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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 November 2005 (24.11.2005)

PCT

(10) International Publication Number
WO 2005/110015 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2005/013252

(22) International Filing Date: 19 April 2005 (19.04.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/563,349 19 April 2004 (19.04.2004) US
60/653,236 15 February 2005 (15.02.2005) US

(71) Applicant *(for all designated States except US)*: **OHIO UNIVERSITY** [US/US]; Technology Transfer Office, Unit 14, 340 West State Street, Athens, OH 45701 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KIELISZEWSKI, Marcia, J.** [US/US]; 5251 Raymar Drive, Albany, OH 45701 (US). **HELD, Michael** [—/US]; 6492 State Road 26 West, West Lafayette, IN 47906 (US). **TAN, Li** [—/US]; 8864 United Lane, Apt. 31, Athens, OH 45701 (US).

(74) Agent: **MYERS-PAYNE, Sean, C.**; Calfee, Halter & Griswold LLP, 1100 Fifth Third Center, 21 East State Street, Columbus, OH 43215 (US).

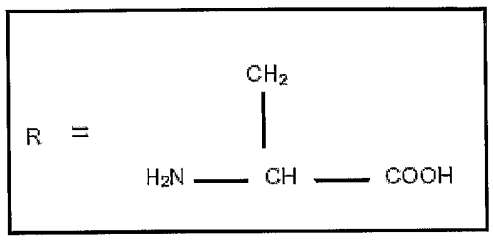
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

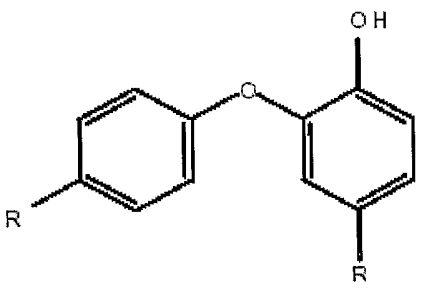
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CROSS-LINKABLE GLYCOPROTEINS AND METHODS OF MAKING THE SAME





(57) Abstract: Non-naturally occurring peptides/polypeptides/proteins comprising the crosslinking motif, Tyr-X-Tyr, wherein X is any amino acid, and methods of preparing the same.

WO 2005/110015 A2

CROSS-LINKABLE GLYCOPROTEINS AND METHODS OF MAKING THE SAME

DESCRIPTION OF THE INVENTION

[001] This application claims priority to U.S. Provisional Patent Application Nos. 60/563,349, filed April 19, 2004, and 60/653,236, filed February 15, 2005. The entire disclosures of these applications are incorporated herein by reference.

Field of the Invention

[002] The invention is directed to methods of making crosslinkable molecules, such as peptides/polypeptides/proteins, by introducing a crosslinking motif into the molecule's structure, and the crosslinkable molecules made according to these methods.

Background of the Invention

The plant cell wall

[003] The plant cell wall is composed of independent, interacting networks of cellulose microfibrils tethered by hemicellulosic polysaccharides, which are embedded in a matrix of pectins, glycoproteins, and phenolic substances (Carpita and Gibeaut 1-30). The plant cell wall is not an immutable substance; rather, it is a self-(re)organizing barrier able to respond to external and internal stimuli to govern cellular defense, growth, and development. For a size perspective, plant cell walls are approximately 0.1 – 10 μm thick depending on the type of plant species. In comparison, the plasma membranes of plant cells are not more than 0.01 μm thick (Stephen C.Fry).

[004] Plant cell walls are generally subdivided into two types: the primary cell wall and the secondary cell wall. Primary cell walls are synthesized and modified during a period of active cell growth and are often referred to as "growing cell walls" (Kerr and Bailey 327-49). Secondary cell walls are deposited after cell

growth has stopped. Since cell wall material is deposited from the inside out, the secondary cell wall is interior to the primary cell wall. Although the bulk of mature plant cell walls are composed of secondary wall material, there is great interest in the components of the primary cell wall, and their controlled assembly and modification, because primary cell walls dictate cell type, shape, and size, and provide a selective barrier to the external environment.

