Production of Recombinant Plant Gum With Tobacco Cell Culture in Bioreactor and Gum Characterization

Jianfeng Xu,[1,2] Elena Shpak,[1] Tingyue Gu,[2]
Murray Moo-Young,[2,3] Marcia Kieliszewski[1]

[1]Department of Chemistry and Biochemistry, Ohio University,
350 W. State St., Athens, Ohio 45701; telephone: 740-597-1225;
fax: 740-597-1772; e-mail: xuj@ohio.edu
[2]Department of Chemical Engineering, Ohio University, Athens, Ohio 45701
[3]Department of Chemical Engineering, University of Waterloo, Waterloo, ON,
Canada, N2L 3G1

Abstract: Many plant gums, such as gum arabic, contain hydroxyproline-rich glycoproteins (HRGPs), which are also abundant components of the plant cell extracellular matrix. Here we expressed in transgenic BY2 *Nicotiana tabacum* (tobacco) cells, a synthetic gene encoding a novel HRGP-based gum, designated gum arabic-8 or $(GA)_8$. $(GA)_8$ encoded eight repeats of the consensus polypeptide sequence of gum arabic glycoprotein (GAGP): Gly-Pro-His-Ser-Pro-Pro-Pro-Pro-Leu-Ser-Pro-Ser-Pro-Thr-Pro-Thr-Pro-Pro-Leu, in which most of the Pro residues were posttranslationally modified to hydroxyproline (Hyp). $(GA)_8$ was expressed as a green fluorescent protein (GFP) fusion protein targeted to the culture medium, $(GA)_8GFP$. The culture of the transgenic cells in a 5-L bioreactor showed that the production of $(GA)_8GFP$ was cell growth-associated. The extracellular yield of $(GA)_8GFP$ was 116.8 mg/L after 14 days of culture and accounted for 87% of the total fusion protein expressed. $(GA)_8GFP$ was purified from the culture medium by a combination of hydrophobic interaction, gel permeation, and reversed phase chromatography. Biochemical characterization indicated that the amino acid composition of the $(GA)_8$ module, after removal of GFP by proteolysis, was virtually identical to that of predicted by the GAGP consensus sequence and that carbohydrate, which occurred as arabinogalactan polysaccharides and small oligoarabinosides O-linked through the Hyp residues, accounted for 84% of the molecules' dry weight. Functional assays showed that $(GA)_8$ exhibited low viscosity in aqueous solution similar to native GAGP. However, neither GFP alone nor the $(GA)_8$ module could emulsify orange oil. However, the fusion protein $(GA)_8GFP$ possessed 1.28-fold better emulsification properties than native GAGP. This work demonstrates the feasibility and potential of a synthetic gene approach to the de novo design of novel glycoprotein-based gums and emulsifiers. © 2005 Wiley Periodicals, Inc.
Keywords: transgenic plants; gum; O-glycosylation; recombinant protein; bioreactor

Correspondence to: Jianfeng Xu

INTRODUCTION

Plant cells are increasingly attractive for the low-cost production of recombinant proteins and other biological products in large quantities and with few health risks to the human consumer (Doran, 2000; Fischer et al., 1999; Giddings et al., 2000). Advances in plant genetic engineering are creating new opportunities to use plants for the production of a range of novel polymers currently produced from petroleum or animal sources (Moire et al., 2003; Poirier, 1999). Examples include the biodegradable plastic polyhydroxyalkanoates (Matsumoto and Doi, 2003; Poirier et al., 1995) and protein-based biopolymers encoded by natural or synthetic genes such as spider silk (Scheller et al., 2001). Another important group of biopolymers amenable to genetic engineering and hyperproduction in plants is the plant gums, which are hydrocolloids often containing hydroxyproline (Hyp)-rich glycoproteins (HRGPs).

HRGPs are highly repetitive, structural glycoproteins unique to green algae and plants, where they contribute to cell wall networks and the plasma membrane cell wall interface (Lamport and Northcote, 1960; Serpe and Nothnagel, 1999). They are highly glycosylated mainly through the Hyp residues, with carbohydrate accounting for as much as 95% of the dry weight of the molecule. Indeed, the polypeptide backbone acts mainly as a template directing the orderly arrangement of glycan side-chains: acidic arabinogalactan polysaccharides are added to clustered noncontiguous Hyp residues, while blocks of contiguous Hyp are arabinosylated with linear chains of L-arabinose generally up to four residues long (Kieliszewski and Lamport, 1994; Shpak et al., 1999, 2001).

Not only are HRGPs crucial for normal plant development, but some members of the HRGP family have practical applications. For hundreds of years, HRGP containing hydrocolloids such as gums arabic, ghatti, tragacanth, and karaya have been widely used as emulsifiers, encapsulators,

and excipients in foods and pharmaceuticals (Garti and Reichman, 1993). Gum arabic, in particular, the exudate of *Acacia senegal* possesses desirable emulsifying properties and is the least viscous and the most soluble of the plant gums. These features and the fact that it is generally recognized as safe (GRAS) have led to its extensive use. However, in recent years an increasing demand for gum arabic has outpaced its uncertain supply; thus, reliable, alternative sources of consistent quality are sought.

Naturally occurring gum arabic is heterogeneous and polydisperse, with HRGP, the emulsifying component, comprising only ~10% of the total gum (Islam et al., 1997; Ray et al., 1995; Qi et al., 1991). This HRGP component, designated gum arabic glycoprotein (GAGP), is an arabinogalactan-protein (AGP), a type of HRGP containing arabinogalactan polysaccharides attached to Hyp residues. Although native GAGP has been isolated and well characterized (Goodrum et al., 2000; Qi et al., 1991), the mechanism by which GAGP emulsifies is unknown and its gene remains uncloned.

The GAGP polypeptide backbone is comprised mainly of hydrophilic amino acids in variations of a repetitive 19-residue consensus sequence: Gly-Pro-His-Ser-Hyp-Hyp-Hyp-Hyp-Leu-Ser-Hyp-Ser-Hyp-Thr-Hyp-Thr-Hyp-Hyp-Leu (Goodrum et al., 2000). Thus, we proposed that the arabinogalactan polysaccharide components rather than the polypeptide of GAGP conferred the emulsifying properties of GAGP, as the glycans possessed hydrophobic features such as terminal rhamnose residues and a β-1-3 galactan backbone (Goodrum et al., 2000). We also hypothesized that a synthetic gene approach might allow us to produce functional GAGP analogs in tobacco, given the similarities between the glycans present in GAGP and transgenic tobacco AGPs (Goodrum et al., 2000; Tan et al., 2003). To test this hypothesis we constructed a GAGP analog, gum arabic-8, $(GA)_8$, that contained a signal sequence and eight repeats of the consensus GAGP motif and expressed it in tobacco BY2 cells as a green fluorescent protein (GFP) fusion protein, designated $(GA)_8GFP$. Transgenic cells propagated in a 5-L stirred tank reactor demonstrated cell growth kinetics typical for suspension-cultured plant cells, with a gum yield in the medium of 116.8 mg/L. Purified $(GA)_8$, characterized after removal of the GFP domain, was hydroxylated and glycosylated similar to naturally occurring GAGP, although it contained more arabinogalactan polysaccharides. However, contrary to our predictions $(GA)_8$ was a poor emulsifier, as was the lone GFP domain. In contrast, $(GA)_8GFP$ surpassed GAGP in emulsifying ability and the stability of the emulsions produced.

