

O-Linked glycosylation in maize-expressed human IgA1

Anton S. Karnoup^{1,2}, Virgil Turkelson², and
W.H. Kerr Anderson³

²Analytical Sciences, The Dow Chemical Company, 1897 Building,
Midland, MI 48667; and ³Dow Biopharma, The Dow Chemical
Company, 1707 Building, Midland, MI 48674

Received on March 4, 2005; revised on May 9, 2005; accepted on
May 10, 2005

O-Linked glycans vary between eukaryotic cell types and play an important role in determining a glycoprotein's properties, including stability, target recognition, and potentially immunogenicity. We describe O-linked glycan structures of a recombinant human IgA1 (hIgA1) expressed in transgenic maize. Up to six proline/hydroxyproline conversions and variable amounts of arabinosylation (Pro/Hyp + Ara) were found in the hinge region of maize-expressed hIgA1 heavy chain (HC) by using a combination of matrix-assisted laser-desorption ionization mass spectrometry (MALDI MS), chromatography, and amino acid analysis. Approximately 90% of hIgA1 was modified in this way. An average molar ratio of six Ara units per molecule of hIgA1 was revealed. Substantial sequence similarity was identified between the HC hinge region of hIgA1 and regions of maize extensin-family of hydroxyproline-rich glycoproteins (HRGP). We propose that because of this sequence similarity, the HC hinge region of maize-expressed hIgA1 can become a substrate for posttranslational conversion of Pro to Hyp by maize prolyl-hydroxylase(s) with the subsequent arabinosylation of the Hyp residues by Hyp-glycosyltransferase(s) in the Golgi apparatus in maize endosperm tissue. The observation of up to six Pro/Hyp hydroxylations combined with extensive arabinosylation in the hIgA1 HC hinge region is well in agreement with the Pro/Hyp hydroxylation model and the Hyp contiguity hypothesis suggested earlier in literature for plant HRGP. For the first time, the extensin-like Hyp/Pro conversion and O-linked arabinosylation are described for a recombinant therapeutic protein expressed in transgenic plants. Our findings are of significance to the field of plant biotechnology and biopharmaceutical industry-developing transgenic plants as a platform for the production of recombinant therapeutic proteins.

Key words: HRGP/IgA1 antibody/maize/mass spectrometry/O-glycosylation

Introduction

The difficulty and cost of manufacturing glycosylated proteins for human therapeutic use has driven the evaluation of

multiple plant and plant cell systems as production vehicles (Giddings, 2001; Larrick *et al.*, 2001; Mayfield *et al.*, 2003; Rosin, 2004). These include transgenic plants, cell cultures, and nontransgenic plants which transiently express proteins following infection with recombinant plant viruses (Giddings, 2001). Plants of many types can assemble and accumulate valuable proteins that can be economically extracted and processed. They often show advantages over other expression systems as factories for the production of pharmaceutical and industrial proteins (Fischer *et al.*, 2004).

Antibodies are critical for the diagnosis, management, and treatment of contagious diseases and cancers (Stoger *et al.*, 2000; Weiner and Carter, 2003). Functional recombinant monoclonal antibodies (MAbs) have been successfully produced in plants including tobacco, rice, wheat, canola, maize, peas, soybean, potatoes, lettuce (Hiatt *et al.*, 1989, 1997, 2001; Ma *et al.*, 1994; Hein *et al.*, 1999, 2002; Stoger *et al.*, 2000; Giddings, 2001; Rosin, 2004), and alfalfa (*Medicago sativa*) (Bardor *et al.*, 2003a; Rosin, 2004). MAbs have also been produced in contained plant systems such as *Lemna* (duckweed) (Gasdaska *et al.*, 2003; Dickey *et al.*, 2004), a moss *Physcomitrella patens* (Wagner, 2003; Gorr *et al.*, 2004), and green alga *Chlamydomonas reinhardtii* (Mayfield *et al.*, 2003) and in suspension cultures of rice and tobacco cells (Hellwig *et al.*, 2004).

Antibodies of the IgA class have the ability to protect against various diseases at mucosal surfaces where they exist in the stable dimeric secretory IgA (sIgA) form (Larrick *et al.*, 2001). To date, there has been little success in the application of IgA class antibodies to therapeutic use because of the difficulty of producing the dimeric form in mammalian cells at economic levels (Chintalacharuvu and Morrison, 1997). This article describes, in part, the work done to express fully human IgA antibodies in transgenic maize for use in topical therapeutic applications.

N- and O-Linked glycosylation are important for biological activity and efficacy (Wright and Morrison, 1997; Bardor *et al.*, 2003a; Weiner and Carter, 2003), in vivo stability (Wright and Morrison, 1997), and maintaining correct tertiary structure (Weiner and Carter, 2003) of glycoproteins. It has also been acknowledged that modification of glycan structures may result in immunogenicity or allergenicity of a therapeutic molecule (Altmann *et al.*, 1993; Ma and Hein, 1995; Larrick *et al.*, 2001; Bardor *et al.*, 2003b). The impact of N-linked glycosylation processes on biomanufacturing is better understood because of the frequency with which this linkage is observed and relative ease of its characterization. (Rayon *et al.*, 1998; Wilson, 2002). N-Linked oligosaccharides can be found attached to Asn residues primarily in the Asn–Xxx–Ser/Thr consensus sequence (where Xxx is any amino acid except Pro). O-Linked glycosylation has been less studied, especially for plant glycoproteins,

¹To whom correspondence should be addressed; e-mail:
askarnoup@dow.com

primarily because of the absence of a common consensus sequence for O-linked glycosylation and characterization challenges (Fitchette-Laine *et al.*, 1998; Van den Steen *et al.*, 1998; Wilson, 2002).

O-Linked glycosylation has been detected in the Fc fragment of human IgA1 (*hIgA1*) molecules (Figure 1). The conserved heavy chain (HC) hinge region includes the following amino acid sequence PSTPPTSPSTPPTSPSPS (Kabat *et al.*, 1991), which is rich with proline (Pro), serine (Ser), and threonine (Thr) residues. Primarily, the O-linked glycans attached to the Thr and Ser residues were NeuAc α 2-3 β Gal β 1-3GalNAc trisaccharide and other neutral and sialylated O-glycans (bi-, tri-, and tetrasaccharides, containing galactose [Gal], *N*-acetylgalactosamine [GalNAc], and sialic acid [NeuAc]) (Mattu *et al.*, 1998). This is a typical mucin-type O-linked glycosylation (Van den Steen *et al.*, 1998). The glycosylation pattern in the *hIgA1* hinge region has been implicated in the pathology of many diseases including IgA neuropathy (reviewed in Mattu *et al.*, 1998). Based on molecular modeling, Mattu *et al.* (1998) suggested that *hIgA1* HC hinge region, because of its high Pro content, forms an extended rod-like conformation in solution and proposed that the heavy O-linked glycosylation of this region provides a continuous coat of glycan along the exposed surface which protects the region against bacterial proteolysis and degradation.

In plants, O-linked glycans are typically represented by Gal attached to Ser (solanaceous lectin-type O-linked glycosylation) and Hydroxyproline (Hyp) and chains of arabinose (Ara) attached to Hyp (linear extensin-type O-linked glycosylation); there is also arabinogalactan-type O-linked glycosylation, in which linear or branched oligosaccharides

consisting of Gal and Ara are attached to Hyp (Lamport, 1969; Lamport *et al.*, 1973; Pope, 1977; Van den Steen *et al.*, 1998; Wilson, 2002). O-Linked glycosylation in plants is believed to be involved in plant development and wound healing (Wilson, 2002).

In this study, we describe O-linked glycosylation of two maize-expressed monomeric *hIgA1* antibodies (and variations thereof purified from different genetic constructs): HX8 *hIgA1* (designated E1) targeted against glycoprotein-D of herpes simplex virus (HSV) (Zeitlin *et al.*, 1997) and H6-3C4 *hIgA1* (designated E2 or "sperm-immobilizing" antibody) targeted against SAGA-1 antigen (Anderson *et al.*, 1987; Briggs *et al.*, 2004; Diekman *et al.*, 2000). Amino acid sequences of the HCs of E1 and E2 *hIgA1* antibodies are shown in Figure 2A and B, respectively.

In this study, for the first time, the extensin-like Hyp/Pro conversion and O-linked arabinosylation are described for a recombinant therapeutic protein expressed in transgenic plants. Our findings may be of great importance to the field of plant biotechnology and to the biopharmaceutical industry developing transgenic plants as a platform for production of recombinant therapeutic proteins.

Results

Identification of modified hinge region fragment in maize-expressed *hIgA1*

The following experimental design was executed to confirm the identity of the HC hinge region fragment (HRF) in maize-expressed *hIgA1* (Figures 3, 4, and 5). Affinity-purified *hIgA1* was digested with trypsin after reduction of the disulfide bonds and subsequent carboxymethylation of the Cys side chains, and the tryptic peptides were fractionated by reversed-phase high-performance liquid chromatography (RP-HPLC). Figure 3 gives an example of the separation, showing only the region of the chromatogram where HRF species eluted. Fractions were collected and eluting peptides identified by matrix-assisted laser-desorption ionization mass spectrometry (MALDI MS). The fractions containing the HRF heterogeneity were pooled together for further analysis, and the glycosylation pattern was analyzed. An aliquot of the HRF-containing sample was treated with peptide *N*-glycosidase A (PNGase-A; Roche Diagnostics, Indianapolis, IN) and examined by MS. The absence of changes in MS ion pattern after PNGase-A treatment suggested that the glycosylation was not N-linked and therefore was likely to be O-linked (data not shown). To directly confirm identity of the HRF and determine the position of the attached glycans relative to a secondary protease (Asp-N) cleavage site, we treated the tryptic glycopeptide sample with endoprotease Asp-N and examined the resulting secondary proteolytic fragments by MS. It was expected that the "glycosylation" ion pattern would downshift and that the mass difference corresponding to the mass of the nonglycosylated secondary proteolytic fragment (Δ pep) would be different depending on the position of the attached glycan(s) relative to the secondary protease (Asp-N) cleavage site (due to the addition of either a proton or a hydroxyl group to the cleaved fragment during proteolysis) (Figures 4 and 5). The cleaved nonglycosylated portion of HRF was further

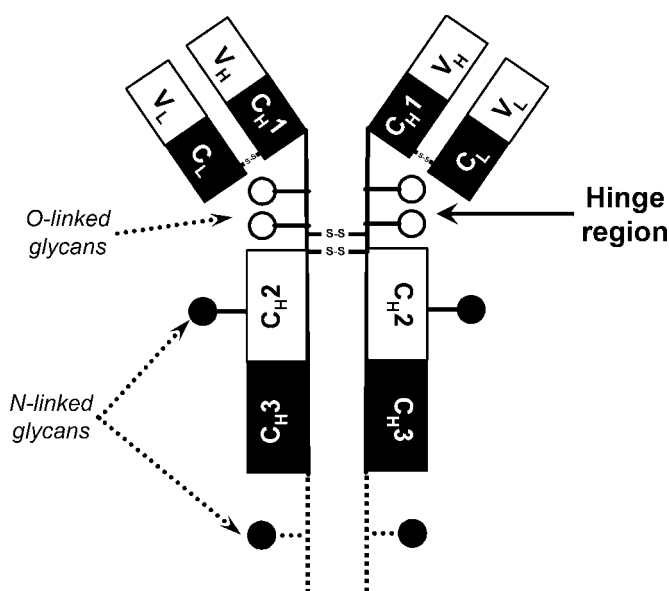


Fig. 1. A schematic representation of an assembled human IgA1 (*hIgA1*) molecule. Hinge region and sites of possible N- and O-linked glycosylation are indicated with arrows. The heavy chain (HC) tailpiece fragment (dotted lines) containing one possible N-linked glycosylation site was not experimentally detected in maize-expressed *hIgA1* samples that were investigated (data not shown).

