agar) supplemented with 250 mg $L^{-1}$ cefotaxime and 50 mg $L^{-1}$ kanamycin for selection. The selection plates were incubated in the dark at standard culture conditions for 3–4 weeks for regeneration of putative transgenic BY-2 calli. Approximately 30 regenerated calli (≈0.5 cm in diameter) per plate were transferred to new selection plates and cultured at the same conditions until the callus size was ≈1 cm. The transgenic BY-2 calli were subcultured and maintained on mLS agar within a 14 days interval for screening of high protein expression lines by immunoblotting.

*Suspension Culture of Transgenic BY-2 Cells*: The initial suspension cultures of transgenic BY-2 cell lines with high protein expression were established by suspending a callus (≈1 cm in diameter) of the selected cell line in 40 mL of mLS media. The liquid cultures were incubated at standard culture conditions with rotary shaking. After 7 days of inoculation, 2 mL of cell suspension was transferred to 100 mL of mLS media. Suspension cultures of transgenic BY-2 cell lines were further maintained by sub-cultivation of 1 mL of culture to 100 mL of mLS media within 7 days interval.

0.25 gram (fresh cell weight, FCW) of 4-day-old transgenic BY-2 cells grown in mLS liquid media was transferred to 100 mL of Schenk and Hildebrandt media,[66] pH 5.8 supplemented with 34 g $L^{-1}$ sucrose and 0.2 mg $L^{-1}$ 2,4-D (SHM media) to study protein secretion.[26] This culture was incubated at standard cultivation conditions with rotary shaking at 120 rpm.

*Growth and Recombinant Protein Expression in Suspension Cultures*: Biomass production, expression, and secretion of recombinant proteins were determined in suspension cultures of transgenic BY-2 cells cultivated in SHM media at different periods of cultivation. One milliliter of suspension culture was collected to the microcentrifuge tube and settled for 30 min for cell sedimentation. Culture media was suctioned from BY-2 cells and fresh cell weight was determined.

Fluorescence intensity in suspension culture and culture media without BY-2 cell was measured by fluorescence detection mode in SpectraMax iD3 Multi-mode microplate reader (Molecular Devices, Tokyo, Japan). GFP fluorescence in 100 µL of sample (cell culture and culture media without BY-2 cell) was determined by the emission and excitation wavelengths of 488 and 530 nm, respectively. The integral time for GFP fluorescence measurement was set at 800 milliseconds. Concentration of recombinant fluorescent proteins in sample was computed from a linear regression equation of recombinant His-GFP as previously described.

*Fluorescence Microscopy*: Subcellular localization of GFP-fused recombinant protein in tobacco cells was observed under a confocal laser-scanning microscope (CLSM) (ZeissLSM880, Carl Zeiss, Oberkochen, Germany). Fluorescence images were taken under a Plan-Apochromat 20×/0.8 M27 objective lane with a digital zoom of 2. The laser power from a diode source was preset at 10. The size of the 16-bit image was 1024 × 1024 pixels. The excitation/emission (ex/em) wavelengths for detection of GFP-fused proteins were 488/510–525 nm. Chlorophyll autofluorescence in tobacco leaf cells was detected at ex/em 488/640–700 nm. The digital fluorescence gains of GFP and chlorophyll were 550 and 600, respectively. For leaf infiltration experiments, CLSM images were taken from the abaxial side of tobacco leaves at 3 days post infiltration.

 

Transgenic BY-2 cells in suspension cultures were prior stained with 10 µM of ER Tracker Blue-White DPX (Invitrogen, Paisley, UK) in the dark at room temperature for 15 min. The stained BY-2 cells were washed three times with 1 mL of mLS media. Fluorescence of ER Tracker fluorescent dye in BY-2 cell was detected at ex/em 405/440–450 nm with the digital gain of 600.

*Strep–Tactin Purification and On Column TEV-Digestion*: Recombinant MaSp2 proteins were purified from cell lysates of tobacco leaves agroinfiltrated with GFP-NR6C and SS-GFP-NR6C expression vectors using a modified protocol. Briefly, total leaf proteins were extracted from transformed tobacco leaves at 3 days post infiltration with extraction buffer (100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, pH 8.0, 1.0% v/v Triton X-100 + cOmplete, EDTA-free Protease Inhibitor Cocktail [Roche Diagnostics GmBH]). Cell lysates were passed through Strep–Tactin Spin Column (IBA GmbH, Goettingen, Germany) and recombinant MaSp2 proteins were consequently purified according to manufacturer's protocol.