The hydroxyproline-rich glycoproteins

[005] The hydroxyproline-rich glycoproteins (HRGPs), which include the extensins, proline-rich proteins, and arabinogalactan-proteins, contribute to the extracellular matrix throughout the plant kingdom and the *Chlorophycean* green algae (Kieliszewski and Lamport 157-72;Showalter and Varner 485-520). HRGPs are involved in all aspects of plant growth and development involving wall architecture (Goodenough et al. 405-17;Roberts 129-46) and wall assembly during embryogenesis (Hall and Cannon 1161-72) as well as responses to biotic and abiotic stress (Merkouropoulos and Shirsat 356-66;Merkouropoulos, Barnett, and Shirsat 212-19;Yoshiba et al. 115-22) that include mechanical stress (Shirsat et al. 618-24); (Hirsinger et al. 343-55), physical wounding (Chen and Varner 2145-51;Han et al. 59-70;Showalter et al. 547-65;Zhou, Rumeau, and Showalter 5-17), pathogenesis (Benhamou et al. 457-67), and symbiosis (Cassab 441-46;Franssen et al. 4495-99;Frueauf et al. 429-38).

[006] HRGPs are extended macromolecules consisting of small repetitive peptide and glycopeptide motifs that form peptide modules and glycomodules of functional significance, as in "mix-and-match" mode they define the molecular properties of the overall macromolecule (Kieliszewski et al. 538-47;Kieliszewski and Lamport 157-72). The glycomodules result from a combination of posttranslational

modifications unique to plants, namely proline hydroxylation (Lamport 1438-40) and its subsequent glycosylation (Lamport 1322-24) that leads either to short arabinooligosaccharide or larger arabinogalactan polysaccharide addition to the Hyp residues. A sequence-dependent O-Hyp glycosylation code directs the precise addition of oligosaccharides and polysaccharides (Kieliszewski 319-23) and there is increasing evidence that other sequence-dependent codes direct inter- and intramolecular crosslinking of HRGPs. Crosslinked HRGPs, including extensins, contribute to wall architecture and defense responses by forming interpenetrating crosslinked networks in the wall. However, the precise identity of the intermolecular crosslink(s) has remained elusive.

The extensins

[007] Extensins are structural HRGPs that are covalently crosslinked into the primary cell wall, rendering them insoluble. Aside from being rich in Hyp, they also are rich in Ser, Tyr, Lys, Val and Thr. They are extensively post-translationally modified with short Hyp-O-oligoarabinosides and typically possess monogalactosyl-serine (Lamport, Katona, and Roerig 125-31;Lamport and Miller 454-56). Extensins are rigid linear molecules that adopt a polyproline II left-handed helical conformation (3 residues per turn, 9.4 Å pitch) (van Holst and Varner 247-51); (Heckman, Terhune, and Lamport 848-56); (Stafstrom and Staehelin 242-46).

[008] Three major types of extensin precursors to the extensin network are widespread in dicot plants, namely Precursor 1 (P1) extensins that are characterized by the repetitive motif: Ser-Hyp-Hyp-Hyp-Hyp-Thr-Hyp-Val-Tyr-Lys, (Smith et al. 1021-30;Smith, Muldoon, and Lamport 1233-39), P2 extensins that contain repeats of the motif Ser-Hyp-Hyp-Hyp-Hyp-Val-Tyr-Lys-Tyr-Lys (Smith et al. 1021-30;Smith, Muldoon, and Lamport 1233-39), and finally, the P3 extensins that contain a major

palindromic (bolded) repeat: Ser-**Hyp-Hyp-Hyp-Hyp-Ser-Hyp-Ser-Hyp-Hyp-Hyp-Hyp**-Tyr-Tyr-Tyr-Lys (Lamport 79-115;Smith et al. 1021-30).