MATERIALS AND METHODS

Synthetic Gene and Plasmid Construction

A synthetic gene, designated $(GA)_2$, encoding two variations of the GAGP consensus sequence, specifically the amino acid sequence Gly-Pro-His-Ser-Pro-Pro-Pro-Pro-Leu-Ser-Pro-Ser-Pro-Thr-Pro-Thr-Pro-Pro-Leu-Gly-Pro-His-Ser-Pro-Pro-Pro-Thr-Leu-Ser-Pro-Ser-Pro-Thr-Pro-Thr-Pro-Pro-Pro, was synthesized by primer extension of two mutually priming oligonucleotides (Integrated DNA Technologies, Coralville, IA): 5′-TAAAAGCTTGCCGGCCCT-CATAGCCCACCTCCACCATTATCACCATCACCTAC-TCCAACTCCTCCTTTGGGACCACACAGTCCA-3′, and 5′-AATAGAATTCACCATGGTCCGCCCGGGGGGTGG-TGTTGGGGGTTGGTGAAGGGGAAAGTGTAGGGGG-TGGACTGTGTGGTCCCAA-3′. The complementary regions of the two oligonucleotides are underlined. The oligonucleotides were designed using the codon bias for known HRGPs from the *Solanaceae* (Murray et al., 1989).
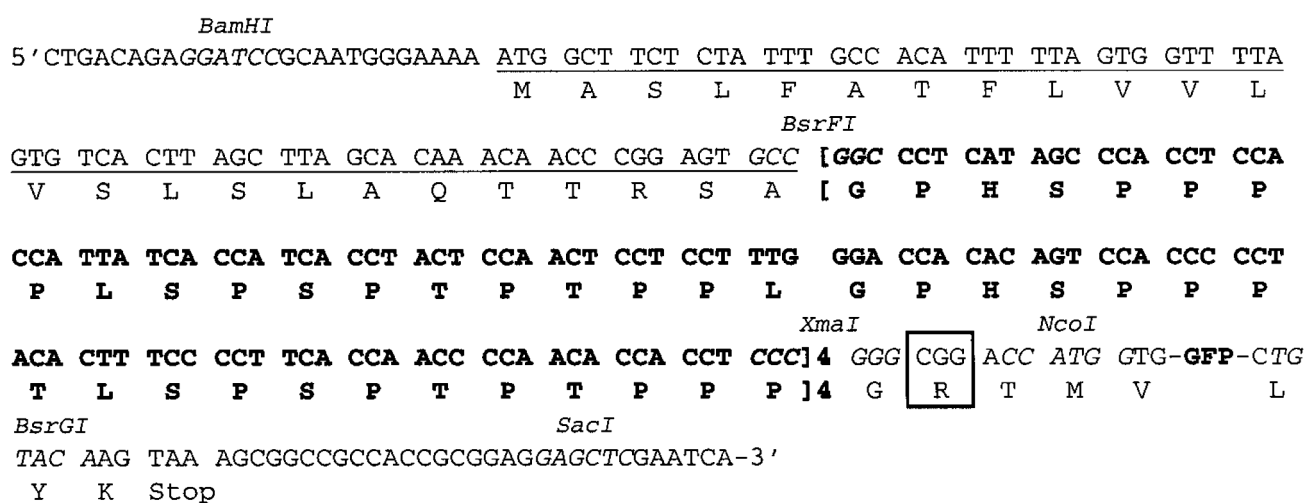
```
          BamHI
5'CTGACAGAGGATCCGCAATGGGAAAA ATG GCT TCT CTA TTT GCC ACA TTT TTA GTG GTT TTA
                             M   A   S   L   F   A   T   F   L   V   V   L
                                                     BsrFI
GTG TCA CTT AGC TTA GCA CAA ACA ACC CGG AGT GCC [GGC CCT CAT AGC CCA CCT CCA
V   S   L   S   L   A   Q   T   T   R   S   A  [G   P   H   S   P   P   P

CCA TTA TCA CCA TCA CCT ACT CCA ACT CCT CCT TTG GGA CCA CAC AGT CCA CCC CCT
P   L   S   P   S   P   T   P   T   P   P   L   G   P   H   S   P   P   P
                                                XmaI            NcoI
ACA CTT TCC CCT TCA CCA ACC CCA ACA CCA CCT CCC]4 GGG CGG ACC ATG GTG-GFP-CTG
T   L   S   P   S   P   T   P   T   P   P   P ]4 G   R   T   M   V       L
BsrGI                                  SacI
TAC AAG TAA AGCGGCCGCCACCGCGGAGGAGCTCGAATCA-3'
Y   K   Stop
```

Figure 1. The DNA sequence of $(GA)_8GFP$. The $(GA)_8$ (in bold) was inserted between the encoded signal sequence (underlined) and *GFP* to give the synthetic gene $(GA)_8GFP$. The restriction sites are in italics. The arginine, or "R" residue (boxed), for tryptic cleavage to remove GFP from the fusion protein was encoded by the *CGG* codon sandwiched between the *XmaI* and *NcoI* restriction sites.

$(GA)_2$ was subcloned into *pUC18* as a *HindIII/EcoRI* restriction fragment to generate the plasmid *pUC-(GA)_2*. The construction of four GAGP consensus repeats, a gene designated $(GA)_4$, involved exploiting compatible but non-regenerable restriction sites (*XmaI* and *BsrFI*) in *pUC-(GA)_2* to double number of consensus repeats as described earlier (Lewis et al., 1996; Shpak, 2001). The process was repeated to build a synthetic gene having eight GAGP consensus repeats, designated $(GA)_8$ (Fig. 1).

Synthetic gene $(GA)_8$ was subcloned into *pUC-SS^{tob}-GFP* (Shpak et al., 1999) as a *BsrFI/NcoI* fragment between *SS^{tob}*, which encodes the extensin signal sequence (SS^{tob}) from tobacco (De Loose et al., 1991), and the gene encoding the enhanced green fluorescent protein (GFP) (ClonTech, Palo Alto, CA) to generate the plasmid designated *pUC-SS^{tob}-(GA)_8-GFP*. A tryptic cleavage site was introduced between encoded $(GA)_8$ and *GFP* (Fig. 1) to enable us to remove GFP from the fusion protein. The SS^{tob}-$(GA)_8$-*GFP* gene was then subcloned as a *BamHI/SacI* fragment into the plant transformation vector *pBI121* (ClonTech) in place of the β-glucuronidase gene, generating the plasmid *pBI-SS^{tob}-(GA)_8-GFP*. Expression of *SS^{tob}-(GA)_8-GFP* was under the control of the 35S cauliflower mosaic virus (CaMV) promoter. DNA sequencing of *SS^{tob}-(GA)_8-GFP* was performed at the Guelph Molecular Supercenter (University of Guelph, Ontario, Canada).

Tobacco Cell Transformation and Cell Line Selection

Plasmid *pBI-SS^{tob}-(GA)_8-GFP* was transferred into *Agrobacterium tumefaciens* strain LBA4404 by the freeze-thaw method (Holsters et al., 1978) and suspension-cultured tobacco cells (*Nicotiana tabacum*, BY2) were subsequently transformed with the *Agrobacterium* as described by An (1985). Transformed tobacco cells were selected on solid Schenk & Hildebrandt (SH) medium (Schenk and Hildebrandt, 1972) containing 30 g/L sucrose, 0.4 mg/L 2,4-dichlorophenoxyacetic acid (2,4-D), 200 mg/L kanamycin (Sigma, St. Louis, MO) and 400 mg/L timentin (SmithKline Beecham, Pittsburgh, PA).

Growth and Maintenance of Suspension Cultures

Transformed tobacco cells were grown in liquid SH medium containing kanamycin but lacking timentin. The pH of the medium was adjusted to 5.8 before sterilization. The stock cell cultures were maintained in 1,000-mL Erlenmeyer flasks containing 400 mL medium. Flasks were placed on a gyratory shaker at 100 rpm under continuous illumination at 25°C. Subculture was carried out every 7 days with a 5% (v/v) inoculum density.

Bioreactor Culture

Determination of cell growth kinetics and recombinant gum production was performed with transformed cells grown in a 5-L Celligen Cell Culture System (New Brunswick Scientific, Edison, NJ). The bioreactor was operated in the stirred tank reactor (STR) mode with a marine impeller having a 50 mm diameter. The agitation rate was maintained at 80 rpm. The culture temperature was maintained at 25°C and the aeration rate was 0.1 vvm. Dissolved oxygen was automatically controlled at 70% of air saturation by manipulating the oxygen concentration in the gas sparging stream. The pH of the medium was allowed to vary during the batch culture and monitored continuously using a pH electrode (Ingold, Bedford, MA). An 8-day-old cell suspension was used as the "seed" for inoculation. Ten-ml samples were withdrawn at 2-day intervals for the determination of the fresh and dry weights of the cells, as well as $(GA)_8$GFP content, and the sugar concentration in the medium.