For on column TEV-digestion, recombinant proteins bound to Strep–Tactin matrix in column were incubated with ten units of TEV protease (Merck KGaA) for 1 h at ambient temperature. TEV-cleaved native MaSp2 proteins were then collected from columns by centrifugation at 15 000 rpm for 5 min. Affinity tags and GFP domains of recombinant proteins bound to the Strep–Tactin matrix were subsequently eluted from columns by extraction buffer containing 50 mM biotin. Each protein fraction was analyzed by immunoblotting.

*Protein Identification*: The identification of the protein was carried out by resolving ≈5 µg of TEV-cleaved recombinant MaSp2 NR6C proteins on SDS-PAGE gels prior to staining with Coomassie Brilliant Blue (CBB) or silver staining to visualize the protein bands. Recombinant hexameric histidine-tagged 6×His-NR6C protein purified from *E. coli* cell lysates was included as a control protein in SDS-PAGE and liquid-mass spectrometry/mass spectrometry (LC-MS/MS) analysis. The target protein bands were excised and digested with ten units of trypsin (Promega). The resulting peptide fragments were analyzed by LC-MS/MS using a Q Exactive mass spectrometer (Thermo Fisher Scientific, Yokohama, Japan) at the RIKEN BMA Mass Spectrometry Service (Proteomic Analysis), Research Resources Division (RRD), Center for Brain Science (CBS). The ion mass data were searched against the SwissProt and TrEMBL databases of *N. tabacum* using the in-house Proteome Discoverer software with the MASCOT search engine. The peptide sequences matched to the ion mass spectra are available in the supporting data (Excel file).

*Immunoblotting*: Total soluble proteins were extracted from plant leaves, transgenic BY-2 calli, and pelleted cells from suspension cultures by 6 M urea cracking solution (6 M urea + 100 mM Tris-HCl, pH 7.0, 20% v/v glycerol, 10% w/v SDS, 5% v/v $\beta$-mercaptoethanol + cOmplete, EDTA-free Protease Inhibitor Cocktail [Roche Diagnostics GmBH]). Culture media was collected from suspension cultures of transgenic BY-2 cells after sedimentation. For dot-blot analysis, two micrograms of total plant proteins was spotted on a 0.2 µm PVDF membrane pre-wet with 1× Tris-glycine buffer, pH 8.0 + 10% v/v methanol and dried in ambient conditions. Ten micrograms of total soluble proteins or ten microliters of culture media was resolved in a 4–15% Mini PROTEAN TGX Precast gel and blotted onto a PVDF membrane using a Tran-Blot SD Semi-Dry Transfer Cell (Bio-Rad Laboratories, Tokyo, Japan). GFP-fused recombinant proteins on the membrane were probed with rabbit anti-GFP polyclonal antibody (NB600-308; Novus Biologicals, Littleton, CO, USA) as a primary antibody at a dilution of 1:5000. The glycosylated recombinant proteins were probed with the mouse anti-Arabinogalactan-2 monoclonal antibody (CCRC-M133; Agrisera, Vännäs, Sweden) as the primary antibody at a dilution of 1:2000. A horseradish peroxidase (HRP)-conjugated polyclonal antibody, either goat anti-rabbit IgG (ab6721; Abcam, Tokyo, Japan) or goat anti-mouse IgG (ab6789; Abcam), was used as the secondary antibody at a dilution of 1:20 000. After antibody reaction, 1 mL of SuperSignal West Pico PLUS chemiluminescence substrate mixture (Life Technologies, Carlsbad, CA, USA) was applied to the membrane prior to detection of the HRP signal using a LuminoGraph I imaging system (ATTO Corporation, Tokyo, Japan). The chemiluminescent intensities of protein bands were quantified using Fiji ImageJ.[67]

*Computational Prediction*: NetOGlyc-4.0 server was used in prediction of possible O-linked glycosylation sites in SS-GFP-MaSp2 proteins.[68]

*Statistical Analysis*: Experimental data are usually presented as bar graphs, which include error bars representing the mean ± standard deviation (SD) collected from three to six biologically independent experiments ($N$). To analyze statistically significant differences among treatments, one-way ANOVA was performed with Tukey's honestly significant difference HSD test at a significance level of $p = 0.05$. In addition, paired samples $t$-tests were conducted with various probability levels ($P$-values) to compare the statistical significance of differences between two samples, using Jamovi version 1.6 (The Jamovi Project, Sydney, Australia).