A

(Pyro-*E*)-VQLVQSGA *E* VKKPGSSVK *V* SCKASGGSF *S* SYAINWVRQ *A* PGQG
 LEWMG *G* LMPIFGTTN *Y* AQKFQDRLT *I* TADVSTSTAY *M*QLSGLTYE *D* TAM
 YYCARV *A* YMLEPTVTA *G* GLDVWGQGT *L* VTVSSASPT *S* PKVFPLSLC *S* T
 QPDGNVVI *A* CLVQGFFPQ *E* PLSVTWSES *G* QGV TARNFP *P* SQDASGDLY *T*
 TSSQLTLPA *T* QCLAGKSVT *C* HVKHYTNPS *Q* DVTVPCPVP *S* TPPTSPST *P*
PTPSPSCCH *P* RLSLHRPAL *E* DLLLGSEAN *N* *L* TCTLTGLRD *A* SGVTFTWTP *S*
 SGKSAVQGP *P* ERDLGCGYS *V* SSVLPGCAE *P* WNHGKTFTC *T* AAYPESKTP
L TATLSKSGN *T* FRPEVHLLP *P* PSEELALNE *L* VTLTCLARG *F* SPKDVLRVW
L QGSQELPRE *K* YLTWASRQE *P* SQGTTTFAV *T* SILRVAEDW *W* KKGDTFSC
M *V* GHEALPLAF *T* QKTIDRLAG *K* PTHV *N* VSVV *M* AEVDGTC *Y*

Fragment	Amino acid residues	Fragment	Amino acid residues
H-T1	1-12	H-T16	303-325
H-T2	13-19	H-T17	326-337
H-T3	20-23	H-T18	338-346
H-T4	24-38	H-T19	347-378
H-T5	39-63	H-T20	379-383
H-T6	64-67	H-T21	384-388
H-T7	68-98	H-T22	389-398
H-T8	99-132	H-T23	399-400
H-T9	133-176	H-T24	401-407
H-T10	177-206	H-T25	408-424
H-T11	207-213	H-T26	425-431
H-T12	214-251	H-T27	432
H-T13	252-278	H-T28	433-452
H-T14	279-293	H-T29	453-456
H-T15	294-302	H-T30	457-478

B

(Pyro-*E*)-VQLQWAGAG LLKPSETLSL TCAVYGGGSF' GYYWSWIRQ' PGKGLE
 WIGE *I* **N**HSGSTNY *N* PSLRSRVITIS VDTSKNQFSZ KLRSVTAAD' AVYYCARG
 FM VRGIMWNYYY MDVWGKGTT' TVSSASPTSP KVFPLSLCST' QPDGNVVI
 A *C* LVQGFFPQE *P* LSVTWSESG *Q* GVTARNFPSP' QDASGDLYT' SSQLTLPAT
Q CLAGKSVTCH VKHYTNPSQD VTVPCVPST PPTSPSTP *P* TPSPSCCHP *R*
 LSLHRPALED LLLGSEAN *L* T CTLTGLRDA *S* GVTFTWTPS *S* GKSAVQGPPE
 RDLCGCGYSV *S* SVLPGCAEP *W* NHGKTFTCT *A* AYPESKTPLT' ATLSKSGNT' *F* R
 PEVHLLPP *P* SEELALNEL *V* TLTCLARGF *S* PKDVLRWL *Q* GSQELPREK *Y* LT
 WASRQEP *S* QGTTTFAVT *S* ILRVAEDWK KGDTFSCMV *G* HEALPLAFT *Q* K
 TIDRLAGK *P* THV *N* VSVVM *A* EVDGTC *Y*

Fig. 2. (A) Translated amino acid sequence of E1 human IgA1 (hIgA1) heavy chain (HC) from corresponding DNA vector. Hinge region (221–251) is underlined. Potential sites of N-glycosylation are highlighted in boldface type. A table of expected tryptic fragments is also shown in the inset, with the tryptic fragment corresponding to hinge region fragment (H-T12) highlighted. (B) Translated amino acid sequence of E2 hIgA1 HC from corresponding DNA vector. Hinge region (220–250) is underlined. Potential sites of N-glycosylation are highlighted in boldface type

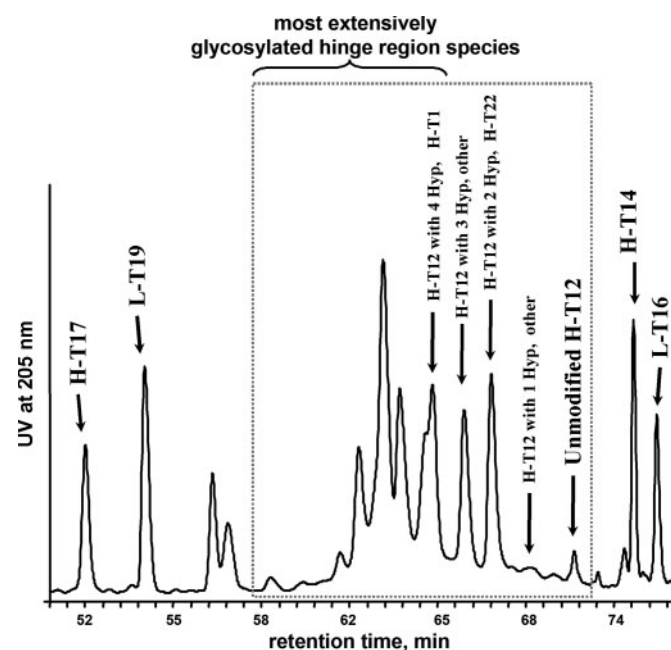


Fig. 3. Separation of tryptic digest of affinity-purified maize-expressed E1 human IgA1 (*hIgA1*) (reduced and carboxymethylated) on a reversed-phase high-performance liquid chromatography (RP-HPLC) column. Only a portion of the chromatogram in the proximity of the O-linked microheterogeneity (boxed region) is shown. Eluting peptides were identified with offline matrix-assisted laser-desorption/ionization mass spectrometry (MALDI MS). LC peak areas corresponding to peptides H-T17, H-T14, and unmodified H-T12 (hinge region fragment [HRF]) (corrected to the corresponding UV extinction coefficients of the peptides) were used to estimate amount of unmodified HRF relative to its modified species (Table I). Sequencing by MALDI postsource decay (MALDI PSD) was performed for unmodified HRF (H-T12) and HRF (H-T12) fragments containing one and two Hyp (Table III).

sequenced (by MALDI PSD [postsorce decay]) to directly confirm HRF identity (data not shown).

The mass-spectral pattern observed in the region of the chromatogram corresponding to elution of partial “O-linked” glycopeptide heterogeneity is shown in Figure 5A. Two main features were observed in these mass-spectral patterns (Figure 6): (1) clusters with 16-Da spacing between peaks consistent with Pro/Hyp heterogeneity of the HRF; and (2) 132-Da spacings between the ion clusters consistent with the presence of one or several pentoses (Pent), which could be due, according to our knowledge, to heterogeneous arabinosylation or xylosylation. Fractions eluted slightly later in the same chromatogram contained ions whose *m/z* were consistent with unmodified HRF and HRF species containing Pro/Hyp heterogeneity, but no modification with Pent (Figure 3). The results of an experiment using secondary digestion with Asp-N combined with sequencing of the cleaved N-terminal fragment of HRF by MALDI PSD confirmed the identity of the HRF as the modified fragment (Figure 5). Figure 6 shows MALDI MS corresponding to the combined “O-linked” HRF fractions, including unmodified HRF. Similar results were obtained for other genetic events of maize-expressed E1 and E2 *hIgA1* (A.S. Karnoup, unpublished data).

Detailed characterization of the O-linked glycosylation of maize-expressed hIgA1

The relative amounts of modified and unmodified HRF species were estimated by further characterization of the fractionated E1 *hIgA* tryptic digest (Figure 3). Assuming complete digestion with trypsin, 1 mol of *hIgA1* HC should give rise to 1 mol of each HC-derived tryptic fragment. The peak areas of two well-separated peptides (H-T14 and H-T17) in proximity to the H-T12 peptide and the peak area under the unmodified H-T12 peptide were normalized with respect to the corresponding calculated extinction coefficients

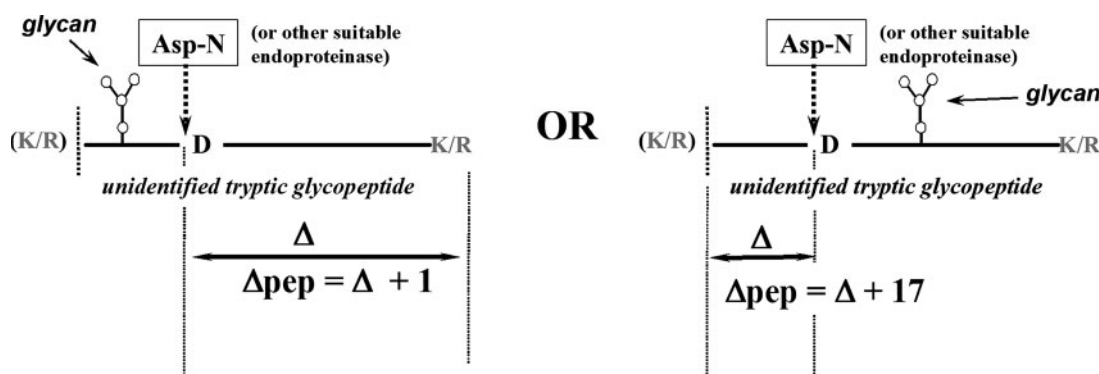


Fig. 4. Schematic summary of the approach used to identify site of O-linked glycosylation. “Fork-like” structure is a schematic representation of an oligosaccharide attached to the peptide. Δ is the mass difference between peaks of the same origin before and after secondary proteolytic cleavage. Δ_{pep} is the mass of the nonglycosylated secondary proteolytic fragment. The sample of isolated suspected hinge region fragment (HRF) is further treated with endoproteinase Asp-N, and the resulting secondary proteolytic fragments are analyzed by mass spectrometry. The observed fragments must be consistent with the rules of cleavage by trypsin and Asp-N: (1) amino acids on the C-terminus of the tryptic peptide of interest, and the peptide immediately preceding it must be Lys (K) or Arg (R); and (2) Asp-N cleaves before Asp (D). The mass-spectral pattern corresponding to glycosylation heterogeneity is expected to downshift after secondary proteolytic cleavage. The value of the mass shift (Δ) should correspond to the portion of the peptide being removed. The mass of the nonglycosylated secondary proteolytic fragment (Δ_{pep}) will be different depending on the position of the attached glycan(s) relative to the secondary protease (Asp-N) cleavage site (due to addition of either a proton or a hydroxyl group to the cleaved fragment during proteolysis). See Figure 5 for illustration of this approach.