[009] The P1 and P2 extensins can be isolated by the salt-elution of intact cells as soluble monomer precursors to the extensin network (Fong et al. 548-52;Smith et al. 1021-30;Smith, Muldoon, and Lamport 1233-39). The P3 extensins, on the other hand, have never been isolated as monomeric precursors to the extensin network, presumably due to their rapid incorporation into the cell wall via covalent intermolecular crosslinking (Mort and Lamport 289-309). Consequently, the molecular properties of soluble monomeric P3 extensins have thus far been inferred from gene sequences, and P3 glycopeptides and peptides rendered from cell walls (Corbin, Sauer, and Lamb 4337-44;Lamport 1155-63;Lamport 27-31;Lamport 79-115;Lamport and Miller 454-56;Showalter et al. 547-65;Showalter and Varner 375-92;Zhou, Rumeau, and Showalter 5-17).

Intramolecular extensin crosslinking

[010] Both P2 and P3 type extensins can undergo intramolecular crosslinking of Tyr residues (Epstein and Lamport 1241-46) (underlined in the sequences above) to form the diphenylether crosslink amino acid, isodityrosine (IDT) (Fig. 1) (Fry 449-55). First observed as an unknown tyrosine derivative in extensin peptides (Lamport 1155-63), and later identified in wall hydrolysates (Fry 449-55), and also as a component of trityrosine in *Ascaris* cuticle collagen (Fujimoto 637-43), IDT was initially hypothesized to be an intermolecular crosslink responsible for transforming the soluble extensin monomeric precursors, P1, P2 and P3, into an insoluble extensin network *in muro* (Fry 449-55); (Lamport and Epstein 73-83). But IDT was identified only as an *intra*molecular crosslink in P2 and P3-derived extensin peptides purified from enzymic digests of cell walls (Epstein and Lamport 1241-46)

and extensin peptides crosslinked by *inter*molecular IDT have never been isolated. Although it has been suggested that extracellular peroxidases catalyze the formation of IDT, to date, the precise mechanism *in muro* remains a mystery (Fry 853-62).

Intermolecular extensin crosslinking

[011] Evidence has suggested that pectin-protein crosslinks may play a role in extensin insolublization *in muro* (Keegstra et al. 188-96;Mort;Qi et al. 1691-701). Although these crosslinks may exist to some extent, the extensin network in cell walls and *in vitro* remains insoluble after complete deglycosylation with anhydrous hydrogen fluoride at 0°C (Mort and Lamport 289-309;Schnabelrauch et al. 477-89). Furthermore, Hyp-rich extensin peptides are only released after proteolytic cleavage of a partially or fully deglycosylated extensin network (Lamport 1155-63;Lamport 79-115;Mort and Lamport 289-309;Qi et al. 1691-701). Thus, the insolubility of cell wall extensins is primarily attributed to a protein-protein and/or protein-phenolic-protein crosslink(s) (Mort and Lamport 289-309).

[012] Peroxidase-catalyzed extensin intermolecular crosslinking has been demonstrated *in vitro* by several groups (Brownleader et al. 1115-23;Everdeen et al. 616-21;Jackson et al. 1065-76;Price et al. 41389-99;Schnabelrauch et al. 477-89), yet the molecular nature of the intermolecular crosslink(s) has not been were not identified. Indeed, Lamport and colleagues (Schnabelrauch et al. 477-89) found no IDT increase in P1 extensins after their crosslinkage *in vitro* by a tomato pI 4.6 extensin peroxidase, although the abundance of Val-Tyr-Lys motifs in several crosslinking extensins, including P1 and P2, suggested the intermolecular crosslinks involved tyrosine and/or lysine (Schnabelrauch et al. 477-89).

[013] More recently, Brady and Fry identified a trimeric tyrosine derivative, pulcherosine (Fig. 2) and the tetrameric tyrosine derivative, di-isodityrosine (diIDT;

Fig. 3) (Brady, Sadler, and Fry 349-53); (Brady, Sadler, and Fry 323-27) in tomato cell wall hydrolysates and speculated that IDT-containing extensins could be insolubilized through *inter*molecular IDT crosslinks forming pulcherosine and di-isodityrosine (Fig. 4). Also, as extensin incorporation into the cell wall increased, hydrolysates of these walls showed that the amounts of IDT decreased and diIDT increased (Brady and Fry 87-92). These findings suggest a significant role for IDT-rich extensins, which are supported by the recent discovery that RSH, an extensin containing 14 intramolecular IDT motifs, is crucially involved in positioning the cell plate during the earliest stages of embryogenesis in *Arabidopsis* (Hall and Cannon 1161-72).