Isolation of $(GA)_8$GFP Fusion Glycoprotein From Medium

The culture medium harvested after 12–14 days of culture was concentrated 10-fold by rotor-evaporation under vacuum at 30°C. A 100–200-ml aliquot of medium adjusted to 2 M sodium chloride was loaded onto a hydrophobic-interaction column (HIC) (Phenyl-Sepharose 6 Fast Flow, 16 × 700 mm, Amersham Pharmacia Biotech, Piscataway, NJ) equilibrated in 2 M sodium chloride, and eluted with a stepwise gradient of 2 M sodium chloride, 1 M sodium chloride, and then distilled water. The green fluorescent fraction containing $(GA)_8$GFP eluted in distilled water and was concentrated by freeze-drying before fractionation on a semipreparative Superose-12 gel permeation (GPC) column (16 × 700 mm, Amersham Pharmacia Biotech) equilibrated in Superose buffer (200 mM sodium phosphate buffer, pH 7, containing 0.005% sodium azide). The visibly green fluorescent fraction collected from the GPC column was further fractionated on a Hamilton PRP-1 semipreparative reversed phase column (10 μm, 7 × 305 mm, Hamilton, Reno, NV) equilibrated with starting Buffer A (0.1% aqueous trifluoroacetic acid). Proteins were eluted over 100 min with a 0–70% linear gradient of Buffer B (0.1% trifluoroacetic acid containing 80% acetonitrile, v/v) at a flow rate of 1.0 ml/min. Absorbance was measured at 220 nm.

Removal of GFP From the Fusion Glycoprotein by Tryptic Digestion

Fifty mg of $(GA)_8$GFP was heat-denatured in boiling water for 2 min, cooled, then combined with an equal volume of freshly prepared 2% (w/v) ammonium bicarbonate containing 10 mM calcium chloride and 500 μg trypsin. After an overnight incubation at room temperature, $(GA)_8$ lacking the GFP unit was purified by gel filtration and reversed-phase chromatography on the Superose-12 and PRP-1 columns, respectively.

Assay of Culture Medium for (GA)$_8$GFP by Gel Permeation Chromatography on the Superose-12 Analytical Column

Cell culture medium (100 µl) was injected directly into a Superose-12 analytical GPC column (10 ×300 mm, Amersham Pharmacia Biotech) equilibrated in Superose buffer (see above) and eluted at a flow rate of 0.2 ml/min. The eluate was monitored by a HP 1100 Series flow-through fluorometer (Hewlett Packard, Corvallis, OR) at an excitation wavelength of 488 nm and an emission wavelength of 520 nm.

Sodium Dodecylsulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Samples (10 µg) were mixed with an equal volume of 2× reducing sample buffer and resolved on a 10% SDS-PAGE gel (BioRad, Hercules, CA) as described by Laemmli (1970).

Hyp-Polysaccharide/Oligosaccharide Isolation and Hyp glycoside profiles assay

Purified (GA)$_8$ ($\sim$ 10 mg) was dissolved in 0.44 N sodium hydroxide and hydrolyzed at 105°C for 18 h. The base hydrolysate was chilled on ice, neutralized with 1 N hydrochloric acid, then applied to a Superdex peptide HR 10/30 column (Amersham Pharmacia Biotech) equilibrated in 20% acetonitrile, and eluted at a flow rate of 0.3 ml/min. Fractions (0.6 ml) were collected, lyophilized, and assayed for Hyp and sugar composition. For Hyp glycoside profiling, the neutralized base hydrolysate was fractionated on a C2 cation exchange column, as described earlier (Shpak et al., 1999).

Colorimetric Hydroxyproline Assays

Sample Hyp content was assayed colorimetrically as described by Kivirikko and Liesmaa (1959).

Biomass and Recombinant (GA)$_8$GFP Quantification

Cultured cells were harvested by vacuum filtration on a sintered funnel and washed three times with distilled water before fresh weight determination. The cells were then dried in an oven at 60°C for 2 days, then weighed to determine dry weight. Biomass was expressed as grams of cell dry weight per liter medium (g/L). (GA)$_8$GFP was quantified by assaying for GFP (Liu et al., 2001). Twenty-five-µL samples adjusted to pH 8 were injected for measurement of the fluorescence intensity. The fluorescence measurements were performed in 200 mM phosphate buffer (pH 8) using an HP 1100 Series flow-through fluorometer at an excitation wavelength of 488 nm and an emission wavelength of 520 nm. Purified, dried, and carefully weighed (GA)$_8$GFP was used as the calibration standard. Culture medium was used to determine the extracellular (GA)$_8$GFP content. For intracellular (GA)$_8$GFP quantification, 0.5 g fresh weight of cells was washed three times with water, resuspended in 5 ml of phosphate buffer (50 mM, pH 8) containing 200 mM sodium chloride, then packed in ice and lysed by sonication for 10 min using a Sonic Dismembrator (Fisher Scientific, Pittsburgh, PA) equipped with a Microtip probe. After centrifugation of the lysate at 12,000g for 15 min, the supernatant was assayed for GFP fluorescence as described above.

Amino Acid Composition and Sequence Analysis

The N-terminal sequence of (GA)$_8$GFP was analyzed at the Michigan State University Macromolecular Facility on a 477-A Applied Biosystems Gas Sequencer. The (GA)$_8$ amino acid composition was determined by reversed-phase HPLC on a Beckman Gold System (Beckman Instruments, Palo Alto, CA) as described by Bergman et al. (1986).

Sugar Analysis

The total sugar content of the medium was assayed colorimetrically by the anthrone method (Dische, 1962). The neutral sugar composition of (GA)$_8$GFP was analyzed as alditol acetates (Bhatti et al., 1970) by gas chromatography using a 6 foot × 2 mm polyethylene glyco succinate 224 column programmed from 130–180°C at 4°C/min. Uronic acid content was assayed colorimetrically using the m-hydroxybiphenyl method (Blumenkrantz and Asboe-Hansen, 1973).

Functional Analysis

The emulsification properties of (GA)$_8$, (GA)$_8$GFP, and GFP alone, including the emulsifying ability (EA) and emulsifying stability (ES) of each, were measured according to the method of Pearce and Kinsella (1978). Orange oil (Sigma) was used as the oil phase. Viscosities were measured at 25°C using a Zeitfuchs Cross-Arm viscometer (Cannon Instrument, State College, PA). The viscometer constant was 0.00965 cSt/s.

RESULTS

Isolation of Transformed Cell Cultures and (GA)$_8$GFP Production

Ten stable transformed cell lines were isolated, based on the GFP fluorescence in the culture medium. (GA)$_8$GFP mainly occurred in the medium and remained intact judging by GPC, which showed only a minor peak corresponding to GFP (Fig. 2A,B) that was presumably cleaved from the (GA)$_8$GFP fusion protein. All the cell lines examined produced products of identical size; however, yields between the lines differed as much as 10-fold (Fig. 3).
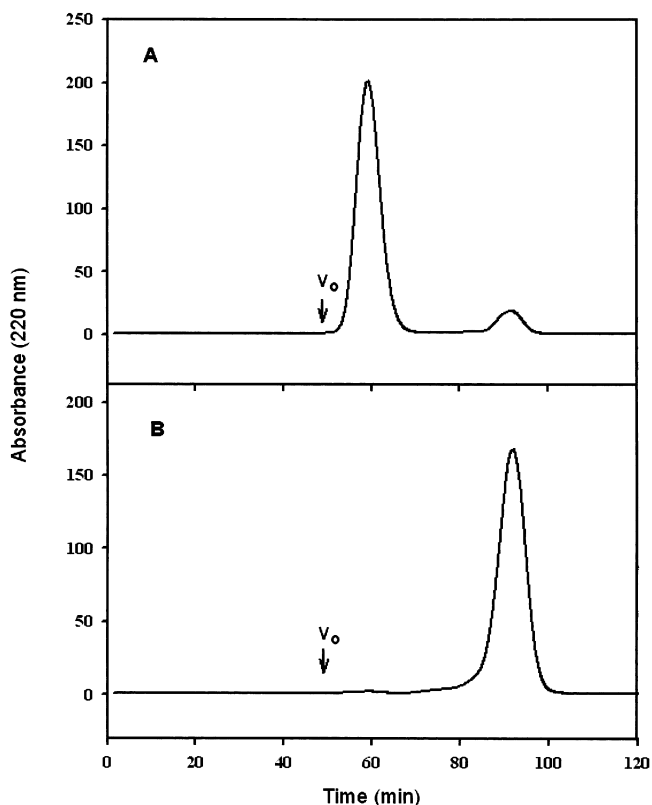
Figure 2. Superose-12 gel permeation chromatography of the culture medium from the culture of transformed cells. **A:** Size fractionation of medium from cells expressing $(GA)_8GFP$ showed a major fluorescent peak eluting at 60 min and a minor peak at 90 min that corresponded to GFP, presumably cleaved from the fusion protein after secretion into the medium. **B:** The GFP standard eluted at 90 min on the Superose-12 column. The column eluates were assayed for fluorescence using a 488 nm excitation wavelength, monitoring the emission at 520 nm.