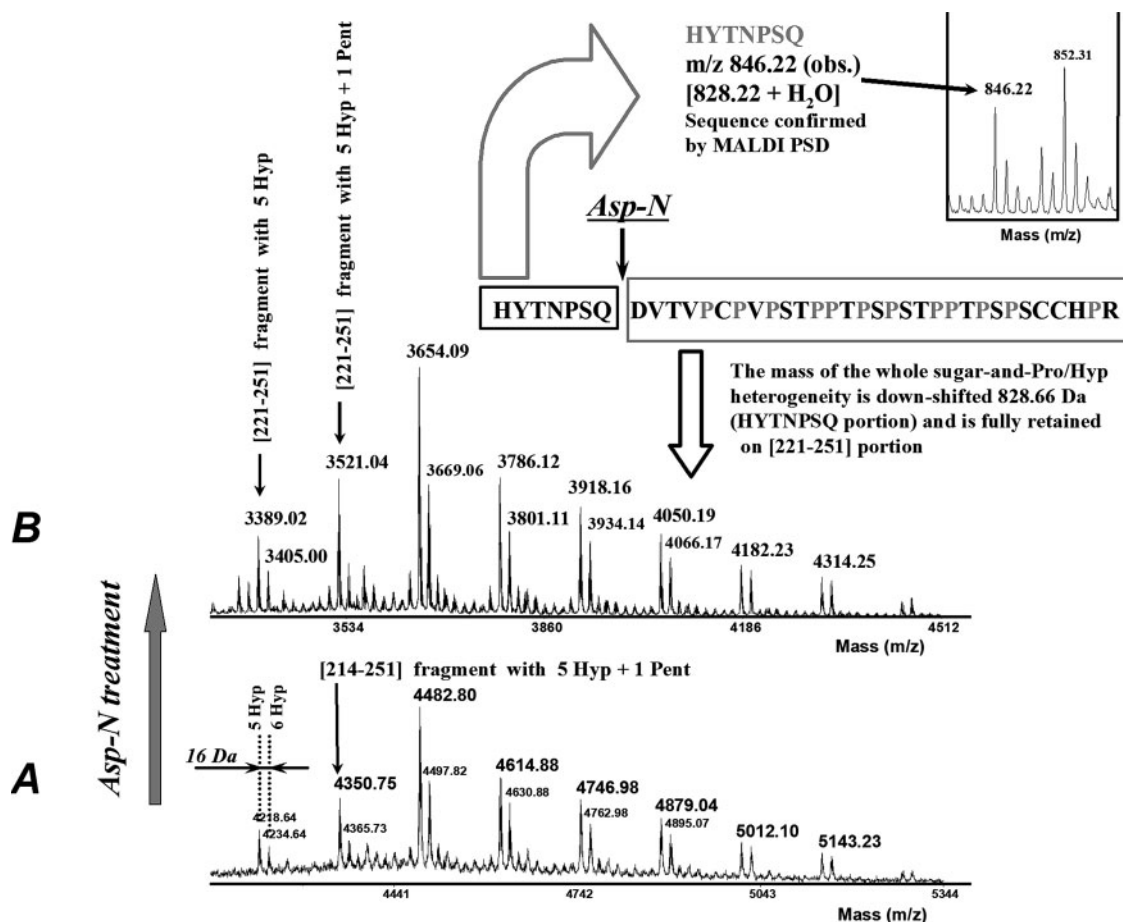


Fig. 5. Identification of O-linked glycosylation in E1 human IgA1 (hIgA1) heavy chain hinge region fragment (HRF). The approach outlined in Figure 4 was used. (A) Portion of partially separated O-linked microheterogeneity observed in tryptic digest of maize-expressed E1 hIgA1. Clusters with 16-Da spacings between peaks are consistent with Pro/Hyp microheterogeneity of the HRF. The 132-Da spacings between the ion clusters are consistent with the attachment of pentose (Pent) sugar residues to the HRF fragment. Fractions eluted later in the chromatogram (not shown) contained ions with *m/z* consistent with unmodified [214–251] fragment (H-T12 peptide) and [214–251] fragment containing Pro/Hyp microheterogeneity, but no Pent attachments. See text for details. (B) Sample shown in (A), treated with endoproteinase Asp-N. The pattern shown in (B) is consistent with the [221–251] fragment, which results from Asn-N cleavage of [214–251] fragment and which contains the same Pro/Hyp + Pent heterogeneity as in (A). Sequence of the “clipped” HYTNPSQ portion of the [214–251] fragment (inset) was confirmed by matrix-assisted laser-desorption/ionization postsource decay (MALDI PSD) (data not shown). Thus, the isolated fragment was identified as the HRF of hIgA1. Further experiments identified the modifying Pent sugar as arabinose (see Figure 7).

at 215 nm (detection wavelength) and used to estimate relative amount of unmodified H-T12 peptide. Approximate values of the extinction coefficients at 215 nm for individual peptides were calculated using the following relationship:

$$\epsilon_{215}^{\text{peptide}} \approx a\epsilon_{215}^{\text{Trp}} + b\epsilon_{215}^{\text{Tyr}} + c\epsilon_{215}^{\text{Phe}} + d\epsilon_{215}^{\text{His}} + e\epsilon_{215}^{\text{pept.bond}} \quad (1)$$

where $\epsilon_{215}^{\text{peptide}}$ is the estimated extinction coefficient of an individual peptide at 215 nm, $\epsilon_{215}^{\text{Trp}}$ is the extinction coefficient of tryptophan at 215 nm, $\epsilon_{215}^{\text{Tyr}}$ is the extinction coefficient of tyrosine at 215 nm, $\epsilon_{215}^{\text{Phe}}$ is the extinction coefficient of phenylalanine at 215 nm, $\epsilon_{215}^{\text{His}}$ is the extinction coefficient of histidine at 215 nm, $\epsilon_{215}^{\text{pept.bond}}$ is the extinction coefficient of a peptide bond at 215 nm, *a* is the number of tryptophan residues in a peptide, *b* is the number of tyrosine residues in a peptide, *c* is the number of phenylalanine res-

idues in a peptide, *d* is the number of histidine residues in a peptide, and *e* is the number of peptide bonds in a peptide. The following approximate values for extinction coefficients of individual amino acids were used: $\epsilon_{215}^{\text{Trp}} = 47,000 \text{ cm}^{-1} \text{ M}^{-1}$, $\epsilon_{215}^{\text{Tyr}} = 10,000 \text{ cm}^{-1} \text{ M}^{-1}$, $\epsilon_{215}^{\text{Phe}} = 10,000 \text{ cm}^{-1} \text{ M}^{-1}$, $\epsilon_{215}^{\text{His}} = 5900 \text{ cm}^{-1} \text{ M}^{-1}$, $\epsilon_{215}^{\text{pept.bond}} = 7000 \text{ cm}^{-1} \text{ M}^{-1}$ (Simons, 1979). Contributions from other chromophores to absorption at 215 nm were assumed negligible (Simons, 1979). The correction of the peak areas corresponding to H-T14, H-T17, and unmodified H-T12 peptides in Figure 3 is summarized in Table I. Similar values of the normalized peak areas were obtained for H-T14 and H-T17 peptides, confirming the validity of the estimation.

Approximately 9% of the HRF (H-T12 peptide) is unmodified implying that ~91% of the HRF was modified by either Pro/Hyp conversion(s) or Pro/Hyp + glycosylation modifications.

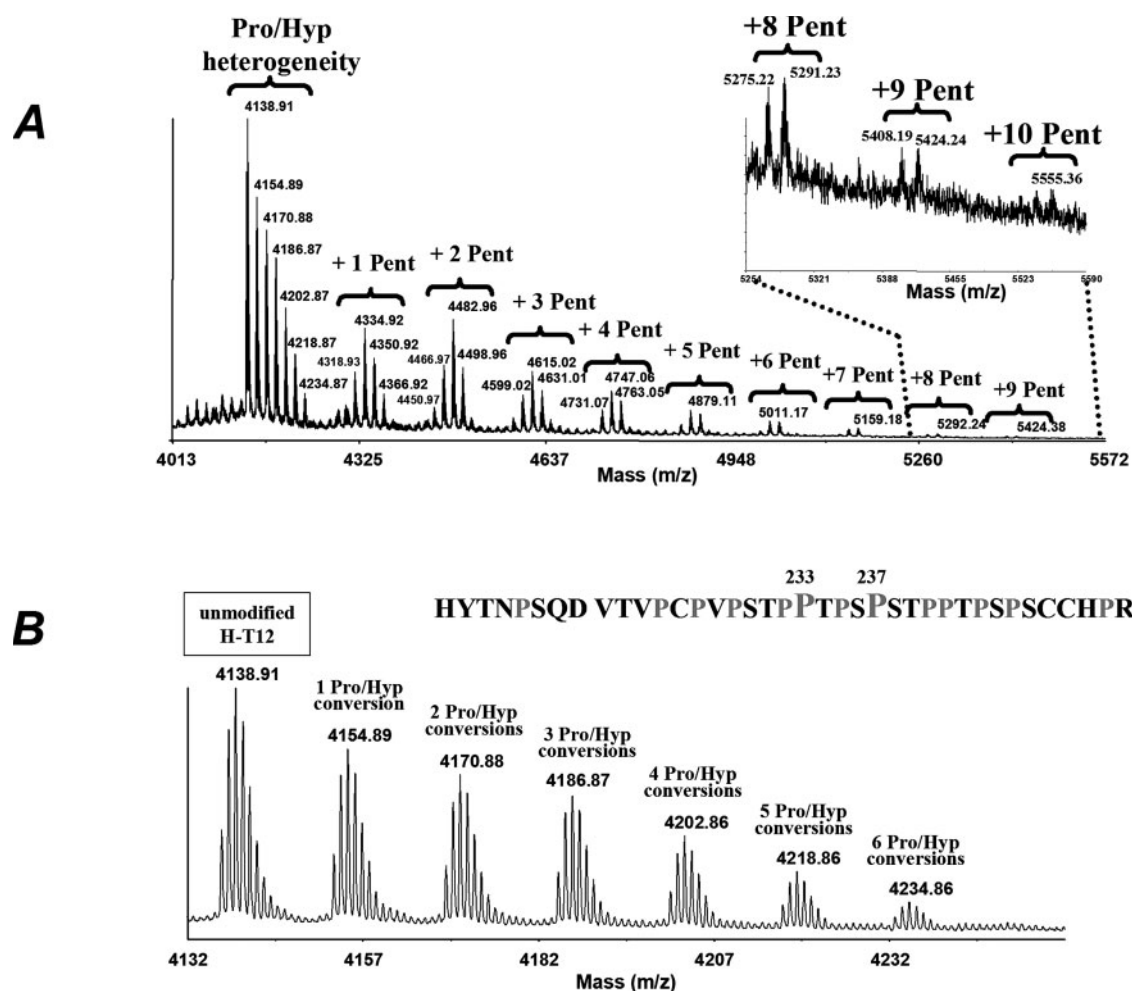


Fig. 6. Matrix-assisted laser-desorption ionization mass spectrometry (MALDI MS) profiling of the pooled hinge region fragment (HRF) heterogeneity. (A) The heterogeneity of the HRF (H-T12 peptide of E1 human IgA1 (*h*IgA1), underlined in Figure 2A) as it appears in MALDI MS of the pooled “O-linked” high-performance liquid chromatography fraction (boxed region in Figure 3). Ion clusters differing from the first one by $n \times 132$ Da may be assigned to Pro/Hyp + pentose heterogeneity, with number of pentose (Pent) units running from $n = 1$ up to 10. Further experiments identified the Pent sugar as arabinose (see Figure 7). (B) The first cluster of ions in (A) may be assigned to Pro/Hyp heterogeneity of the HRF (H-T12 peptide) without glycosylation.

The identity and quantity of the sugar that was attached to HRF species was determined by high performance anion-exchange with pulsed amperometric detection (HPAE/PAD) after release by acid hydrolysis of monosaccharides from peptides contained in the “O-linked” HRF RP-HPLC fraction (Figure 7). A reference standard was injected before each sample injection to maximize analysis accuracy. Quantitation of the Ara peaks observed in several independent injections yielded a value of 2.5 ± 0.2 μ g Ara for the “O-linked” HRF fraction hydrolysate. The chromatogram in Figure 7A also shows peaks in the regions where Gal, glucose (Glc), and xylose (Xyl) are known to elute. The peak corresponding to Xyl is negligible (this peak calculated to ~ 0.04 μ g Xyl, which was within the error of detection). These results confirm the identity of the modifying sugar as Ara. The minor peaks eluting in the “Gal” and “Glc” regions of the sample chromatograms (Figure 7A) are probably artifacts, because no hexoses (Gal or Glc) were detected among the HRF modifications by MS. HPAE/PAD gave an estimate

of ~ 2.5 μ g of Ara per amount of sample resulted from ~ 420 μ g of initial E1 *h*IgA1. Therefore, an estimated average molar ratio of at least six Ara units per each molecule of E1 *h*IgA1 was determined.

The heterogeneity of the HRF (H-T12 fragment) species as it appears in the MALDI mass spectrum of the pooled “O-linked” fraction (boxed region in Figure 3) is presented in Figure 6. In Figure 6A, the first cluster of ions was assigned to Pro/Hyp heterogeneity without glycosylation, and ion clusters differing from the first one by $n \times 132$ Da were assigned to Pro/Hyp + Ara heterogeneity, with the number of Ara units running from $n = 1$ up to 10 (see also Figure 5 and Table II). In Figure 6B, a close-up on the ion cluster assigned to Pro/Hyp heterogeneity without glycosylation shows up to six Pro/Hyp conversions. The first three peptides in this heterogeneity (unmodified HRF [H-T12] and HRF with one and two Pro/Hyp conversions) were sequenced by using MALD-PSD, and the first two Pro/Hyp conversions were assigned to Pro237 and Pro233

Table I. Estimate of relative quantity of unmodified hinge region fragment (H-T12 peptide of E1 human IgA1 HC) by reversed-phase high-performance liquid chromatography (RP-HPLC) of tryptic digest (Figure 3)

Tryptic fragment	Sequence	Coefficients in equation 1 ^b	Calculated $\epsilon_{215}^{\text{peptide}}$ ($\text{cm}^{-1} \text{M}^{-1}$)	Peak area, A (Arb unit)	Normalized peak area (calculated) (Arb unit), $N = A/\epsilon_{215}^{\text{peptide}}$	% of expected amount ^a
H-T14	DASGVTFWTWPSSGK	$a = 1$ $b = 0$ $c = 1$ $d = 0$ $e = 14$	155,000	1752208	11.30	112
H-T17	TFTCTAAYPESK	$a = 0$ $b = 1$ $c = 1$ $d = 0$ $e = 11$	97,000	857934	8.85	88
H-T12 (hinge region fragment) (unmodified)	HYTNPSQDVTVPVPCVPSTPPTPSPSTPPTPSPSCCHPR	$a = 0$ $b = 1$ $c = 0$ $d = 1$ $e = 37$	274,900	254180	0.92	9

^aAverage $N = (11.30 + 8.85)/2 = 10.08$ equivalent to 100% of expected amount was used in the calculation.