[014] However, to date, there has been no direct demonstration of an extensin *inter*molecular crosslink that involves either Tyr or Lys. The results of *in vitro* assays have been difficult to interpret, as the substrate extensins, P1 and P2, contain both Tyr and Lys and the amino acids formed by crosslinking were not identified (Brownleader et al. 1115-23;Everdeen et al. 616-21;Fujimoto 637-43;Hall and Cannon 1161-72;Schnabelrauch et al. 477-89). Other approaches involving the isolation of intermolecularly crosslinked peptides from the cell wall itself have also proven intractable (Epstein and Lamport 1241-46).

SUMMARY OF THE INVENTION

[015] The present invention advances the art by, among other things, identifying a residue involved in glycoprotein crosslinking.

[016] In order to simplify the results of *in vitro* crosslinking assays and to test for the involvement of Tyr and Lys in crosslinking, we created a series of extensin mutants using the synthetic gene approach described earlier (Kieliszewski 319-23; Shpak et al. 11272-78; Shpak et al. 11272-78; Shpak, Leykam, and

Kieliszewski 14736-41; Tan L 1362-69; Tan, Leykam, and Kieliszewski 1362-69). The P3 extensin was of particular interest in view of its repetitive structure widespread in the plant kingdom and therefore functionally significant although frustratingly recalcitrant to isolation in precursor form; furthermore, it contains Tyr, Lys, and an intramolecular IDT motif. The synthetic gene approach allows us to explore the roles of IDT, Tyr, and Lys in crosslinking, and eventually to test the proposed role of the palindromic motif in molecular recognition (Kieliszewski and Lamport 157-72).

[017] Thus, we designed a series of synthetic genes, one encoding twenty and the other eight repeats of the P3 repetitive subdomain: Ser-Hyp-Hyp-Hyp-Hyp-Ser-Hyp-Ser-Hyp-Hyp-Hyp-Hyp-Tyr-Tyr-Tyr-Lys, as well as variants containing Tyr→Phe and Lys→Leu mutations. Overexpression of these genes as enhanced green fluorescent protein (EGFP) fusion proteins in BY-2 tobacco cells yielded secreted transgenic extensins. They were glycosylated as predicted by the Hyp contiguity hypothesis and contained IDT similar to native P3-type extensins. Furthermore, the pI 4.6 extensin peroxidase (E.C. 1.11.1.7) (Schnabelrauch et al. 477-89) catalyzed intermolecular crosslinking only between those P3 extensin analogs which contained Tyr. The present invention advances the art by, among other things, identifying a particular residue, i.e., tyrosine, involved in crosslinking.

[018] Thus, by using consensus sequences that are targeted for crosslinking, novel molecules can be created, which will form crosslinks. The consensus sequence for crosslinking is Tyr-X-Tyr, where X is any amino acid. In some embodiments, X is chosen from Tyr, Lys, and Val. The crosslinks that can form with this motif can be intermolecular or intramolecular. By inserting these sequences into proteins, crosslinkable proteins are generated.

[019] The present invention provides, for example, crosslinked fibers, emulsifiers, moleclular scaffolds, etc. Other specific applications of the invention include, but are not limited to, biologically active films. For example, a crosslinkable chimera involving an enzyme at one end (e.g., nitric oxide synthase or peroxidase or any other enzyme) and crosslinking extensin repeats at the other, would allow for an immobilized enzyme in a film to make a biosensor or for implantation at a wound site. Other immobilized proteins include, but are not limited to, lectins (which bind carbohdrate residues) and protein sequences that can chelate metals, etc. It is also contemplated that arabinogalactan modules (X-Hyp-X-Hyp) with extensin arabinosylation modules ($X\text{-}Hyp_n$) can be combined with crosslinking modules and with non-HRGP modules to create novel combinations.

[020] The present invention provides non-naturally occurring proteins comprising the amino acid sequence Tyr-X-Tyr, wherein X is chosen from any amino acid. In some embodiments, X is chosen from Tyr, Lys, and Val. In some embodiments, the protein is intermolecularly crosslinked to a non-naturally occurring protein comprising the amino acid sequence Tyr-X-Tyr. In some embodiments, the non-naturally occurring protein is intramolecularly crosslinked to itself.