The highest-yielding cell line, designated G6, produced 125.4 mg/L of $(GA)_8GFP$ in medium and was chosen for further characterization.

Bioreactor Culture

Culture of the G6 cells was scaled up to the 5-L bioreactor and the kinetics of cell growth and $(GA)_8GFP$ production and properties determined (Fig. 4). The transgenic cells grew well in the 5L Celligen Cell Culture System, reaching a maximum biomass of 13.8 g/L on day 12. No detrimental effects from agitation were observed using the marine impeller rate of 80 rpm. The growth kinetics followed a typical plant cell growth pattern with a lag phase of 2 days, an exponential growth phase between 2–10 days, followed by a stationary phase after 10 days. The growth kinetic data are summarized in Table I. The production of $(GA)_8GFP$ was growth-associated: about 87% of it was secreted into the medium after 10 days in culture. The maximum $(GA)_8GFP$ yield obtained in the medium was 116.8 mg/L on day 14, which was comparable to that obtained from cells grown in shake flasks. The medium pH decreased
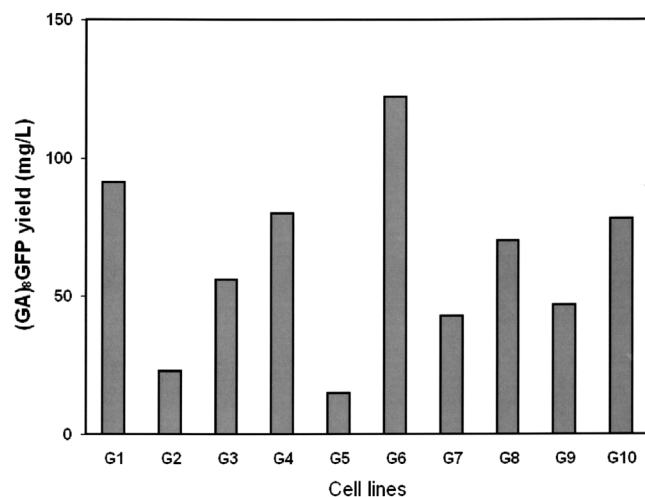

Figure 3. The extracellular $(GA)_8GFP$ yields of 10 different tobacco cell lines. Medium samples were assayed after 14 days of culture and the $(GA)_8GFP$ yields calculated from the GFP fluorescence in the medium.

from 5.8 to 4.5 over the first 4 days, then increased slowly up to 6.4 by stationary phase, which is typical for many plant cell cultures (Stoner et al., 1997; Xu et al., 1998). Sugar in the medium was almost completely consumed
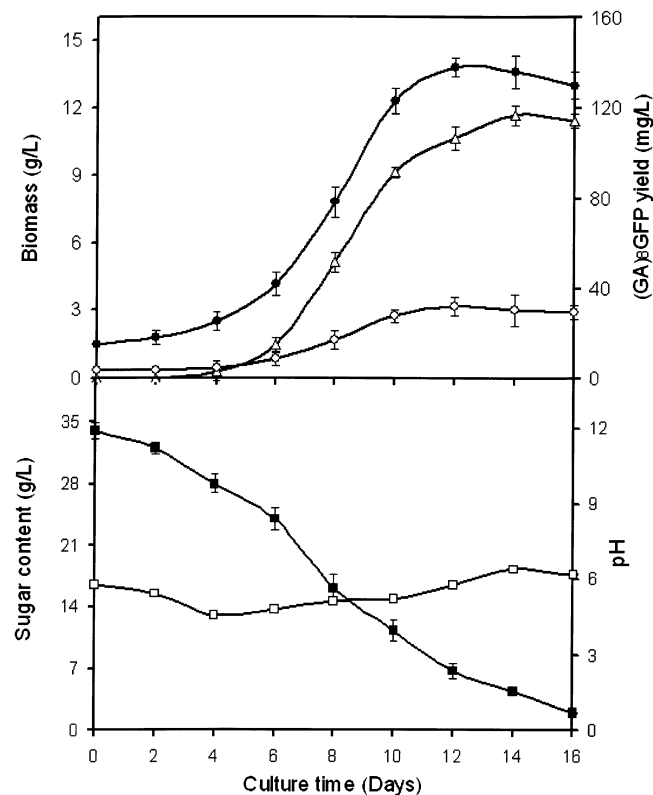

Figure 4. Time course of cell growth and $(GA)_8GFP$ production of the transgenic tobacco cell (G6) culture in a 5-L bioreactor. G6 cells were assayed for biomass production (●); extracellular $(GA)_8GFP$ content (△); intracellular $(GA)_8GFP$ content (○); sugar content in the medium (■); and medium pH (□). Error bars represent the standard error of three measurements.

Table I. Summary of growth kinetics data and $(GA)_8$GFP yield from cell line G6.

X_0 g/L	X_{max} g/L	$Y_{x/s}$	μ day^{-1}	$T_{1/2}$ day	Y_{gum} mg/L	$Y_{gum/x}$ mg/g
1.5	13.8	0.45	0.38	2.6	134.2	9.7

X_0: Inoculation concentration (g/L); X_{max}: Maximum biomass (g/L); $Y_{x/s}$: Yield coefficient; μ: Specific growth rate during exponential growth phase (day^{-1}); $T_{1/2}$: Doubling time (day); Y_{gum}: Total $(GA)_8$GFP yield, including intracellular and extracellular $(GA)_8$GFP (mg/L); $Y_{gum/x}$: Specific $(GA)_8$GFP yield (mg/g), total $(GA)_8$GFP yield (mg/L)/maximum biomass (g/L).

during the culture cycle. The average yield coefficient was calculated to be $Y_{x/s} = 0.45$.

Isolation of $(GA)_8$GFP and $(GA)_8$

The protein recoveries of $(GA)_8$GFP after fractionation on the HIC and GPC columns was 82% and 85%, respectively. $(GA)_8$GFP recoveries from the final fractionation step, the PRP-1 reversed-phase column, was 74%. Both $(GA)_8$GFP and $(GA)_8$ recovered after cleavage of $(GA)_8$GFP with trypsin and GPC on the Superose 12 column, produced single symmetric peaks on the reversed phase column (Fig. 5).
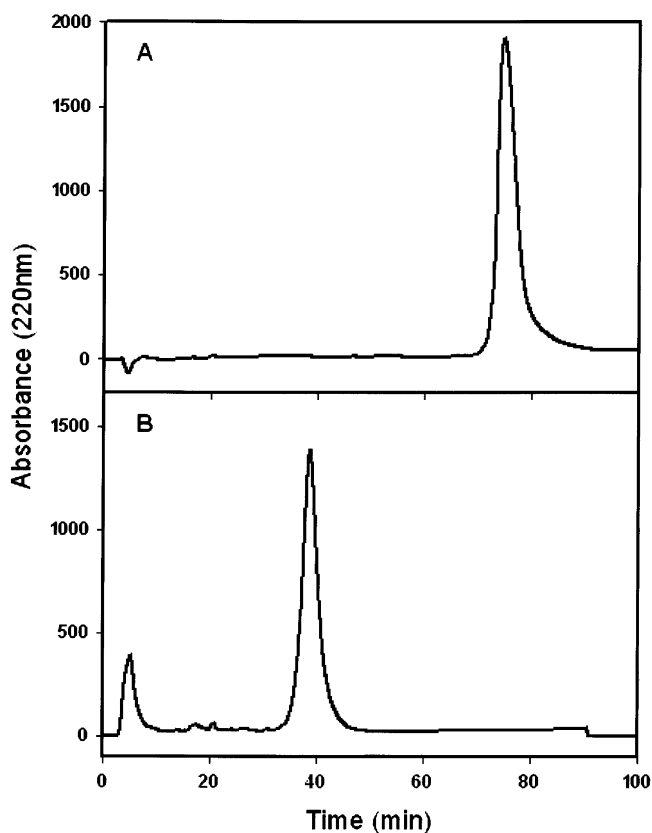

Figure 5. Purification of the $(GA)_8$GFP and $(GA)_8$ by reversed-phase chromatography. **A:** $(GA)_8$GFP eluted at 75 min. **B:** Removal of the relatively hydrophobic GFP domain from the fusion protein resulted in a $(GA)_8$ module that eluted much earlier from the column (40 min).

From one liter of culture medium, we isolated 65.5 mg of $(GA)_8$GFP, or 32.0 mg of $(GA)_8$ gum module.