^bSee text for details.

residues in the E1 hIgA1 HC sequence (Table III). MALDI PSD fragmentation pattern of two other peptides (m/z 4186.9 and 4202.9, Figure 6B) from the HRF heterogeneity was also consistent with the HRF amino acid sequence, but, due to insufficient signal intensities, for these HRF variants, it was not possible to unambiguously identify positions of further Pro/Hyp modifications (data not shown).

In addition, the presence of Hyp in maize-expressed hIgA1 HC was determined by amino acid analysis (Table IV). The average number of Hyp residues per E1 hIgA1 HC molecule was between two and three.

To obtain more detail on the distribution of the modified species within the heterogeneous pool of the HRF, by further purification with hydrophilic interaction chromatography (HILIC), we removed the peptides that coeluted with the HRF species during the first pass of RP-HPLC performed on the E1 hIgA1 tryptic digest (Figure 8A). HILIC (Alpert, 1990; Zhang and Wang, 1998) was chosen as a second-dimension technique because of the polar nature of the HRF and its glycosylated variants. Fractions were collected, desalted, and analyzed by MALDI MS to identify eluting peaks. Fractions containing HRF species were pooled, concentrated, and reinjected onto the RP-HPLC column to obtain additional information on the distribution of the modified HRF species. Figure 8B shows the second-pass RP-HPLC on the purified HRF heterogeneity. Fractions were collected and analyzed by MALDI MS to identify eluting peaks. The semiquantitative results are summarized in the inset in Figure 8B (also see Table V). The most abundant HRF species were found to be H-T12

peptides containing an average of five Hyp and extensive arabinosylation (up to 10 Ara) and an average of four Hyp and less extensive arabinosylation (up to six Ara).

Estimates of relative abundance of various HRF species in the HRF heterogeneity made by measuring relative intensities of HRF-related peaks in the MALDI MS yielded results comparable to those obtained by RP-HPLC and RP-HPLC/HILIC/RP-HPLC methods (Tables II and V).

The abundance of the unmodified HRF species in maize-expressed E2-480 hIgA1 and E2-540 hIgA1 was ~10% of the total HRF heterogeneity, based on signal intensity in the MALDI MS profiles (Table II). This is similar to results of maize-expressed E1 hIgA1. The pattern of O-linked arabinosylation in both E2 hIgA1 samples was essentially identical to that described for E1 hIgA1.

A BLAST database search performed on the PSTPPTPSP-STPPTPSP sequence (Pro-rich hinge region of the hIgA1 HC; a major part of the HRF) yielded multiple matches with proline-/hydroxyproline-rich glycoproteins (HRGP) from plants as well as the expected match with IgA HC. Among them, there were matches to maize extensins and cell-wall proteins that are believed to undergo conversion of Pro to Hyp in the Pro-rich repeat regions, with subsequent arabinosylation of Hyp residues (e.g., see Stiefel *et al.*, 1988).

Discussion

Glycosylation of several maize-expressed transgenic hIgA1 samples was investigated. In this study, we focused on

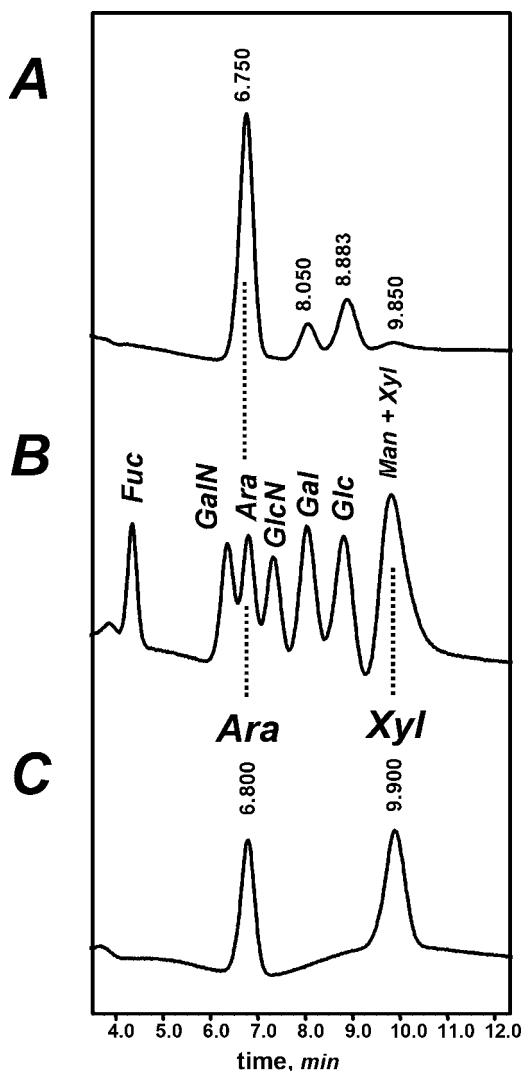


Fig. 7. Identification of the modifying sugar attached to the hinge region fragment (HRF). (A) High performance anion-exchange with pulsed amperometric detection (HPAE/PAD) chromatogram of trifluoroacetic acid-hydrolyzate of the isolated HRF (H-T12 peptide of E1 human IgA1 [*h*IgA1] heavy chain [HC]). The modifying pentose (Pent) sugar was identified as arabinose (Ara) by the HPAE/PAD analysis. (B) HPAE/PAD chromatogram of the mixture of the monosaccharide standards. (C) HPAE/PAD chromatogram of Ara and Xyl standards.

describing the O-linked glycosylation of the transgenic maize-expressed *h*IgA1. The N-linked glycosylation of the maize-expressed *h*IgA1 analyzed was typical of a maize-expressed protein. The most abundant N-glycan was truncated paucimannose type, (GlcNAc)₂(Man)₃Xyl (A.S. Karnoup, unpublished results).

Pro/Hyp + Ara heterogeneity in the HC hinge region of maize-expressed *h*IgA1 was discovered using a combination of MALDI MS, chromatography, and amino acid analysis. Up to six Pro/Hyp conversions were detected in the HC HRF of maize-expressed *h*IgA1. The first two Pro/Hyp conversions were assigned to Pro237 and Pro233 residues. The identity of Ara as a modifying sugar unit was confirmed by HPAE/PAD. No Xyl or Gal was found attached

via O-linkage to the HRF of maize-expressed *h*IgA1 HC. Only ~10% of total E1 *h*IgA1 species were found to be free of Pro/Hyp and/or Pro/Hyp + Ara modifications in the HRF region; a major fraction of the modified E1 *h*IgA1 was found to contain Ara.

O-Linked glycosylation of many proteins, including those of mammalian origin, is often associated with attachment of single sugar residues (typically Gal, GalNAc, or *N*-acetylglucosamine [GlcNAc]) or oligosaccharides (typically containing Gal, GalNAc, and GlcNAc) to Ser and Thr amino acid residues of the protein (Van den Steen *et al.*, 1998). Our data did not indicate such involvement of Ser or Thr residues in the O-linked glycosylation of maize-expressed *h*IgA1 HRF: (1) MALDI PSD of HRF species and amino acid analysis of *h*IgA1 HC did not indicate modifications of Ser and Thr; (2) no evidence for the presence of hexose (Hex) (162 Da) or *N*-acetylhexosamine (HexNAc) (203 Da) (e.g., Gal, GalNAc, or GlcNAc) sugar residues attached to HRF was found in MALDI MS profiles of HRF; and (3) no appreciable amount of Gal, GlcNAc, or any other sugar was detected by monosaccharide analysis of isolated HRF.

Strong sequence similarity was found between the HC hinge region of *h*IgA1 and maize extensin-family HRGP (Figure 9B and references therein). In light of published literature on plant HRGP, we propose the following scheme for the possible pathway for posttranslational “O-linked” modification of maize-expressed *h*IgA1 (Figure 9A). We suggest that because of the sequence similarity to maize threonine/hydroxyproline-rich glycoproteins (THRGP) and extensins, the HC hinge region (HRF) of maize-expressed *h*IgA1 becomes a substrate for posttranslational conversion of Pro to Hyp by maize prolyl-hydroxylase(s) with the subsequent arabinosylation of the Hyp residues by Hyp-glycosyltransferase(s) in the endoplasmic reticulum and/or the Golgi apparatus in maize endosperm tissue. Hydroxylation of Pro (P) to Hyp (O) is common for plant proline-rich proteins and is a sequence-specific event that typically occurs in characteristic repetitive Pro-rich structural motifs such as -XPPX- (arabinogalactan motif), -SP_{*n*}- (typically *n* = 2–4), and variants of -XPPXX- (Shpak *et al.*, 2001). The previously reported examples of repeat sequences undergoing Pro/Hyp hydroxylations in plants, such as a maize extensin repeat -SPKPPTP- (Showalter, 1993) and a maize HRGP repeat -TPSPKPPTPKPTPPTY- (Stiefel *et al.*, 1990), bear a striking similarity to the *h*IgA1 HRF subsequence -PSTPPTPSPSTPPTPSP-, in which the motif -STPPTPSP- is repeated twice (Figure 9B). Previously identified peptide sequences and corresponding Hyp-glycoside profiles of selected HRGP indicated that arabinosylation is correlated with Hyp contiguity (Kieliszewski and Lamport, 1994; Kieliszewski, 2001; Shpak *et al.*, 2001). The Hyp contiguity hypothesis (Kieliszewski and Lamport, 1994) that predicts preferential arabinosylation of contiguous Hyp residues (where contiguity begins with dipeptidyl Hyp, such as in a sequence -X-Hyp_{*n*}-, where *n* ≥ 2) and preferential galactosylation and often attachment of large arabinogalactan polysaccharides to clustered noncontiguous Hyp residues (such as in a sequence -X-Hyp-X-Hyp-X-) was largely confirmed by experiments with natural glycopeptides from plant cell wall glycoproteins (Kieliszewski *et al.*, 1995) and synthetic polypeptides that were recombinantly expressed in plants (Kieliszewski, 2001;

Table II. Summary of O-linked modifications observed in maize-expressed human IgA1 heavy chain