[021] The non-naturally occurring proteins of the invention can further comprise the amino acid sequence $X\text{-}Hyp_n$, wherein X is any amino acid, and n is from 1 to about 1000. In some embodiments, X is chosen from Ser, Ala, Val, and Thr. The non-naturally occurring proteins of the invention can also further comprise the amino acid sequence X-Hyp-X-Hyp, wherein X is any amino acid. In some embodiments, X is chosen from Ser, Ala, Val, and Thr.

[022] The non-naturally occurring proteins of the invention can be crosslinked by reactions that involve the oxidation of tyrosine. Such reactions can be

catalyzed by peroxidases, including but not limited to, horseradish peroxidase, tomato peroxidase, extensin peroxidase, etc.

[023] Additional objects and advantages of the invention will be set forth in part in the description which follows, and in part will be obvious from the description, or may be learned by practice of the invention. The objects and advantages of the invention will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims.

[024] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed.

[025] The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate one (several) embodiment(s) of the invention and together with the description, serve to explain the principles of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[026] Figure 1 shows the chemical structure of isodityrosine (IDT). (MW = 360; (Fry, S.C. 1982))

[027] Figure 2 shows the chemical structure of pulcherosine. (MW = 539; (Brady, J.D. et al. 1998))

[028] Figure 3 shows the chemical structure of diisodityrosine (diIDT). (MW = 718; (Brady, J.D. et al. 1996))

[029] Figure 4 illustrates a possible *in muro* mechanism of pulcherosine and diIDT formation. From Brady et al. 1996, 1997, and 1998, pulcherosine must come from the oxidative coupling of 1 IDT with yr; DiIDT must come from the oxidative coupling of either pulcherosine and Tyr or 2 IDTs, as biphenyl linked dityrosine (upper right, boxed) is not found in plants.

[030] Figure 5 shows oligonucleotide sets for construction of the synthetic genes: (A) *FK9*, (B) *YK*, (C) *YL*, and (D) *FL*. (A) Construction of the *FK9* gene involved three sets of partially overlapping, complementary oligonucleotide pairs that were polymerized head-to-tail as described earlier (McGrath, K.P. et al.1990) (Shpak, E. et al.1999). Nine internal repeats and one of each linker set made up the *FK9* synthetic gene. The restriction sites used for subcloning are highlighted and labeled. The *YK, YL,* and *FL* genes were constructed through primer extension of the sense and antisense oligonucleotide sets shown in (B), (C), and (D) which involved elongation through use of the complementary but nonregenerable *BbsI* and *BsmFI* restriction sites. As the enzymes *BbsI* and *BsmFI* do not restrict the site that they recognize, both the recognition sequence and the sequence that is restricted are highlighted with the same color, the label directly placed above the recognition site. The underlined sequences in (B) (C) and (D) indicate the 24 base pair overlapping sections of the oligonucleotide sets.

[031] Figure 6 illustrates a plasmid map of the *pUC18-FK9* plasmid. The *FK9* synthetic gene was inserted as *BamHI-EcoRI* fragment into a *pUC18* vector.

[032] Figure 7 illustrates the construction of the *YK20*, *YK8*, and *YL8* using oligonucleotides containing complementary, nonregenerable restriction sites. This approach used general methods developed earlier (Lewis, R.V. et al.1996). Primer extension of the sense and antisense overlapping oligonucleotide sets shown in Fig. 5B, C, and D (here designated *YK-ss* and *YK-as*, respectively) initially gave genes having a single repeat (upper left) which were inserted into *pUC18* as *XmaI/SacI* fragments and then sequenced. The resulting plasmid, here designated *pUC18-YK*, was divided into two aliquots; one aliquot was restricted with *BbsI* and *ScaI* and the other with *BsmFI* and *ScaI*. The *BbsI/ScaI* and *BsmFI/ScaI* fragments were isolated,

annealed to each other through the complementary nonregenerable *BbsI/BsmFI* site and the *ScaI* site, then ligated to give rise to *pUC18-YK2*. Longer P3 analog genes were made by repeating the procedure.