Visualization of purified $(GA)_8$GFP and $(GA)_8$ after SDS-PAGE (Fig. 6) showed fuzzy bands typical of HRGPs which possess extensive glycan microheterogeneity and also bind SDS poorly due to their glycosylation and biased amino acid compositions. SDS-PAGE estimated the relative molecular mass range as 85–100 kDa for $(GA)_8$ and 110–140 kDa for $(GA)_8$GFP.

$(GA)_8$ Amino Acid Composition and N-Terminal Sequence Analyses

The amino composition of $(GA)_8$ (Table II) closely matched that predicted by the gene, except that most of the Pro residues were posttranslationally modified to Hyp. The N-terminal amino acid sequence of $(GA)_8$, Gly-Pro-His-Ser-Hyp-Hyp-Hyp-Hyp-Leu-Ser-Hyp-Ser-Hyp-Thr-Hyp-Thr-Hyp-Hyp-Leu-Gly-Pro-His..., was consistent with the gene sequence and amino acid composition.

Glycosylation of $(GA)_8$

Table III shows the monosaccharide composition of $(GA)_8$, which was consistent with the composition of native GAGP and of other known AGPs that are rich in galactose (Gal) and arabinose (Ara) with lesser amounts of uronic acids (Ura) and rhamnose (Rha) (Goodrum et al., 2000; Qi et al., 1991; Serpe and Nothnagel, 1999; Zhao et al., 2002). Sugar accounted for 84% of the total dry weight of $(GA)_8$, or 65%
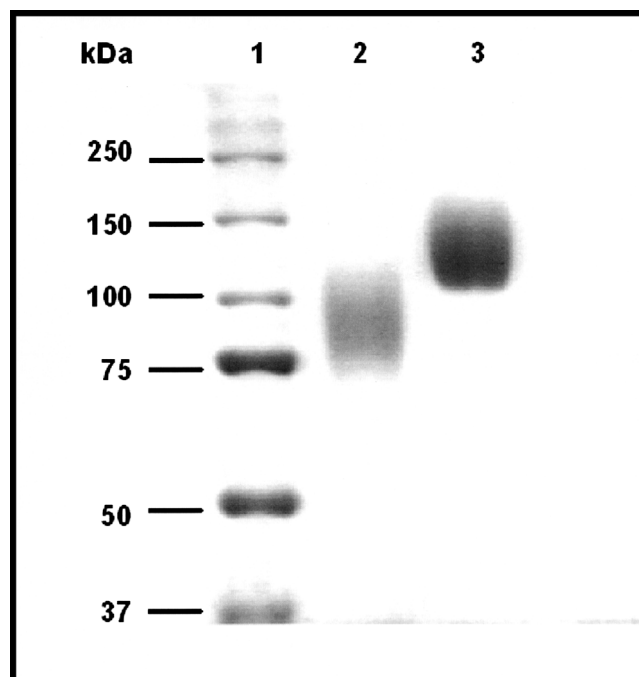

Figure 6. SDS-PAGE of the purified $(GA)_8$GFP and $(GA)_8$ module. The glycoproteins were detected by silver staining. Lane 1: molecular weight markers; lane 2: $(GA)_8$; Lane 3: $(GA)_8$GFP.

Table II. Amino acid composition (mol%) analysis of $(GA)_8$ and GAGP.

	$(GA)_8$		
Amino acid	Measured	Predicted[a]	GAGP[b]
Hyp	44.1	⎫	40.0
Pro	7.5	⎬ 51.3	8.0
Thr	12.4	12.8	10.2
Ser	16.7	16.0	2.2
Gly	5.1	5.8	4.5
His	5.3	5.1	6.6
Ala	0.5	0.6	1.2
Leu	8.1	7.7	6.4
Phe	0.0	0.0	0.5
Arg	0.5	0.6	0.0
Ile	0.0	0.0	0.2

[a]Amino acid composition was calculated from the $(GA)_8$ DNA sequence.
[b]Data from Goodrum et al. (2000).

Table IV. Hyp-glycoside profiles of $(GA)_8$ and native GAGP.

	Percentage of total Hyp	
Hyp-glycoside	$(GA)_8$	GAGP[a]
Hyp-polysaccharide	43	25
Hyp-Ara$_4$[b]	0	5
Hyp-Ara$_3$	21	31
Hyp-Ara$_2$	21	17
Hyp-Ara$_1$	8	10
Non-glycosylated Hyp	7	12

[a]Qi et al. (1991).
[b]Hyp-Ara$_{1-4}$: Hyp-arabinosylation with 1–4 arabinose residues.

of the $(GA)_8$GFP fusion protein. However, the distribution of Hyp in the different glycosides differed significantly, with GAGP having more arabinosylated Hyp and less polysaccharide addition to Hyp than $(GA)_8$, as shown in Table IV.

Superdex size fractionation of $(GA)_8$ base hydrolysates separated the resulting Hyp glycosides into two major size categories (Fig. 7), a large molecular component that eluted near the column void volume (peak 1) and a much smaller component (peak 2). Sugar composition analyses of the large molecular weight fraction indicated it contained Ara, Gal, Rha, and Ura, while the smaller molecular weight glycans contained only Ara (Table V). The molar ratio of sugar/Hyp indicated that the large molecular weight polysaccharide component contained an average of 20 sugar residues per Hyp and had the empirical formula: Gal_{10}-$Ara_4Ura_4Rha_2$.

Calculation of Molecular Mass of $(GA)_8$

Judging by the gene sequence and amino acid composition, the molecular mass of the $(GA)_8$ polypeptide was 15 kDa. Sugar analyses indicated that sugar accounted for 84% of the dry weight of the glycosylated $(GA)_8$. Thus, we estimated the molecular mass of glycosylated $(GA)_8$ to be 15 kDa/(100%-84%) or 94 kDa. The molecular mass of the fusion protein $(GA)_8$GFP was therefore 94kDa+27kDa

(GFP) or 121 kDa, estimations that were consistent with those estimated by SDS-PAGE.

Functional Analysis of the Recombinant Gum

$(GA)_8$ exhibited lower viscosity than native GAGP (Fig. 8) and inferior emulsification properties (Fig. 9), its emulsifying ability being only 1/10 of that of native GAGP and the emulsions formed by $(GA)_8$ separated immediately into two phases at room temperature. In contrast, the fusion glycoprotein $(GA)_8$GFP was 1.28 times more effective than native GAGP and the emulsion was as stable.

DISCUSSION

Genetic engineering offers a facile route to the production of novel protein-based biopolymers that cannot be made through chemical synthesis (Scheller et al., 2001; Moire et al., 2003). Here we have extended the "genetic" approach to include the use of synthetic genes to engineer O-linked glycoproteins of controlled structure, including the glycan substituents. We chose to make an analog of gum arabic glycoprotein (GAGP) because GAGP is a highly repetitive HRGP and is therefore well suited to glycoprotein engineering using a synthetic gene approach. Furthermore, native GAGP from *Acacia senegal* possesses desirable properties that can be easily measured for comparison to GAGP analogs produced in tobacco cells. Tobacco cells were chosen to host the GAGP analog, $(GA)_8$GFP, because tobacco cells are readily transformed and possess the posttranslational machinery to make HRGPs that are hydroxylated and glycosylated like native

Table III. Sugar analysis of $(GA)_8$ and native GAGP.[a]

	Molecular	Sugar composition (mole %)				Sugar content
Sample	mass (kDa)	Ara	Gal	Rha	Ura	(% dry weight)
$(GA)_8$	94	53.2 ± 2.1	25.3 ± 1.1	13.2 ± 1.4	9 ± 0.5	84.0 ± 0.6
GAGP[b]	1500	36	46	10	9	>80

[a]Ara: arabinose; Gal: galactose; Rha: rhamnose; Ura: Uronic acids.
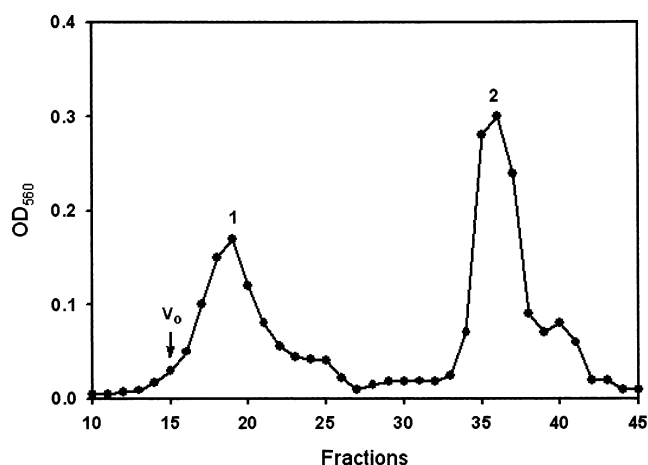[b]Refer to Qi et al. (1991).