Glycopeptide (O-linked) ^a	<i>m/z</i> theoretical	Genetic construct					
		E1 (HX8) IgA1 (pooled)		E2-480 IgA1		E2-540 IgA1	
		<i>m/z</i> observed	% abundance (MALDI)	<i>m/z</i> observed	% abundance (MALDI)	<i>m/z</i> observed	% abundance (MALDI)
Unmodified	4139.84	4138.91	15.0	4139.57	20.3	4140.05	5.8
1 Hyp	4155.84	4154.89	10.9	4155.36	10.1	4155.51	7.0
2 Hyp	4171.83	4170.88	9.7	4171.36	7.4	4171.53	5.5
3 Hyp	4187.83	4186.87	8.2	4187.33	7.8	4188.09	5.8
4 Hyp	4203.82	4202.86	5.8	4203.33	5.8	4203.57	5.2
5 Hyp	4219.82	4218.86	3.4	4219.22	4.1	4219.58	4.4
6 Hyp	4235.81	4234.86	1.7	4235.33	1.4	4235.59	1.4
1 Hyp + 1 Ara	4287.88	4286.90	0.6	ND	0.0	4288.66	Trace
2 Hyp + 1 Ara	4303.87	4302.90	0.9	4303.40	0.7	4303.20	1.1
3 Hyp + 1 Ara	4319.87	4318.93	2.6	4319.42	2.4	4319.64	2.3
4 Hyp + 1 Ara	4335.86	4334.92	4.9	4335.41	4.1	4335.65	5.2
5 Hyp + 1 Ara	4351.86	4350.92	3.4	4351.41	2.7	4351.71	3.4
6 Hyp + 1 Ara	4367.85	4366.93	1.5	4367.42	1.4	4367.69	1.9
1 Hyp + 2 Ara	4419.92	ND	0.0	ND	0.0	ND	0.0
2 Hyp + 2 Ara	4435.91	ND	0.0	ND	0.0	ND	0.0
3 Hyp + 2 Ara	4451.91	4450.97	0.9	4451.50	0.7	4452.76	1.2
4 Hyp + 2 Ara	4467.90	4466.96	3.2	4467.47	3.0	4467.72	3.8
5 Hyp + 2 Ara	4483.90	4482.96	5.4	4483.47	4.1	4483.81	6.3
6 Hyp + 2 Ara	4499.89	4498.96	3.0	4499.47	2.4	4499.83	3.3
7 Hyp + 2 Ara	4515.89	4514.97	0.4	4515.69	0.5	4515.81	1.2
1 Hyp + 3 Ara	4551.96	ND	0.0	ND	0.0	ND	0.0
2 Hyp + 3 Ara	4567.95	ND	0.0	ND	0.0	ND	0.0
3 Hyp + 3 Ara	4583.95	4583.02	0.4	4583.49	0.3	ND	0.0
4 Hyp + 3 Ara	4599.94	4599.02	1.7	4599.53	2.4	4599.79	3.8
5 Hyp + 3 Ara	4615.94	4615.02	3.0	4615.55	3.0	4615.93	5.1
6 Hyp + 3 Ara	4631.93	4631.01	1.9	4631.53	1.7	4631.92	3.0
7 Hyp + 3 Ara	4647.93	4647.02	0.4	4646.75	1.0	4647.99	1.0
8 Hyp + 3 Ara	4663.92	4663.01	0.2	4662.70	0.3	4663.88	0.8
1 Hyp + 4 Ara	4684.00	ND	0.0	ND	0.0	ND	0.0
2 Hyp + 4 Ara	4699.99	ND	0.0	ND	0.0	ND	0.0
3 Hyp + 4 Ara	4715.99	ND	0.0	ND	0.0	ND	0.0
4 Hyp + 4 Ara	4731.98	4731.07	1.1	4731.61	1.7	4731.84	2.3
5 Hyp + 4 Ara	4747.98	4747.06	2.1	4747.60	2.4	4748.03	3.7
6 Hyp + 4 Ara	4763.97	4763.05	1.5	4763.58	1.4	4764.05	2.5
7 Hyp + 4 Ara	4779.97	4779.09	0.2	4778.82	0.3	4780.08	1.1
8 Hyp + 4 Ara	4795.96	ND	0.0	4794.80	Trace	4796.04	Trace
1 Hyp + 5 Ara	4816.04	ND	0.0	ND	0.0	ND	0.0
2 Hyp + 5 Ara	4832.03	ND	0.0	ND	0.0	ND	0.0
3 Hyp + 5 Ara	4848.03	ND	0.0	ND	0.0	ND	0.0
4 Hyp + 5 Ara	4864.02	4864.11	0.4	4863.84	0.7	4863.93	0.8
5 Hyp + 5 Ara	4880.02	4880.11	1.1	4879.86	1.7	4880.13	2.7
6 Hyp + 5 Ara	4896.01	4896.11	0.9	4895.89	0.7	4896.14	1.4
7 Hyp + 5 Ara	4912.01	4912.10	0.2	4911.88	Trace	4912.11	0.4

Table II. Continued

Glycopeptide (O-linked) ^a	<i>m/z</i> theoretical	Genetic construct					
		E1 (HX8) IgA1 (pooled)		E2-480 IgA1		E2-540 IgA1	
		<i>m/z</i> observed	% abundance (MALDI)	<i>m/z</i> observed	% abundance (MALDI)	<i>m/z</i> observed	% abundance (MALDI)
1 Hyp + 6 Ara	4948.08	ND	0.0	ND	0.0	ND	0.0
2 Hyp + 6 Ara	4964.07	ND	0.0	ND	0.0	ND	0.0
3 Hyp + 6 Ara	4980.07	ND	0.0	ND	0.0	ND	0.0
4 Hyp + 6 Ara	4996.06	4996.18	0.2	4995.72	Trace	4996.00	0.5
5 Hyp + 6 Ara	5012.06	5011.17	0.9	5011.90	1.0	5012.27	1.4
6 Hyp + 6 Ara	5028.05	5027.17	0.6	5027.99	0.7	5028.20	1.1
7 Hyp + 6 Ara	5044.05	5043.17	0.2	5044.02	0.3	ND	0.0
1 Hyp + 7 Ara	5080.12	ND	0.0	ND	0.0	ND	0.0
2 Hyp + 7 Ara	5096.11	ND	0.0	ND	0.0	ND	0.0
3 Hyp + 7 Ara	5112.11	ND	0.0	ND	0.0	ND	0.0
4 Hyp + 7 Ara	5128.10	ND	0.0	ND	0.0	5129.07	Trace
5 Hyp + 7 Ara	5144.10	5143.18	0.4	5143.99	0.9	5144.33	1.1
6 Hyp + 7 Ara	5160.09	5159.18	0.4	5159.96	0.5	5160.30	0.8
1 Hyp + 8 Ara	5212.16	ND	0.0	ND	0.0	ND	0.0
2 Hyp + 8 Ara	5228.15	ND	0.0	ND	0.0	ND	0.0
3 Hyp + 8 Ara	5244.15	ND	0.0	ND	0.0	ND	0.0
4 Hyp + 8 Ara	5260.14	ND	0.0	ND	0.0	ND	0.0
5 Hyp + 8 Ara	5276.14	5276.22	0.2	5275.89	0.3	5276.32	1.0
6 Hyp + 8 Ara	5292.13	5292.23	0.2	ND	0.0	5292.31	0.7
1 Hyp + 9 Ara	5344.20	ND	0.0	ND	0.0	ND	0.0
2 Hyp + 9 Ara	5360.19	ND	0.0	ND	0.0	ND	0.0
3 Hyp + 9 Ara	5376.19	ND	0.0	ND	0.0	ND	0.0
4 Hyp + 9 Ara	5392.18	ND	0.0	ND	0.0	ND	0.0
5 Hyp + 9 Ara	5408.18	5408.19	Trace	ND	0.0	ND	0.0
6 Hyp + 9 Ara	5424.13	5424.24	Trace	ND	0.0	ND	0.0
1 Hyp + 10 Ara	5476.24	ND	0.0	ND	0.0	ND	0.0
2 Hyp + 10 Ara	5492.23	ND	0.0	ND	0.0	ND	0.0
3 Hyp + 10 Ara	5508.23	ND	0.0	ND	0.0	ND	0.0
4 Hyp + 10 Ara	5524.22	ND	0.0	ND	0.0	ND	0.0
5 Hyp + 10 Ara	5540.22	5539.36	Trace	ND	0.0	ND	0.0
6 Hyp + 10 Ara	5556.17	5555.36	Trace	ND	0.0	ND	0.0

MALDI, matrix-assisted laser-desorption ionization; ND, not detected.

^aHinge region fragment DVTVPVCPVPSTPPTPSPSTPPTPSPSCCHPR with Pro converted to Hyp and arabinosylated.

Shpak *et al.*, 2001; Tan *et al.*, 2003). Interesting detail relating to our research was demonstrated by Shpak *et al.* using synthetic polypeptides recombinantly expressed in *Nicotiana tabacum*: ~45% Gal and ~28% Ara were attached *in vivo* to the repeating sequence -Ser-Hyp-Hyp-, whereas the repeating sequence -Ser-Hyp-Hyp- was modified with 100% Ara (Shpak *et al.*, 2001). Although clustered noncontiguous Hyp residues in the repeating motifs of plant HRGP, extensins, and arabinogalactan proteins (AGPs) were shown to have a lesser probability of being arabinosylated, the extent of their arabinosylation may still be as significant as ~28% (Kieliszewski *et al.*, 1995; Shpak *et al.*, 2001). The

HRF of *h*IgA1 HC contains four contiguous Pro/Hyp residues and seven clustered noncontiguous Pro/Hyp residues (Figures 6 and 9). Therefore, our observation of up to six Pro/Hyp hydroxylations combined with extensive arabinosylation in the *h*IgA1 HC hinge region is well in agreement with the Pro/Hyp hydroxylation model and the Hyp contiguity hypothesis described above. We assigned the first two Pro/Hyp conversions to Pro237 and Pro233 residues. At this moment, it is not clear which Pro residues at which other positions in the HRF amino acid sequence are also converted to Hyp, and further work is needed to determine specific positions of these hydroxylations. However, based

Table III. Determination of positions of Pro/Hyp substitutions in unmodified HRF and first two hydroxylated HRF species (MALDI PSD).**precursor ion m/z 4138.9**

Sequence		H	Y	T	N	P	S	Q	D	V	T	V	P	C	P	V	P	S	T	P	P	T	P	S	P	S	T	P	P	T	P	S	P	S	C	C	H	P	R	
y-ions	theor.	4139.8	4139.8	4002.8	3839.7	3738.7	3624.6	3627.6	3440.6	3312.5	3197.5	3098.4	2997.3	2898.3	2801.2	2640.2	2543.2	2444.1	2347.0	2260.0	2159.0	2061.9	1964.8	1863.8	1766.7	1679.7	1582.7	1495.6	1394.6	1297.5	1200.5	1099.4	1002.4	915.3	818.3	731.3	570.2	409.2	272.2	175.1
	obs.	4139.6	4003.3	nd	nd	nd	3626.0	nd	3440.4	3313.2	3198.4	3098.0	2997.8	2898.8	nd	2640.3	nd	2445.2	nd	2260.0	2159.1	2062.4	nd	1865.4	nd	1682.4	nd	1495.7	1395.5	1298.9	1200.5	1100.2	nd	917.1	818.6	731.6	572.5	410.5	271.9	nd
b-ions	theor.	138.1	301.1	402.2	516.2	613.3	700.3	828.4	943.4	1042.5	1143.5	1242.6	1339.6	1500.6	1597.7	1696.8	1793.8	1880.8	1981.9	2078.9	2176.0	2277.1	2374.1	2461.1	2558.2	2642.2	2746.3	2843.3	2940.4	3041.4	3138.5	3225.5	3322.6	3409.6	3570.6	3731.6	3868.7	3965.7	4121.8	4121.4
	obs.	nd	nd	nd	nd	613.4	700.3	828.0	944.5	nd	11142.9	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	4121.4

precursor ion m/z 4154.9

Sequence		H	Y	T	N	P	S	Q	D	V	T	V	P	C	P	V	P	S	T	P	P	T	P	S	O	S	T	P	P	T	P	S	P	S	C	C	H	P	R																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
y-ions	theor.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
	obs.	4155.0	4155.8	4020.4	4018.8	3856.6	3855.7			3754.9	3754.7	3642.4	3640.6	3543.6	3456.5	3328.5	3328.5	3214.4	3213.5	3114.0	3114.4	3013.2	3013.4	2914.9	2914.3	2817.2	2655.6	2656.2	2559.2	2461.3	2460.1	2363.0	2276.2	2276.0	2175.4	2174.9	2078.5	2077.9	1980.8	1879.8	1782.7	1695.7	1582.7	1495.6	1394.6	1299.0	1297.5	1200.5	1099.4	1009.4	915.3	818.3	731.3	570.2	409.2	272.2	175.1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
b-ions	theor.	138.1	301.1	402.2	516.2	613.3	700.3	828.4	943.4	1042.5	1143.5	1242.6	1339.6	1500.6	1597.7	1696.8	1793.8	1880.8	1981.9	2078.9	2176.0	2277.1	2374.1	2461.1	2574.2	2661.2	2762.3	2859.3	2956.4	3057.4	3154.5	3241.5	3338.6	3425.6	3586.6	3747.6	3884.7	3981.7	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	4137.8	41

precursor ion m/z 4170.9

Sequence		H	Y	T	N	P	S	Q	D	V	T	V	P	C	P	V	P	S	T	P	O	T	P	S	O	S	T	P	P	T	P	S	P	S	C	H	P	R		
y-ions	theor.																																							
	obs.	4172.0	4171.8	4034.8	3871.7	3770.6	3656.6	3559.6	3472.5	3344.5	3229.5	3130.6	3029.3	2930.3	2833.2	2672.2	2575.1	2476.1	2379.0	2292.0	2190.9	2093.9	1981.6	1879.8	1782.7	1695.7	1582.7	1495.6	1394.6	1297.5	1200.5	1099.4	1002.4	915.3	818.3	731.3	570.2	409.2	272.2	175.1
b-ions	theor.	138.1	301.1	402.2	516.2	613.3	700.3	828.4	943.4	1042.5	1143.5	1242.6	1339.6	1500.6	1597.7	1696.8	1793.8	1880.8	1981.9	2078.9	2192.0	2293.0	2390.1	2477.1	2590.2	2677.2	2778.3	2875.3	2972.4	3073.4	3170.5	3257.5	3354.5	3441.6	3602.6	3763.6	3900.7	3997.7	4153.8	4153.8
	obs.	nd	301.2	nd	nd	614.0	nd	828.8	944.8	nd	11143.2	1244.5	1338.2	1502.0	nd	nd	nd	1881.8	nd	nd	2192.3	2292.3	nd	2478.0	nd	nd	22778.5	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	4153.5

nd, not detected; O, Hyp (Hydroxyproline); P, proline.