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

The authors thank Dr. Ali D. Malay for providing pET15b-*NR6C* and pET15b-*NR12C* vectors. This research was financially supported by the Japan Science and Technology Agency Exploratory Research for Advanced Technology (JST-ERATO; K.N., JPMJER1602). The authors also thank the Support Unit for Bio-Material Analysis, RIKEN CBS Research Resources Division for providing technical support.

## Conflict of Interest

The authors declare no conflict of interest.

## Data Availability Statement

The data that support the findings of this study are available in the Supporting Information of this article.

## Keywords

BY-2 cell, plant expression system, recombinant protein, secretory protein, spider silk

Received: January 10, 2023
Revised: May 15, 2023
Published online: July 6, 2023

[1] A. Rising, J. Johansson, *Nat. Chem. Biol.* **2015**, *11*, 309.
[2] M. Heim, D. Keerl, T. Scheibel, *Angew. Chem., Int. Ed.* **2009**, *48*, 3584.
[3] F. Vollrath, D. P. Knight, *Nature* **2001**, *410*, 541.
[4] T. Scheibel, *Microb. Cell Fact.* **2004**, *3*, 14.
[5] M. Ramezaniaghdam, N. D. Nahdi, R. Reski, *Front. Bioeng. Biotechnol.* **2022**, *10*, 835637.
[6] A. D. Malay, T. Suzuki, T. Katashima, N. Kono, K. Arakawa, K. Numata, *Sci. Adv.* **2020**, *6*, eabb6030.
[7] D. R. Whittall, K. V. Baker, R. Breitling, E. Takano, *Trends Biotechnol.* **2021**, *39*, 560.
[8] C. H. Bowen, B. Dai, C. J. Sargent, W. Bai, P. Ladiwala, H. Feng, W. Huang, D. L. Kaplan, J. M. Galazka, F. Zhang, *Biomacromolecules* **2018**, *19*, 3853.
[9] X. X. Xia, Z. G. Qian, C. S. Ki, Y. H. Park, D. L. Kaplan, S. Y. Lee, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 14059.

 