Figure 7. Size fractionation of the $(GA)_8$ base hydrolysate on the Superdex peptide column. Each fraction was assayed for Hyp. The two major peaks correspond to the large molecular weight Hyp-arabinogalactan polysaccharides (peak 1) and the smaller Hyp-oligoarabinosides (peak 2), respectively.

Figure 8. The viscosity of $(GA)_8$ and GAGP solutions measured at varying concentrations. ($\bigcirc$) GAGP solution; ($\bullet$) $(GA)_8$ solution. Error bars represent the standard error of three measurements.

GAGP. Indeed, the legumes, which include *A. senegal*, and the *Solanaceae* share the same Hyp-O-glycosylation code and Hyp-glycan structures judging from recent work characterizing "designer" HRGPs from transgenic tobacco (Tan et al., 2003; Shpak et al., 1999, 2001).

Beginning with the 19-residue repetitive consensus motif of GAGP (Goodrum et al., 2000), we designed the synthetic gene, *(GA)₂* that encoded two naturally occurring variations (variant amino acids are underlined) of the GAGP consensus sequence above, specifically Gly-Pro-His-Ser-Pro-Pro-Pro-Pro-Leu-Ser-Pro-Ser-Pro-Thr-Pro-Thr-Pro-Pro-Leu-Gly-Pro-His-Ser-Pro-Pro-Pro-Thr-Leu-Ser-Pro-Ser-Pro-Thr-Pro-Thr-Pro-Pro-Pro. Genes encoding four GAGP repeats, *(GA)₄* and then eight repeats *(GA)₈* were constructed from *(GA)₂* using an approach described earlier involving complementary, nonregenerable restrictions sites (Lewis et al., 1996; Shpak, 2001). Although native GAGP is a large AGP, containing at least 20 highly glycosylated consensus repeats, we characterized a much smaller GAGP analog here, as earlier work with GAGP analogs showed tobacco cells gave increasingly lower yields as the number of consensus repeats increased from 8 to 20 and 36 repeats (Shpak, 2001), with 8 being optimum. $(GA)_8$ was expressed
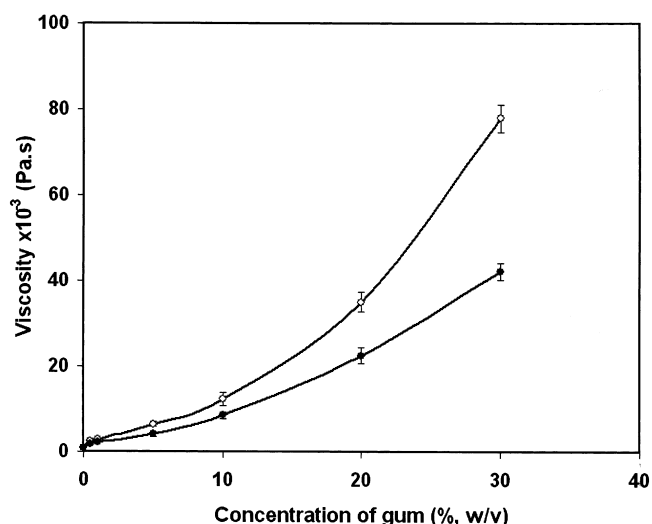
as a GFP fusion protein to facilitate selection of highly expressing tobacco cell lines through visualization of GFP fluorescence in the medium (Fig. 2). GFP also enhanced $(GA)_8$ hydrophobicity, enabling purification of the glycoprotein, as individual AGPs are notoriously difficult to purify due to similarities in amino acid sequence and glycosylation (Shpak et al., 1999; Zhao et al., 2002). $(GA)_8$GFP was readily targeted to the ER/Golgi membrane system (Ituriaga et al., 1989) and out to the culture medium by the tobacco extensin signal sequence (Shpak et al., 1999), where more than 80% of the total expressed gum occurred.

Although proteolysis of recombinant proteins is often a significant problem in transgenic plant cell culture (Kusnadi et al., 1997), there was little degradation of $(GA)_8$GFP judging by size fractionation of the culture medium, which gave one major, symmetrical fluorescent peak corresponding to $(GA)_8$GFP and only a very minor peak corresponding to the GFP cleavage product (Fig. 2). The N-terminal sequence also corroborated the integrity of the polypeptide backbone. Undoubtedly the extensive glycosylation (Jentoft, 1990; Hart, 1992; Dealwis and Wall, 2004) and abundance of secondary amino acids protected the $(GA)_8$ polypeptide from proteolysis, while GFP is also

Table V. Sugar composition of Hyp-polysaccharide and Hyp-oligosaccharide isolated from the base hydrolysis of $(GA)_8$ and native GAGP.

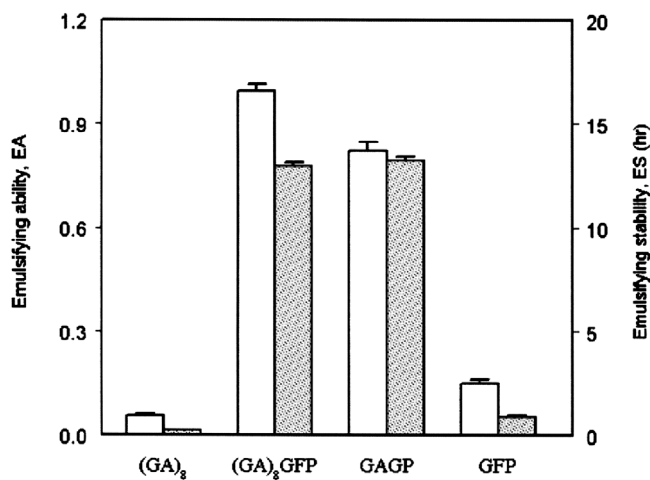| Sugar residue	$(GA)_8$ (mol%)		GAGP[a] (mol%)	
	Polysaccharide	Oligosaccharide	Polysaccharide	Oligosaccharide
Ara	35	100	38	100
Gal	41	0	34	0
Rha	12	0	13	0
Ura	12	0	15	0
Sugar/Hyp	20.1	2.1	23.4	2.3

[a]Refer to Qi et al. (1991).

Figure 9. The emulsification properties of $(GA)_8$, $(GA)_8GFP$, GAGP, and GFP. The emulsifying ability (EA) was determined from the turbidity measured immediately after the emulsion had formed. The emulsion stability (ES) was estimated by calculating the half-life of the turbidity. White bars: EA; Shaded bars: ES. Error bars represent the standard error of three measurements.

known to be fairly resistant (Ward, 1981). The susceptible Arg residue engineered between the HRGP and GFP domains of the fusion protein probably gave rise to the small amount of GFP in the culture medium of transformed cells (Fig. 2A). Stability is a feature of HRGPs in general, presumably as part and parcel of the structural role they play in the rugged extracellular environment (Showalter, 2001), and is a feature that might prove useful in the production of recombinant glycoproteins.

The high-yield G6 cell line grew well in the 5-L bioreactor, with a specific growth rate of 0.38 day^{-1}, which was comparable to rates reported for other plant cell cultures (Stoner et al., 1997; Xu et al., 1998). Suspension cultured plant cells are fragile in comparison to microorganisms, as plant cells are large and often damaged during culture in a bioreactor (Takeda et al., 1994). However, judging by staining with Evan's blue, the cell viability observed here remained high until the culture reached the stationary phase, after which lysis occurred (data not shown). We observed no detrimental agitation effects from a marine impeller rate of 80 rpm. Furthermore, the cells produced 116.8 mg/L of $(GA)_8GFP$, or 40.9 mg/L of the $(GA)_8GFP$ polypeptide backbone (35% of the dry weight of $(GA)_8GFP$), a 2–5-fold yield increase over other reports of recombinant protein production in plant cell cultures with a constitutive 35S CaMV promoter, such as 5–10 mg/L of Guy's 13 mouse IgG$_1$ antibody (Sharp and Doran, 2001a), 12 mg/L of GFP (Su et al., 2004), and 20–22 mg/L of vaccine hepatitis B surface antigen (Smith et al., 2002). We expect $(GA)_8GFP$ production could be considerably enhanced by optimizing the culture conditions and operational parameters, or by employing "super" promoters for boosting gene expression (Choi et al., 2003; Kusnadi et al., 1997; Sharp and Doran, 2001b; Shin et al., 2003).