Cysteine residues were carboxymethylated.

Bolded and underlined values indicate altered masses of the fragment ions due to oxidation (Pro/Hyp conversion).

on the Hyp contiguity hypothesis, it is possible to predict that Pro232, Pro240, and Pro241 may also become hydroxylated, because they are situated in contiguous Pro regions (two consecutive Pro). According to the Hyp contiguity hypothesis, in addition to the identified Pro237, any of the other clustered noncontiguous Pro residues in the HRF sequence (Pro225, Pro227, Pro229, Pro235, Pro243, and Pro245) may also become hydroxylated, although it is not clear which ones of them are actually hydroxylated and

which ones are not. According to the same hypothesis, Pro250 is not likely to become hydroxylated, because it is neither contiguous nor clustered noncontiguous.

We determined that up to 10 Ara units can be attached to each hIgA1 molecule in the hinge region, with an average molar ratio of at least six Ara units per hIgA1 molecule. The data presented in this work are not sufficient to make conclusions about the length of the Ara chains attached to Hyp residues. From Table II, it is apparent that some of the

Table IV. Summary of amino acid analysis for E1 human IgA1 heavy chain

Amino acid	Theoretical mole % (full-length heavy chain) ^a	Observed mole %		
		Preparation 1	Preparation 2	Preparation 3 ^b
Asn + Asp	5.4	8.00	6.95	7.22
Gln + Glu	9.2	7.17	3.40	3.22
Ser	10.7	10.46	11.30	14.86
His	1.7	1.28	1.28	1.80
Gly	7.5	15.18	13.82	16.00
Thr	10.9	9.69	8.14	7.25
Ala	7.3	8.79	8.37	7.84
Arg	3.1	3.25	5.11	7.00
Tyr	2.7	2.54	2.18	2.18
Cys ^c	3.6	0.71	0.00	0.00
Val	8.2	7.14	6.56	5.39
Met	1.5	0.00	1.31	0.97
Trp ^c	1.9	0.00	0.00	0.00
Phe	2.9	4.90	3.06	3.01
Ile	1.3	2.62	2.14	2.00
Leu	9.4	9.70	8.86	6.47
Lys	3.8	4.38	4.25	3.48
Pro	9.0	3.85	12.82	10.53
Hyp	—	0.34^d	0.43^e	0.77^f
HyLys	—	0.00	0.00	0.00

^aMole % is based on assumption that the heavy chain is “full length” (478 residues in “theoretical sequence”).

^bTruncated version of heavy chain (HC): ~40 kDa band from SDS-PAGE under reducing conditions. N-terminal Edman degradation sequencing positively identified this protein band as E1 human IgA1 HC.

^cThese amino acids (Trp and Cys) are typically destroyed during strong acid hydrolysis and therefore not reliably detected.

^dThis corresponds to ~1.67 (rounded up to two) Hyp residues. Number of amino acid residues is calculated assuming “full length” heavy chain. (50,000 Da molecular weight).

^eThis corresponds to ~2.13 (rounded up to two) Hyp residues. Number of amino acid residues is calculated assuming “full length” heavy chain. (50,000 Da molecular weight).

^fThis corresponds to ~2.91 (rounded up to three) Hyp residues. Number of amino acid residues was calculated assuming the molecular weight of “truncated” heavy chain ~40 kDa (portion of “theoretical” sequence 1–378 residues [theoretical 39,896 Da] was used in calculations).

Hyp residues become modified with single Ara and that there must be at least two Ara on some Hyp residues. However, it is also possible that only one or two Hyp residues in the HRF become modified with longer chains of Ara, and the rest of the Hyp residues may not be arabinosylated at all. This aspect of arabinosylation of maize-expressed *h*IgA1 will have to be addressed in the future work. Further work needs to be done to show the positions of the Hyp residues that become arabinosylated in maize-expressed *h*IgA1, which may be important for prediction of such modification in other plant-expressed proteins.

For the first time, the extensin-like Hyp/Pro conversion and O-linked arabinosylation have been described for a recombinant therapeutic protein expressed in transgenic plants. Surprisingly, part of the amino acid sequence in the conserved hinge region of *h*IgA1 HC bears a striking similarity to Pro-rich repeat sequences of plant extensins and HRGP. We suggest that, in accordance with Pro/Hyp hydroxylation model and the Hyp contiguity hypothesis,

the *h*IgA1 hinge region undergoes extensin-like Hyp/Pro conversion and O-linked arabinosylation, when it is expressed in plants. This modification may occur in a similar way in other therapeutic proteins containing contiguous and clustered noncontiguous Pro residues in their amino acid sequences. Our findings suggest the need to examine the amino acid sequences of other candidate transgenic bio-therapeutic proteins for possible plant-specific posttranslational modifications.

All of the recombinant *h*IgA1 samples studied in this work were produced in the endosperm tissue of maize seeds. It is important to further investigate the extent of Pro/Hyp + Ara posttranslational modifications such as that occurring to *h*IgA1 in plant tissues other than endosperm and in different plant types and plant production systems. It is possible that these conversions occur with lower frequency in other environments.

At this time, it is not known if the Pro/Hyp + Ara modification has any impact on the stability, biological activity,

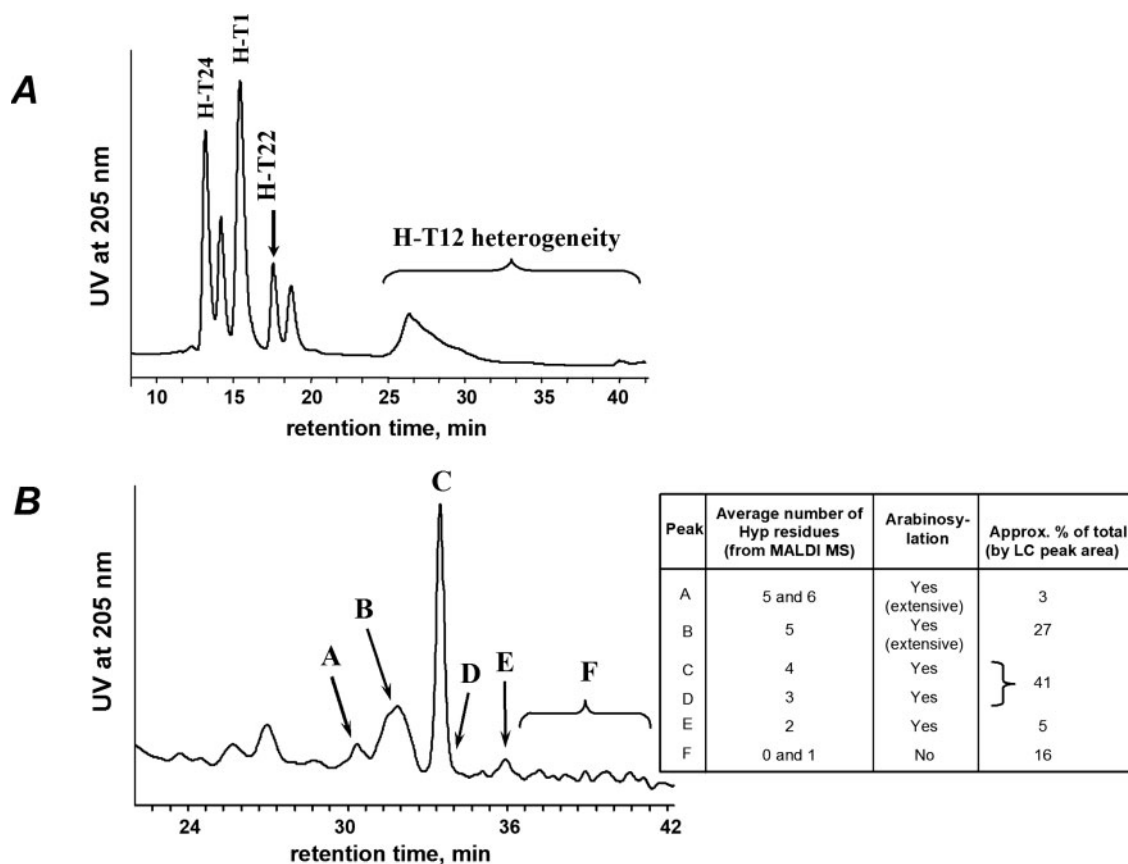


Fig. 8. Use of RP-HPLC/HILIC/RP-HPLC approach to analyze hinge region fragment (HRF) heterogeneity. (A) Further purification of HRF (H-T12 peptides of E1 human IgA1[hIgA1] heavy chain [HC]) by HILIC. Eluting peptides were identified with offline matrix-assisted laser-desorption/ionization mass spectrometry (MALDI MS). (B) The purified HRF pool that was collected off the HILIC column was concentrated and rerun on a RP-HPLC column. This gave another semiquantitative estimate of the species distribution within the HRF heterogeneity (see inset and Table V).

Table V. Abundance of maize-expressed human IgA1 antibodies with modified hinge region

Human IgA1 antibody	Estimated % of modified hinge region (Pro/Hyp + Ara)		
	by RP-HPLC of tryptic digest	by RP/NP/RP-HPLC	by MALDI MS
E1 (HX8), pooled	91.0	>84.0	85.0
E2-480	NA	NA	79.7
E2-540	NA	NA	94.2

MALDI MS, matrix-assisted laser-desorption/ionization mass spectrometry; NA, not applicable (not tested); RP-HPLC, reversed-phase high-performance liquid chromatography.

or efficacy of the IgA molecule, or whether these sugars could be a source of immunogenicity or allergenicity in patients. These issues will have to be clarified before any plant-produced IgA1 is taken into the clinic. The presence of these plant-specific glycan structures may require changing the antibody isotype to IgA2 or modification of the hinge region of an IgA1 before production in transgenic plants for human therapeutic use.