[10] K. V. Sidoruk, L. I. Davydova, D. G. Kozlov, D. G. Gubaidullin, A. V. Glazunov, V. G. Bogush, V. G. Debabov, *Appl. Biochem. Microbiol.* **2015**, *51*, 766.
[11] S. R. Fahnestock, L. A. Bedzyk, *Appl. Microbiol. Biotechnol.* **1997**, *47*, 33.
[12] J. Scheller, K.-H. Gührs, F. Grosse, U. Conrad, *Nat. Biotechnol.* **2001**, *19*, 573.
[13] J. Yang, L. A. Barr, S. R. Fahnestock, Z. B. Liu, *Transgenic Res.* **2005**, *14*, 313.
[14] A. Lazaris, S. Arcidiacono, Y. Huang, J. F. Zhou, F. Duguay, N. Chretien, E. A. Welsh, J. W. Soares, C. N. Karatzas, *Science* **2002**, *295*, 472.
[15] S. Grip, A. Rising, H. Nimmervoll, S. Erica, S. J. McQueen-Mason, N. Pouchkina-Stantcheva, F. Vollrath, W. Engstrom, A. Fernandez-Arias, *Cancer Genomics Proteomics* **2006**, *3*, 83.
[16] H. T. Xu, B. L. Fan, S. Y. Yu, Y. H. Huang, Z. H. Zhao, Z. X. Lian, Y. P. Dai, L. L. Wang, Z. L. Liu, J. Fei, N. Li, *Anim. Biotechnol.* **2007**, *18*, 1.
[17] F. Teulé, Y. G. Miao, B. H. Sohn, Y. S. Kim, J. J. Hull, M. J. Fraser, R. V. Lewis, D. L. Jarvis, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 923.
[18] C. G. Copeland, B. E. Bell, C. D. Christensen, R. V. Lewis, *ACS Biomater. Sci. Eng.* **2015**, *1*, 577.
[19] H. Wen, X. Lan, Y. Zhang, T. Zhao, Y. Wang, Z. Kajiura, M. Nakagaki, *Mol. Biol. Rep.* **2010**, *37*, 1815.
[20] A. Sponner, W. Vater, S. Monajembashi, E. Unger, F. Grosse, K. Weisshart, *PLoS One* **2007**, *2*, e998.
[21] J. R. A. dos Santos-Pinto, G. Lamprecht, W.-Q. Chen, S. Heo, J. G. Hardy, H. Priewalder, T. R. Scheibel, M. S. Palma, G. Lubec, *J. Proteomics* **2014**, *105*, 174.
[22] J. K. C. Ma, P. M. W. Drake, P. Christou, *Nat. Rev. Genet.* **2003**, *4*, 794.
[23] M. J. B. Burnett, A. C. Burnett, *Plants, People, Planet* **2020**, *2*, 121.
[24] M. Benchabane, C. Goulet, D. Rivard, L. Faye, V. Gomord, D. Michaud, *Plant Biotechnol. J.* **2008**, *6*, 633.
[25] T. Wu, S. M. Kerbler, A. R. Fernie, Y. Zhang, *Plant Commun.* **2021**, *2*, 100235.
[26] N. Zhang, M. Gonzalez, B. Savary, J. Xu, *Biotechnol. J.* **2016**, *11*, 497.
[27] J. Xu, S. Okada, L. Tan, K. J. Goodrum, J. J. Kopchick, M. J. Kieliszewski, *Transgenic Res.* **2010**, *19*, 849.
[28] J. Xu, L. Tan, K. J. Goodrum, M. J. Kieliszewski, *Biotechnol. Bioeng.* **2007**, *97*, 997.
[29] N. Zhang, T. Wright, X. Wang, U. Karki, B. J. Savary, J. Xu, *Plant Biotechnol. J.* **2019**, *17*, 1130.
[30] V. Hauptmann, N. Weichert, M. Menzel, D. Knoch, N. Paege, J. Scheller, U. Spohn, U. Conrad, M. Gils, *Transgenic Res.* **2013**, *22*, 369.
[31] N. Weichert, V. Hauptmann, C. Helmold, U. Conrad, *Front. Plant Sci.* **2016**, *7*, 6.
[32] A. D. Malay, H. C. Craig, J. Chen, N. A. Oktaviani, K. Numata, *Biomacromolecules* **2022**, *23*, 1827.
[33] E. Shpak, J. F. Leykam, M. J. Kieliszewski, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 14736.
[34] R. Menassa, H. Zhu, C. N. Karatzas, A. Lazaris, A. Richman, J. Brandle, *Plant Biotechnol. J.* **2004**, *2*, 431.
[35] P. Shah, Y. Ding, M. Niemczyk, G. Kudla, J. B. Plotkin, *Cell* **2013**, *153*, 1589.
[36] G. Marjanac, M. Karimi, M. Naudts, T. Beeckman, A. Depicker, S. de Buck, *New Phytol.* **2009**, *184*, 851.
[37] N. Weichert, V. Hauptmann, M. Menzel, K. Schallau, P. Gunkel, T. C. Hertel, M. Pietzsch, U. Spohn, U. Conrad, *Plant Biotechnol. J.* **2014**, *12*, 265.