Biochemical characterization of purified $(GA)_8GFP$ and native GAGP allowed us to compare the prolyl hydroxylase specificities, Hyp-glycosylation codes, and glycan structures of tobacco to those of *Acacia*. Like native GAGP, which contains 90% sugar on a dry weight basis (Qi et al., 1991), $(GA)_8$ was heavily O-glycosylated (84% dry weight) and both contained Ara and Gal as major components, with lesser amounts of Ura and Rha, although $(GA)_8$ contained more Ara and less Gal (Table III). The amount of Hyp and the type of glycans attached to the Hyp residues were also similar in GAGP and $(GA)_8GFP$ (Tables II, III, V); however, the distribution of Hyp in the different glycosides differed significantly, with GAGP having more arabinosylated Hyp and less polysaccharide addition to Hyp than $(GA)_8$ (Table IV). This is likely due to differences in the prolyl hydroxylase specificities between tobacco and *Acacia*, which in turn would influence the Hyp-glycosylation profiles as follows.

Earlier work demonstrated that blocks of tripeptidyl Pro are completely hydroxylated in *Acacia* (Goodrum et al., 2000) and therefore are Hyp-arabinosylation sites; however, in tobacco blocks of tripeptidyl Pro are not completely hydroxylated (Shpak et al., 1999, 2001). This incomplete Pro hydroxylation of tripeptidyl Pro blocks gives rise to a mixture of contiguous and noncontiguous Hyp residues and thus a mixture of arabinosylated Hyp and polysaccharide adducts on Hyp at these sites in tobacco. $(GA)_8$ encoded a polypeptide containing four blocks of tetrapeptidyl Pro, four dipeptidyl Pro blocks, and eight instances of tripeptidyl Pro. From the $(GA)_8$ amino acid composition and known prolyl hydroxylase specificities for tobacco, we calculated that $(GA)_8$ encoded 152 amino acids, 80 of them Pro. The amino acid composition (Table II) indicated there are 11 Pro residues in $(GA)_8$ (7.5 mole percent) and about 69 Hyp residues. Eight Pro residues likely occur in the Gly-Pro-His tripeptide that is not hydroxylated in $(GA)_8$ or in native GAGP, judging by amino acid sequence analysis (Goodrum et al., 2000). Therefore, the remaining three Pro residues occur somewhere in the eight tripeptidyl Pro/Hyp motifs of $(GA)_8$ and, judging from earlier work, will result in mixed Hyp-glycosylation at those sites. We predict that 24 of the 69 Hyp residues in $(GA)_8$ occur in Ser-Hyp$_4$ and Hyp-Hyp$_2$ blocks and are therefore arabinosylated or remain nonglycosylated; 24 Hyp residues occur in Ser-Hyp-Ser-Hyp-Thr-Hyp sequences and are arabinogalactosylated with polysaccharide adducts; the remaining 21 Hyp residues occur in the blocks of tripeptidyl Pro/Hyp and, judging by earlier work (Shpak et al., 2001), should be arabinosylated and arabinogalactosylated in a ratio of ~2:1, respectively. This predicts that $(GA)_8$ contained 24 + 14, or 38 arabinogalactosylated Hyp residues and 24 + 7, or 31 arabinosylated and nonglycosylated Hyp residues, which is consistent with the empirically determined profile shown in Table IV.

Functional assay of the expressed gum showed that the viscosity of $(GA)_8$ was lower than that of native GAGP, presumably due to the smaller size of $(GA)_8$ compared to native GAGP, which was estimated to consist of at least

400 amino acids (Qi et al., 1991). The low viscosity of GAGP solutions at high gum concentration, a rare feature for gums, may be due to its fairly rigid, highly repetitive structure. Such regularity may enhance molecular packing and self-assembly and give rise to the unusually low viscosity of gum arabic (Goodrum et al., 2000).

We earlier speculated that the GAGP polysaccharide components were responsible for the emulsifying properties of GAGP, as the polypeptide backbone contains very few hydrophobic amino acids, whereas the polysaccharides contain hydrophobic features like terminal rhamnose residues and a β-1,3 galactan backbone (Goodrum et al., 2000). Thus, we predicted that $(GA)_8$ would mimic the emulsifying properties of GAGP. However, $(GA)_8$ was a poor emulsifier, discounting our suggestion that the hydrophobic glycans conferred emulsifying ability and the suggestion that the polypeptide, as represented by the GAGP consensus sequence, was sufficient to enable emulsification (Dickinson et al., 1990). Likewise, GFP was also a poor emulsifier, although it was considerably more hydrophobic than $(GA)_8$. In contrast, the fusion glycoprotein, $(GA)_8GFP$, exhibited an even higher emulsifying ability than native GAGP. This is consistent with other protein-polysaccharide conjugates, which are often excellent emulsifiers (Khan et al., 1999; Nakamura and Kato, 2000; Shu et al., 1998), presumably because the protein component adsorbs to the oil droplet while the H-bond forming polysaccharides interact with aqueous phase and stabilize the droplets through steric hindrance. In the case of $(GA)_8GFP$, however, we speculate that GFP probably interacted with the oil phase while the $(GA)_8$ glycoprotein domain interacted with the aqueous phase. Furthermore, the relatively low molecular mass of $(GA)_8GFP$ may allow it to diffuse more quickly than GAGP to the oil–water interface during homogenization, giving rise to its better emulsification properties than GAGP.

The molecular specifics that confer emulsifying properties to GAGP remain a mystery, probably at least until the complete amino acid sequence of the GAGP polypeptide is elucidated from a clone. The highly biased, simple amino acid composition of GAGP reflects the composition of the consensus sequence from which $(GA)_8$ was derived (Goodrum et al., 2000), arguing against the likelihood that the GAGP polypeptide per se adsorbs to oil droplets. However, as most AGPs are initially membrane-anchored by a glycosylphosphatidyl inositol anchor (Youl et al., 1998), it seems more feasible that lipid plays a role in GAGP-based emulsions.

CONCLUSION

We used a synthetic gene approach to design and produce in tobacco cells a novel recombinant plant gum, $(GA)_8$, that was hydroxylated and glycosylated very similar to native GAGP and exhibited low viscosity in aqueous solution. However, the recombinant gum did not emulsify orange oil due to lack of hydrophobic moiety. This work has demonstrated a unique avenue to the de novo design of glycoprotein-based biopolymers of controlled structure and with properties tailor-made for specific applications. Moreover, the growth of transgenic plant cells in bioreactors facilitates the production of gum with controlled purity and quality. Large-scale production of transgenic GAGP and eventually of other valuable "designer gums" in the future is foreseeable.

References

An G. 1985. High efficiency transformation of cultured tobacco cells. Plant Physiol 79:568–570.

Bergman T, Carlquist M, Jornvall H. 1986. Amino acid analysis by high performance liquid chromatography of phenylthiocarbamyl derivatives. In: Wittmann-Liebold B, editor. Advanced Methods in protein microsequence analysis. Berlin: Springer. p 45–55.

Bhatti T, Chambers RE, Clamp JR. 1970. The gas chromatographic properties of biologically important N-acetylglucosamine derivatives, monosaccharides, disaccharides, trisaccharides, tetrasaccharides and pentasaccharides. Biochim Biophys Acta 222:339–347.

Blumenkrantz N, Asboe-Hansen G. 1973. New methods for quantitative determination of uronic acids. Anal Biochem 54:484–489.

Choi SM, Lee OS, Kwon SY, Kwak SS, Yu DY, Lee HS. 2003. High expression of a human lactoferrin in transgenic tobacco cell cultures. Biotechnol Lett 25:213–218.

De Loose M, Gheysen G, Tire C, Gielen J, Villarroel R, Genetello C, Van Montagu M, Depicker A, Inze D. 1991. The extensin signal peptide allows secretion of a heterologous protein from protoplasts. Gene 99:95–100.

Dealwis C, Wall J. 2004. Towards understanding the structure-function relationship of human amyloid disease. Curr Drug Targets 5:159–171.

Dickinson E, Calazka VB, Anderson DMW. 1990. Emulsifying behavior of acacia gums. In: Phillips GO, Williams PA, Wedlock DJ, editors. Gums and stabilizers for the food industry, vol. 5. London: IRL Press. p 41–44.

Dische Z. 1962. Color reactions of carbohydrates. In: Whistler RL, Wolfrom RM, editors. Methods in carbohydrate chemistry. New York: Academic Press. p 475–514.