Materials and methods

Affinity purification of hIgA1 from extracts of transgenic maize seeds

Maize-expressed E1 and E2 hIgA1 were purified from pooled extracts of transgenic maize seeds using an affinity column with immobilized anti- α (heavy) chain-specific antibody. POROS-A resin (44 mL) (Applied BioSystems, Foster City, CA) was washed with phosphate buffered saline (PBS) buffer (Sigma, St. Louis, MO), and the resin was incubated at room temperature with 100 mg of goat anti-human-IgA (α chain specific, Southern Biotech, Birmingham, AL) for 1 h, after which the resin was sedimented by centrifugation, and the supernatant removed. The capture antibody was attached to the resin using 30 mM dimethyl pimelimidate (Sigma) in 100 mM triethanolamine (Sigma), pH 8.5 (~140 mL). After brief centrifugation and removal of the supernatant, the cross-linking reaction was quenched with 56 mL of 10 mM monoethanolamine (Sigma), pH 9.0. The resin was washed twice with PBS buffer and packed into column. The column was equilibrated with 15 column volumes of PBS buffer, pH 7.4, containing 0.01% sodium azide, at 5 mL/min. The filtered transgenic maize extract containing recombinant hIgA1 was applied to the column. The column was then

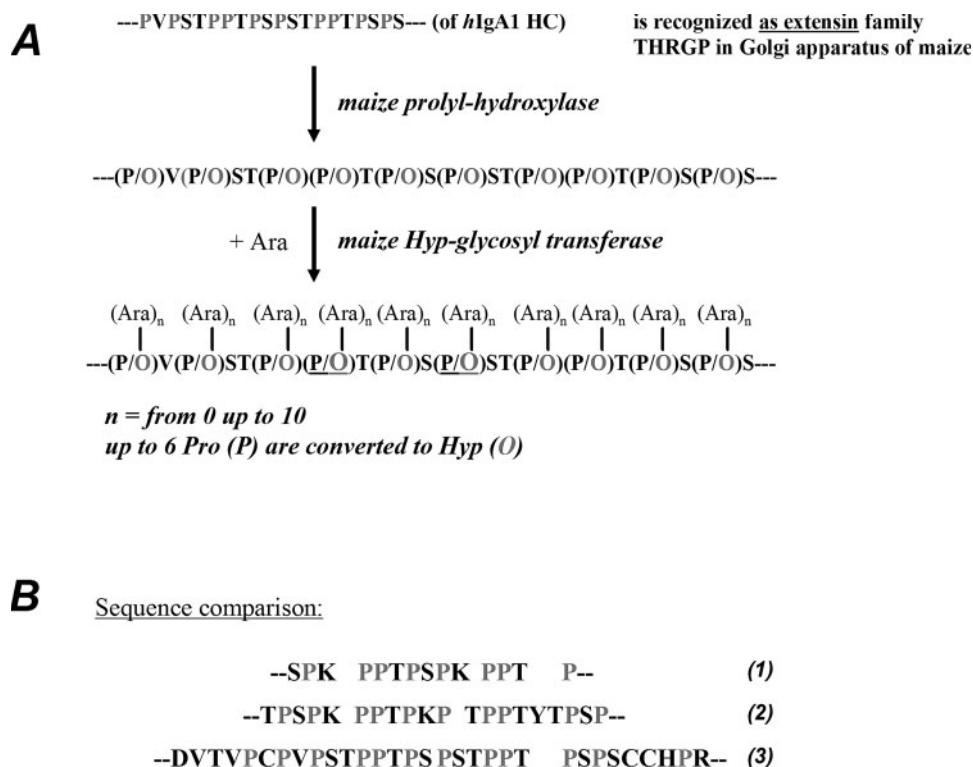


Fig. 9. (A) Suggested scheme for the *in vivo* modification of the hinge region of maize-expressed human IgA1 (*hIgA1*) heavy chain (HC). See text for detail. (B) Partial amino acid sequence comparison between (1) selected maize threonine/hydroxyproline-rich glycoproteins (THRGP) and extensins (Showalter, 1993), (2) maize HRGP, THRGP, and extensins (Stiefel *et al.*, 1990; Showalter, 1993), and (3) *hIgA1* HC hinge region fragment (Kabat *et al.*, 1991).

washed with 15 column volumes of PBS/0.01% sodium azide. The *hIgA1* was eluted from the column with seven column volumes of 100 mM Glycine/100 mM NaCl/0.02% NaN₃, pH 2.5. All the chromatography steps were performed at a constant flow rate of 5 mL/min. The collected *hIgA1* was neutralized with 1 M Tris-HCl buffer, pH 9.0, and quantified by Bradford assay (Bio-Rad, Hercules, CA) and enzyme-linked immunosorbent assay (ELISA).

Purity of the isolated *hIgA1* was examined by reducing and nonreducing 4–20% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) (precast Criterion gels, Bio-Rad). SDS–PAGE gels were stained with Coomassie Blue G-250 stain (Bio-Rad). Precision protein standards from Bio-Rad were used as molecular weight markers.

Reduction, carboxymethylation, digestion, and fractionation of purified hIgA1

Two milligrams of affinity-purified *hIgA1* was dissolved in 0.5 mL of 6 M guanidine hydrochloride/0.4 M ammonium bicarbonate, reduced by addition of 50 µL of aqueous 0.1 M dithiothreitol (DTT) and incubation at 65°C for 1 h. The sample was then alkylated by addition of 100 µL of aqueous 0.2 M iodoacetamide (at room temperature and for 2 h in the dark). The reaction was quenched by addition of 200 µL of DTT. The reduced and alkylated protein was desalted using NAP-5 cartridges (Amersham, Piscataway, NJ), according to the manufacturer's procedure, and

digested with trypsin (Roche Diagnostics) at 1:40 enzyme-to-protein ratio (16 h at 37°C).

A Luna C18(2), 4.6 mm × 150 mm (Phenomenex, Torrance, CA) column, and a Hitachi LC system were used for the separation. Chromatography was run at a constant flow rate of 2 mL/min and at room temperature. Four hundred microliters of the tryptic digest mixture was injected. The separation of peptides was accomplished using the following gradient: 100% solvent A (3% acetonitrile/0.06% trifluoroacetic acid [TFA]) isocratic for 5 min, 0–50% solvent B (80% acetonitrile/0.05% TFA) for 165 min, and 50–100% solvent B for 10 min. The column was then washed with 100% solvent B for 2 min and reequilibrated in 100% solvent A and then washed with 100% solvent A for 5 min. Elution of peptides was monitored by UV absorption at 215 nm. One-and-a-half-milliliter fractions were collected in siliconized microcentrifuge tubes and dried in a centrifugal evaporator. Before analysis by MALDI MS, fractions were redissolved in 4 µL of 50% acetonitrile/0.1% TFA, and 1 µL of the material in each fraction was examined by MALDI MS.

Tryptic/Asp-N digest of fractions containing hIgA1 hinge region fragment microheterogeneity

Fractions of separated tryptic digest of *hIgA1* containing glycopeptides with suspected O-linked glycosylation were first treated with PNGase-A (Roche Diagnostics) to release

any possible N-linked glycans according to the following procedure: fractions (dried in a centrifugal evaporator) were redissolved in 5 μ L of 20 mM ammonium acetate buffer, pH 5.0; then 5 μ L of PNGase-A solution was added; and the resulting solution was incubated at 37°C for 16 h. After examining aliquots of the samples by MALDI MS (no released N-glycans and no change in mass-spectral pattern of the HRF were detected), fractions were further treated with Asp-N protease as follows. Two micrograms of Asp-N protease (Roche Diagnostics) was resuspended in 50 μ L of Milli-Q water. Five microliters of 400 mM ammonium bicarbonate buffer, pH 7.8, then 5 μ L of the Asp-N solution was added to dried peptides, and the mixture was incubated at 37°C for 16 h. Before MALDI MS, peptides were desalted using C18 Zip-Tips (Millipore, Billerica, MA), according to manufacturer's procedure, and eluted onto MALDI plate in 50% acetonitrile/0.1% TFA.

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry

Voyager DE-STR (Applied BioSystems) MALDI ionization time-of-flight (MALDI-Tof) mass spectrometer operated in reflectron mode was used. The acceleration voltage was set to 20 kV. The grid voltage was set to 66% of the acceleration voltage. The delay time varied between 215 and 350 nsec. Five hundred acquisitions were averaged in each spectrum. The mass scale was calibrated using a Sequazyme peptide mass standard kit (Applied BioSystems). One microliter of sample of purified peptides was deposited onto a MALDI sample plate, overlaid with 1 μ L of α -cyano-hydroxycinnamic acid (CHCA), and air-dried.

MALDI PSD spectra were recorded using mirror voltage ratio 1.12; the following mirror ratios were used: 1, 0.85, 0.75, 0.65, 0.55, 0.4, 0.3, 0.2, 0.1, 0.05.

MALDI MS and MALDI PSD data were analyzed using Data Explorer v4.0 software (Applied BioSystems). Molecular weights and amino acid sequences of peptides and glycopeptides were attributed to the amino acid sequence of E1 and E2 hIgA1 using MassLynx v3.4 software (Micromass, Manchester, UK).

HILIC purification of hIgA1 HC HRF microheterogeneity, followed by RP-HPLC and MALDI MS analysis

A poly-hydroxyethyl aspartamide column (100 $\times$ 4.6 mm, PolyLC, Columbia, MD) with guard cartridge (35 $\times$ 4.6 mm, PolyLC) was used to separate hIgA1 HC HRF (H-T12 peptide, Figure 2A) microheterogeneity from other peptides that coeluted with HRF species during fractionation of hIgA1 tryptic digest. The following chromatographic conditions were used: column was washed with 100% mobile phase A (85% acetonitrile/10 mM triethylamine, pH 6.0 [adjusted with perchloric acid]) for 1 min, then content of mobile phase B (5% acetonitrile/10 mM triethylamine, pH 6.0 [adjusted with perchloric acid]) was increased to 22% in 20 min, then content of mobile phase B was increased to 100% in 5 min and kept at 100% for 20 min, and then column was reequilibrated to 100% mobile phase A. One-and-a-half-milliliter fractions were collected, dried in a centrifugal evaporator, and redissolved in 0.1% aqueous TFA; aliquots were desalted using C18 Zip-Tips (Millipore)

and analyzed by MALDI MS. Fractions, containing the HRF microheterogeneity, were pooled and rechromatographed using a Luna C18(2), 4.6 mm $\times$ 150 mm column (Phenomenex). Chromatography was run at a constant flow rate of 2 mL/min. Four hundred microliters of the sample was injected. The separation of peptides was accomplished using the following gradient: 100% solvent A (3% acetonitrile/0.06% TFA) isocratic for 5 min, 0–12% solvent B (80% acetonitrile/0.05% TFA) in 10 min, 12–24% solvent B in 45 min, 24–100% solvent B in 5 min, washed with 100% solvent B for 5 min, then column was reequilibrated to 100% solvent A. Elution of peptides was monitored by UV absorption at 205 nm. One-and-a-half-milliliter fractions were collected and dried in a centrifugal evaporator. Before analysis by MALDI MS, fractions were redissolved in 2 μ L of 50% acetonitrile/0.1% TFA.

Release of monosaccharides from isolated hIgA1 HC HRF fraction and monosaccharide analysis

A portion of the E1 hIgA1 HC HRF-containing fraction (after first-pass RP-HPLC) was dried in a centrifugal evaporator. The dry residue was dissolved in 320 μ L of Milli-Q deionized water (18 m Ω , ~20 ppb total organic carbon), and TFA was added to the solution to a final TFA concentration of 20% v/v. The tube with sample was tightly sealed and incubated in a sand bath at 103°C for 7 h. The “blank” (0.4 mL of 20% TFA) and “heated standard” (solution of 100 μ g Ara and 100 μ g Xyl in 0.4 mL of 20% TFA) samples were prepared in parallel to the sample using the same procedure. The samples were cooled to room temperature and dried in a centrifugal evaporator. The residue in each tube was dissolved in 300 μ L of Milli-Q deionized water, and the preparations were analyzed using HPAE/PAD chromatography. Carbohydrates were detected using pulsed amperometric detection with a gold electrode (Dionex ED50A, Dionex, Sunnyvale, CA). A Dionex DX-500 LC system equipped with Dionex PeakNet 6.3 software, a Dionex CarboPac PA10 column (250 mm $\times$ 4 mm; anion exchange), and a Dionex AminoTrap precolumn (50 mm $\times$ 4 mm) were used. The system was run isocratically using 18 mM NaOH as eluent. Flow rate was 1.2 mL/min. Injection volume was 10 μ L. High purity carbohydrate standards were obtained from Aldrich Chemical (St. Louis, MO). The eluent was freshly prepared on the day of analysis and was blanketed with helium during the course of analysis. The use of an AminoTrap precolumn is necessary to delay the elution of amino acids and small peptides that could be present in a peptide hydrolysate.

Amino acid analysis for the presence of Hyp

A sample of maize-expressed E1 hIgA1 was first separated by SDS-PAGE. Following staining and destaining of the gel, the appropriate bands were desorbed from each gel into 100 mM Hepes/0.1% SDS buffer, pH 8.0. The desorbed protein preparations corresponding to HC, light chain, and “truncated HC” were pooled separately, and total protein in each protein preparation was then determined by amino acid analysis (AAA). Analysis for Hyp required the following modifications to the standard AAA procedure. Desorbed

protein preparation was dialyzed against water, then transferred to a pyrolyzed glass tube and dried. The sample was hydrolyzed in gas phase in 6 N HCl (110°C; 20 h) in a sealed evacuated glass tube. Following hydrolysis, the sample was dried and dissolved in 150 μ L of loading buffer, and from this 5 μ L were subjected to analysis. Detection of Hyp was achieved by only allowing derivatization of the amino acids with Fmoc-Cl which eliminates detection of virtually all other residues that elute in proximity of Hyp.