[38] M. Niehaus, H. Straube, P. Künzler, N. Rugen, J. Hegermann, P. Giavalisco, H. Eubel, C. P. Witte, M. Herde, *Plant Physiol.* **2020**, *182*, 1194.
[39] H. T. Phan, U. Gresch, U. Conrad, *Front. Bioeng. Biotechnol.* **2018**, *6*, 115.
[40] T. G. M. Schmidt, A. Skerra, *Nat. Protoc.* **2007**, *2*, 1528.
[41] S. Hellwig, J. Drossard, R. M. Twyman, R. Fischer, *Nat. Biotechnol.* **2004**, *22*, 1415.
[42] D. E. Webster, M. C. Thomas, *Biotechnol. Adv.* **2012**, *30*, 410.
[43] Z. Sun, J. L. Brodsky, *J. Cell Biol.* **2019**, *218*, 3171.
[44] S. P. Ferris, V. K. Kodali, R. J. Kaufman, *Dis. Model. Mech.* **2014**, *7*, 331.
[45] M. Ochsner, *J. Photochem. Photobiol., B* **1997**, *39*, 1.
[46] A. Hambrock, C. Löffler-Walz, U. Quast, *Br. J. Pharmacol.* **2002**, *136*, 995.
[47] C. Zhao, O. Zayed, Z. Yu, W. Jiang, P. Zhu, C. C. Hsu, L. Zhang, W. A. Tao, R. Lozano-Durán, J. K. Zhu, *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 13123.
[48] Y. H. Park, C. Choi, E. M. Park, H. S. Kim, H. J. Park, S. C. Bae, I. Ahn, M. G. Kim, S. R. Park, D. J. Hwang, *Plant Cell Rep.* **2012**, *31*, 1845.
[49] B. Y. Cheon, H. J. Kim, K. H. Oh, S. C. Bahn, J. H. Ahn, J. W. Choi, S. H. Ok, J. M. Bae, J. S. Shin, *Transgenic Res.* **2004**, *13*, 541.
[50] S. T. Häkkinen, L. Reuter, N. Nuorti, J. J. Joensuu, H. Rischer, A. Ritala, *Front. Plant Sci.* **2018**, *9*, 45.
[51] U. Karki, H. Fang, W. Guo, C. Unnold-Cofre, J. Xu, *Plant Cell Rep.* **2021**, *40*, 1087.
[52] H. Kang, Y. Park, Y. Lee, Y. J. Yoo, I. Hwang, *Sci. Rep.* **2018**, *8*, 4612.
[53] R. Saberianfar, J. J. Joensuu, A. J. Conley, R. Menassa, *Plant Biotechnol. J.* **2015**, *13*, 927.
[54] K. Augsten, P. Mühlig, C. Herrmann, *Scanning* **2000**, *22*, 12.
[55] A. Weiskopf, K. Senecal, P. Vouros, D. Kaplan, C. M. Mello, *Glycobiology* **1996**, *6*, 1703.
[56] J. R. A. dos Santos-Pinto, H. A. Arcuri, G. Lubec, M. S. Palma, *Biochim. Biophys. Acta, Proteins Proteomics* **2016**, *1864*, 1444.
[57] W. G. Kaelin, *Annu. Rev. Biochem.* **2005**, *74*, 115.
[58] J. R. A. dos Santos-Pinto, H. A. Arcuri, F. G. Esteves, M. S. Palma, G. Lubec, *Sci. Rep.* **2018**, *8*, 14674.
[59] J. E. Garb, N. A. Ayoub, C. Y. Hayashi, *BMC Evol. Biol.* **2010**, *10*, 243.
[60] A. Zhao, T. Zhao, Y. Zhang, Q. Xia, C. Lu, Z. Zhou, Z. Xiang, M. Nakagaki, *Transgenic Res.* **2010**, *19*, 29.
[61] S. Sasaki, I. Nakagaki, *Membr. Biochem.* **1980**, *3*, 37.
[62] D. Long, W. Lu, Y. Zhang, Q. Guo, Z. Xiang, A. Zhao, *FEBS J.* **2015**, *282*, 89.
[63] K. Tanaka, S. Inoue, S. Mizuno, *Insect Biochem. Mol. Biol.* **1999**, *29*, 269.
[64] D. Weigel, J. Glazebrook, *Cold Spring Harb. Protoc.* **2006**, https://doi.org/10.1101/pdb.prot4666.
[65] H. Buschmann, P. Green, A. Sambade, J. H. Doonan, C. W. Lloyd, *New Phytol.* **2011**, *190*, 258.
[66] R. U. Schenk, A. C. Hildebrandt, *Can. J. Bot.* **1972**, *50*, 199.
[67] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, *Nat. Methods* **2012**, *9*, 676.
[68] C. Steentoft, S. Y. Vakhrushev, H. J. Joshi, Y. Kong, M. B. Vester-Christensen, K. T. B. G. Schjoldager, K. Lavrsen, S. Dabelsteen, N. B. Pedersen, L. Marcos-Silva, R. Gupta, E. Paul Bennett, U. Mandel, S. Brunak, H. H. Wandall, S. B. Levery, H. Clausen, *EMBO J.* **2013**, *32*, 1478.

 