Doran PM. 2000. Foreign protein production in plant tissue cultures. Curr Opin Biotechnol 11:199–204.

Fischer R, Emans N, Schuster F, Hellwig S, Drossard J. 1999. Towards molecular farming in the future: using plant cell suspension cultures as bioreactors. Biotechnol Appl Biochem 30:109–112.

Garti N, Reichman D. 1993. Hydrocolloids as food emulsifiers and stabilizers. Food Struct 12:411–421.

Giddings G, Allison G, Brooks D, Carter A. 2000. Transgenic plants as factories for biopharmaceuticals. Nat Biotechnol 18:1151–1155.

Goodrum LJ, Patel A, Leykam JF, Kieliszewski MJ. 2000. Gum arabic glycoprotein contains glycomodules of both extensin and arabinogalactan-glycoproteins. Phytochemistry 54:99–106.

Hart G. 1992. Glycosylation. Curr Opin Cell Biol 4:1017–1023.

Holsters M, de Waele D, Depicker A, Messens E, van Montagu M, Schell J. 1978. Transfection and transformation of *Agrobacterium tumefaciens*. Mol Gen Genet 163:181–187.

Islam AM, Phillips GO, Sljivo A, Snowden MJ, William PA. 1997. A review of recent developments on the regulatory, structural and functional aspects of gum arabic. Food Hydrocoll 11:493–505.

Ituriaga G, Jefferson RA, Bevan M. 1989. Endoplasmic reticulum targeting and glycosylation of hybrid proteins in transgenic tobacco. Plant Cell 1:381–390.

Jentoft N. 1990. Why are proteins glycosylated? Trends Biochem Sci 15: 291–294.

Khan MAS, Babiker EFE, Azakami H, Kato A. 1999. Molecular mechanism of the excellent emulsifying properties of phosvitin-galactomannan conjugate. J Agric Food Chem 47:2262–2266.

Kieliszewski MJ, Lamport DTA. 1994. Extensin: repetitive motifs, functional sites, posttranslational codes and phylogeny. Plant J 5:157–172.

Kivirikko KI, Liesmaa M. 1959. A colorimetric method for determination of hydroxyproline in tissue hydrolysates. Scand J Clin Lab 11: 128–131.

Kusnadi AR, Nikolov ZL, Howard JA. 1997. Production of recombinant proteins in transgenic plants: practical considerations. Biotechnol Bioeng 56:473–484.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685.

Lamport DTA, Northcote DH. 1960. Hydroxyproline in primary cell walls of higher plants. Nature 188:665–666.

Lewis RV, Hinman M, Kothakota S, Fournier M. 1996. Expression and purification of a spider silk protein: a new strategy for producing repetitive proteins. Protein Expr Purif 7:400–406.

Liu S, Bugos RC, Dharmasiri N, Su WW. 2001. Green fluorescent protein as a secretory reporter and a tool for process optimization in transgenic plant cell cultures. J Biotechnol 87:1–16.

Matsumoto K, Doi Y. 2003. Biosynthesis of biodegradable polyester: current research and future direction. J Syn Org Chem Jpn 61:489–495.

Moire L, Rezzonico E, Poirier Y. 2003. Synthesis of novel biomaterials in plants. J Plant Physiol 160:831–839.

Murray E, Lotzer J, Eberle M. 1989. Codon usage in plant genes. Nucl Acids Res 17:477–498.

Nakamura A, Kato A. 2000. Multi-functional biopolymer prepared by covalent attachment of galactomannan to egg-white proteins through naturally occurring *Maillard* reaction. Nahrung 44:201–206.

Pearce KN, Kinsella JE. 1978. Emulsifying properties of proteins: evaluation of a turbidimetric technique. J Agric Food Chem 26:716–723.

Poirier Y. 1999. Production of new polymeric compounds in plants. Cure Opin Biotechnol 10:181–185.

Poirier Y, Nawrath C, Somerville C. 1995. Production of polyhydroxyalkanoates, a family of biodegradable plastics and elastomers, in bacteria and plants. Bio/Technol 12:142–150.

Qi W, Fong C, Lamport DTA. 1991. Gum arabic glycoprotein is a twisted hairy rope, a new model based on O-galactosyl hydroxyproline as the polysaccharide attachment site. Plant Physiol 96:848–855.

Ray AK, Bird PB, Lacobucci GA, Clark BC Jr. 1995. Functionality of gum arabic. Fractionation, characterization and evaluation of gum fractions in citrus oil emulsions and model beverages. Food Hydrocoll 9:123–131.

Scheller J, Guhrs KH, Grosse F, Conrad U. 2001. Production of spider silk proteins in tobacco and potato. Nat Biotechnol 19:573–577.

Schenk RU, Hildebrandt AC. 1972. Medium and techniques for induction and growth of monocotyledonous and dicotyledonous plant cell cultures. Can J Bot 50:199–204.

Serpe MD, Nothnagel EA. 1999. Arabinogalactan-proteins in the multiple domains of the plant cell surface. Adv Bot Res 30:207–289.

Sharp JM, Doran PM. 2001a. Characterization of monoclonal antibody fragments produced by plant cells. Biotechnol Bioeng 73:338–346.

Sharp JM, Doran PM. 2001b. Strategies for enhancing monoclonal antibody accumulation in plant cell and organ cultures. Biotechnol Prog 17:979–992.

Shin YJ, Hong SY, Kwon TH, Jang YS, Yang MS. 2003. High level expression of recombinant human granulocyte-macrophage colony stimulating factor in transgenic rice cell suspension culture. Biotechnol Bioeng 82:778–783.

Shpak E. 2001. Synthetic genes for the elucidation of hydroxyproline-O-glycosylation codes. PhD thesis, Athens, OH: Ohio University.

Shpak E, Leykam JF, Kieliszewski MJ. 1999. Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-o-glycosylation code. Proc Natl Acad Sci U S A 96:14736–14741.

Shpak E, Barbar E, Leykam JF, Kieliszewski MJ. 2001. Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in *Nicotiana tabacum*. J Biol Chem 276:11272–11278.

Showalter AM. 2001. Arabinogalactan-proteins: structure, expression and function. Cell Mol Life Sci 58:1399–1417.

Shu Y, Maki S, Nakamura S, Kato A. 1998. Double-glycosylated lysozyme at positions 19 and 49 constructed by genetical modification and its surface functional properties. J Agric Food Chem 46:2433–2438.

Smith ML, Mason HS, Shuler ML. 2002. Hepatitis B surface antigen (HbsAg) expression in plant cell culture: kinetics of antigen accumulation in batch culture and its intracellular form. Biotechnol Bioeng 80:812–822.

Stoner MR, Humphrey CA, Danielle JC, Shih NJR, McDonald KA, Jackman AP. 1997. Kinetics of growth and ribosome-inactivating protein production from *Trichosanthes kirilowii* plant cell cultures in a 5-L bioreactor. Biotechnol Prog 13:799–804.

Su WW, Guan P, Bugos RC. 2004. High-level secretion of functional green fluorescent protein from transgenic tobacco cell cultures: characterization and sensing. Biotechnol Bioeng 85:610–619.

Takeda T, Seki M, Furusaki S. 1994. Hydrodynamic damage of cultured cells of *Carthamus tinctorius* in a stirred tank reactor. J Chem Eng Jpn 27:466–471.

Tan L, Leykam JF, Kieliszewski MJ. 2003. Glycosylation motifs that direct arabinogalactan addition to arabinogalactan-proteins. Plant Physiol 132:1362–1369.

Ward WW. 1981. Properties of the coelenterate green-fluorescent protein. In: Deluca MA, McElroy WD, editors. Bioluminescence and chemiluminescence. New York: Academic Press. p 235–242.

Xu JF, Su ZG, Feng PS. 1998. Suspension culture of compact callus aggregate of *Rhodiola sachalinensis* for improved salidroside production. Enzyme Microb Technol 23:20–27.

Youl JJ, Bacic A, Oxley D. 1998. Arabinogalactan-proteins from *Nicotiana alata* and *Pyrus communis* contain glycosylphosphatidylinositol membrane anchors. Proc Natl Acad Sci U S A 95:7921–7926.

Zhao ZD, Tan L, Showalter AM, Lamport DTA, Kieliszewski MJ. 2002. Tomato LeAGP-1 arabinogalactan-protein purified from transgenic tobacco corroborates the Hyp contiguity hypothesis. Plant J 31: 431–444.