N-terminal sequencing

A sample of maize-expressed E1 hIgA1 was first separated by SDS-PAGE. Following staining and destaining of the gel, the appropriate bands were desorbed from each gel into 100 mM Hepes/0.1% SDS buffer, pH 8.0. Each gel band preparation was dialyzed against water, and the resulting material lyophilized and taken up into the appropriate buffer. Chemically released amino acids were detected as the corresponding phenylthiohydantoin derivatives by HPLC. Twelve cycles of N-terminal sequencing were carried out for each protein sample. Before first cycle of sequencing, protein preparations corresponding to E1 hIgA1 HC (~50 kDa band) and "truncated" HC (~40 kDa band) were treated with pyro-Glu amino peptidase to deacylate N-terminal pyro-Glu and allow for its chemical release and derivatization in the first cycle of sequencing. After obtaining the N-terminal sequences, BLAST database was searched to confirm the identity of the investigated protein components. MALDI PSD sequencing of the N-terminal fragments resulting from E1 and E2 hIgA1 tryptic digests (fractionated by RP-HPLC) was also performed, and identity of pyro-Glu as the N-terminal amino acid as well as the rest of the amino acid sequence of the N-terminal HC tryptic fragments was directly confirmed.

Acknowledgments

We are indebted to Dr. Jean Roberts and her team at Dow AgroSciences for producing and providing extracts of transgenic maize. We greatly appreciate Dr. Scott A. Young for critically reviewing our manuscript and for his valuable insights.

Abbreviations

Ara, arabinose; Gal, galactose; GalNAc, N-acetyl-galactosamine; Glc, glucose; GlcNAc, N-acetylglucosamine; HC, heavy chain; Hex, hexose; HexNAc, N-acetylhexosamine; HILIC, hydrophilic interaction chromatography; HPAE, high-performance anion exchange; HPLC, high-performance liquid chromatography; HRGP, hydroxyproline-rich glycoproteins; HRF, hinge region fragment; MALDI, matrix-assisted laser-desorption ionization; MS, mass spectrometry; PAD, pulsed amperometric detection; PNGase, peptide N-glycosidase; PSD, postsource decay; PAGE, polyacrylamide gel electrophoresis; PBS, phosphate buffered saline; RP, reversed-phase; SDS, sodium dodecyl sulfate; sIgA, secretory IgA; TFA, trifluoroacetic acid; THRGP, threonine/hydroxyproline-rich glycoproteins; Xyl, xylose.

References

- Alpert, A. (1990) Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr.*, **499**, 177–196.
- Altmann, F., Tretter, V., Kubelka, V., Staudacher, E., Marz, L., and Becker, W.M. (1993) Fucose in α 1–3 linkage to the N-glycan core forms an allergenic epitope that occurs in plant and insect glycoproteins. *Glycoconj. J.*, **10**, 103.
- Anderson, D.J., Johnson, P.M., Alexander, N.J., Jones, W.R., and Griffin, P.D. (1987) Monoclonal antibodies to human trophoblast and sperm antigens: report of two WHO-sponsored workshops June 30, 1986. *J. Reprod. Immunol.*, **10**, 231–257.
- Bardor, M., Loutelier-Bourhis, C., Paccalet, T., Cosette, P., Fitchette, A.-C., Vezina, L.-P., Trepanier, S., Dargis, M., Lemieux, R., Lange, C., and others (2003a) Monoclonal C5–1 antibody produced in transgenic alfalfa plants exhibits a N-glycosylation that is homogenous and suitable for glycol-engineering into human-compatible structures. *Plant Biotechnol. J.*, **1**, 451–462.
- Bardor, M., Favreau, C., Fitchette, A.-C., Gilbert, D., Galas, L., Trottein, F., Faye, L., and Lerouge, P. (2003b) Immunoreactivity in mammals of two typical glycol-epitopes, core α (1,3)-fucose and core xylose. *Glycobiology*, **13**, 427–434.
- Briggs, K., Glancy, T., Hein, M.B., Hiatt, A.C., Karnoup, A.S., Anderson, W.H.K., Pareddy, D., Petolino, J., Rubin-Wilson, B., Taylor, D., Roberts, J.L. (2004) Patent WO-2004050838; PCT/WO/CAN/141/52833.
- Chintalacharuvu, K.A. and Morrison, S.L. (1997) Production of secretory immunoglobulin A by a single mammalian cell. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 6364–6368.
- Dickey, L.F., Gasdaska, J.R., Cox, K.M., Peele, C.G., and Spencer, D. (2004) Patent US 2004261148; US 2004-794615.
- Diekmann, A.B., Norton, E.J., Westbrook, V.A., Klotz, K.L., Naaby-Hansen, S., and Herr, J.C. (2000) Anti-sperm antibodies from infertile patients and their cognate sperm antigens: a review. Identity between SAGA-1, H6–3c4 Antigen, and CD52. *Am. J. Reprod. Immunol.*, **43**, 134–143.
- Fischer, R., Stoger, E., Schillberg, S., Christou, P., and Twyman, R.M. (2004) Plant-based production of biopharmaceuticals. *Curr. Opin. Plant Biol.*, **7**, 152–158.
- Fitchette-Laine, A.-C., Denmat, L.-A., Lerouge, P., and Faye, L. (1998) Analysis of N- and O-glycosylation of plant proteins. *Methods Biotechnol.*, **3**, 271–290.
- Gasdaska, J.R., Spencer, D., and Dickey, L. (2003) Advantages of therapeutic protein production in aquatic plant *Lemma*. *BioProcessing J.*, **Mar/Apr**, 49–56.
- Giddings, G. (2001) Transgenic plants as protein factories. *Curr. Opin. Biotechnol.*, **12**, 450–454.
- Gorr, G., Launhardt, H., and Berg, B. (2004) Patent WO 2004024927; WO 2003-EP9959.
- Hein, M.B. and Hiatt, A. (2002) Patent US 6417429.
- Hein, M.B., Hiatt, A., and Ma, J.K.-C. (1999) Patent US 5959177.
- Hellwig, S., Drossard, J., Twyman, R.M., and Fischer, R. (2004) Plant cell cultures for the production of recombinant proteins. *Nature Biotechnol.*, **22**, 1415–1422.
- Hiatt, A.C. and Hein, M.B. (1997) Patent US 5639947.
- Hiatt, A.C., Cafferkey, R., and Bowdish, K. (1989) Production of antibodies in transgenic plants. *Nature*, **342**, 76–78.
- Hiatt, A.C., Ma, J.K.-C., Lehner, T., and Mostov, K.E. (2001) Patent US 6303341.
- Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S., and Foeller, C. (1991) *Sequences of Proteins of Immunological Interest*, 5th ed. US Department of Health and Human Services, NIH, Bethesda, MD, **1**, 663, 681.
- Kieliszewski, M.J. (2001) The latest hype on Hyp-O-glycosylation codes. *Phytochemistry*, **57**, 319–323.

- Kieliszewski, M.J. and Lamport, D.T. (1994) Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J.*, **5**, 157–172.
- Kieliszewski, M.J., O'Neil, M., Leykam, J., and Orlando, R. (1995) Tandem mass spectrometry and structural elucidation of glycopeptides from a hydroxyproline-rich plant cell wall glycoprotein indicate that contiguous hydroxyproline residues are the major sites of hydroxyproline O-arabinosylation. *J. Biol. Chem.*, **270**, 2541–2549.
- Lamport, D.T. (1969) The isolation and partial characterization of hydroxyproline-rich glycopeptides obtained by enzymic degradation of primary cell walls. *Biochemistry*, **8**, 1155–1163.
- Lamport, D.T., Katona, L., and Roerig, S. (1973) Galactosylserine in extensin. *Biochem. J.*, **133**, 125–132.
- Larrick, J.W., Yu, L., Naftzger, C., Jaiswal, S., and Wycoff, K. (2001) Production of secretory IgA antibodies in plants. *Biomol. Eng.*, **18**, 87–94.
- Ma, J.K.C. and Hein, M.B. (1995) Immunotherapeutic potential of antibodies produced in plants. *Trends Biotechnol.*, **13**, 522–527.
- Ma, J.K., Lehrner, T., Stabila, P., Fux, C.I., and Hiatt, A. (1994) Assembly of monoclonal antibodies with IgG1 and IgA heavy chain domains in transgenic tobacco plants. *Eur. J. Immunol.*, **24**, 131–138.
- Mattu, T.S., Pleass, R.J., Willis, A.C., Kilian, M., Wormald, M.R., Lellouch, A.C., Rudd, P.M., Woof, J.M., and Dwek, R.A. (1998) The glycosylation and structure of human serum IgA1, Fab, and Fc regions and the role of N-glycosylation on Fc α receptor interactions. *J. Biol. Chem.*, **273**, 2260–2272.
- Mayfield, S.P., Franklin, S.E., and Lerner, R.A. (2003) Expression and assembly of a fully active antibody in algae. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 438–442.
- Pope, D.G. (1977) Relationships between hydroxyproline-containing proteins secreted into the cell wall and medium by suspension-cultured *Acer pseudoplatanus* cells. *Plant Physiol.*, **59**, 894–900.
- Rayon, C., Lerouge, P., and Faye, L. (1998) The protein N-glycosylation in plants. *J. Exp. Bot.*, **49**, 1463–1472.
- Rosin, L. (2004) Transgenic plants: bio-farming for the future. *Bioprocess Int.*, **2** (Suppl. 2), 52–61.
- Showalter, A.M. (1993) Structure and function of plant cell wall proteins. *Plant Cell*, **5**, 9–23.
- Shpak, E., Barbar, E., Leykam, J.F., and Kieliszewski, M.J. (2001) Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in *Nicotiana tabacum*. *J. Biol. Chem.*, **276**, 11272–11278.
- Simons, W. (1979) *The Sadtler Handbook of Ultraviolet Spectra*. Sadtler Research Laboratories, Philadelphia, PA; see also <http://www.bio.utk.edu>.
- Stiefel, V., Perez-Grau, L., Albericio, F., Giralt, E., Ruiz-Avila, L., Ludevid, M.D., and Puigdomenech, P. (1988) Molecular cloning of cDNAs encoding a putative cell wall protein from *Zea mays* and immunological identification of related polypeptides. *Plant Mol. Biol.*, **11**, 483–493.
- Stiefel, V., Ruiz-Avila, L., Raz, R., Valles, M.P., Gomez, J., Pages, M., Martinez-Izquierdo, J.A., Ludevid, M.D., Langdale, J.A., Nelson, T., and Puigdomenech, P. (1990) Expression of a maize cell wall hydroxyproline-rich glycoprotein gene in early leaf and root vascular differentiation. *Plant Cell*, **2**, 785–793.
- Stoger, E., Vaquero, C., Torres, E., Sack, M., Nicholson, L., Drossard, J., Williams, S., Keen, D., Perrin, Y., Christou, P., and Fischer, R. (2000) Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies. *Plant Mol. Biol.*, **42**, 583–590.
- Tan, L., Leykam, J., and Kieliszewski, M.J. (2003) Glycosylation motifs that direct arabinogalactan addition to arabinogalactan-proteins. *Plant Physiol.*, **132**, 1362–1369.
- Van den Steen, P., Rudd, P.M., Dwek, R.A., and Opdenakker, G. (1998) Concepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.*, **33**, 151–208.
- Wagner, S. (2003) Biopharmaceutical production via moss bioreactors. *Innovations Pharm. Technol.*, **3**, 56,58,60.
- Weiner, L.M. and Carter, P. (2003) The rollercoaster ride to anti-cancer antibodies. *Nature Biotechnol.*, **21**, 510–511.
- Wilson, I.B.H. (2002) Glycosylation of proteins in plants and invertebrates. *Curr. Opin. Biotechnol.*, **12**, 569–577.
- Wright, A. and Morrison, S.L. (1997) Effect of glycosylation on antibody function: implications for genetic engineering. *Trends Biotechnol.*, **15**, 26–32.
- Zeitlin, L., Whaley, K.J., Hegarty, T.A., Moench, T.R., and Cone, R.A. (1997) Tests of vaginal microbicides in the mouse genital herpes model. *Contraception*, **56**, 329–335.
- Zhang, J. and Wang, D. (1998) Quantitative analysis and process monitoring of site-specific glycosylation microheterogeneity in recombinant human interferon-gamma from Chinese hamster ovary cell culture by hydrophilic interaction chromatography. *J. Chromatogr. B Biomed. Sci. Appl.*, **712**, 73–82.