[033] Figure 8 illustrates a plasmid map of *pUC18-YK*. The *YK* synthetic gene was inserted as an *XmaI-SacI* fragment into the *pUC18* plasmid. This procedure was also employed for the synthesis of *pUC18-YL* and *pUC18-FL*.

[034] Figure 9 illustrates a plasmid map of the *modified pUC-SStob-YK20-EGFP*. Synthetic genes *YK20* (shown here), *YK8, YL20, YL8,* and *FL8* were inserted as *XmaI-NcoI* fragments into the *modified pUC-SStob-EGFP* vector between *SStob* and the *EGFP* reporter gene.

[035] Figure 10 illustrates a plasmid map of recombinant pBI121-SStob-YK20-EGFP. Synthetic genes SStob-YK20-EGFP (shown here), SStob-YK8-EGFP, SStob-YL20-EGFP, SStob-YL8-EGFP, SStob-FL8-EGFP, and SStob-FK9-EGFP were inserted as BamHI-SacI fragments into the pBI121 binary plant transformation vector replacing the β-glucuronidase gene as described by Shpak, E.et al. 1999. Nos-pro: nopaline synthesis promoter; NOS-ter: nopaline synthesis terminator; NPTII: Neomycin Phosotransferase II gene.

[036] Figure 11 shows oligonucleotide sequences of PCR and DNA sequencing primers. (a) *M13 forward* primer is complimentary to 21 bp upstream of the MCS region of *pUC 18* and was used for sequencing synthetic genes in *pUC* vectors. (b) The tobacco signal sequence primer (*SSeq4Sen*) binds 4 bp into the signal sequence and was used for PCR of genomic DNA from transformed cell lines as well as the DNA sequencing of the PCR products. (c) The EGFP primer (*EGFP566as*) is complimentary to the sense strand of the EGFP gene 566 bp from the origin.This oligonucleotide was used for PCR of genomic DNA and the

sequencing of synthetic genes in both *pUC* and *pBI121* vectors. All oligonucleotide primers were ordered from Integrated DNA Technologies.

[037] Figure 12 is a flowchart for the isolation of EGFP fusion glycoprotein. (diagram courtesy of Dr. Li Tan). Culture medium of transformed tobacco cells was concentrated and processed by HIC, gel filtration and reversed phase HPLC.

[038] Figure 13 shows the DNA and encoded protein sequences of the P3-type extensin analogs. (A) *SStob-YK8-EGFP* and *SStob-YK20-EGFP*, (B) *SStob-YL8-EGFP* and *SStob-YL20-EGFP* , (C) *SStob-FK9-EGFP* and (D) *SStob- FL8-EGFP*. Each gene encoded a signal sequence, a P3 extensin analog, and EGFP, which is presented only in part. The *XmaI* and *NcoI* sites allowed insertion of the genes into the *modified pUC18* vector described earlier (Shpak, E. et al. 1999) between the signal sequence-encoding region (*SStob*) and *EGFP*. The *BamHI* site together with a *SacI* site (not shown) located at the 3'end of *EGFP* allowed insertion of *SStob-YK20-EGFP* and the other constructions into the binary plant transformation vector, *pBI121*.

[039] Figure 14 shows PCR amplification of genomic DNA from cell lines NtYK20-EGFP and NtYL8-EGFP. Insertion of the synthetic genes *SStob-YK20-EGFP* and *SStob-YL8-EGFP* into the tobacco genome was verified by PCR amplification with oligonucleotide primers complimentary to the antisense strand of *SStob* and the sense strand of *EGFP*. Bands at 783 bp and 1359 bp from lanes 1 and 2 respectively were eluted from the gel and sequenced. Lane 1: *SStob-YL8-EGFP* PCR product; Lane 2: *SStob-YK20-EGFP* PCR product; Lane 3: untransformed tobacco PCR product (control); Lane 4: 100 bp ladder.

[040] Figure 15 is a visualization of cell lines (A) NtYK20-EGFP, (B) NtYK8-EGFP, (C) NtYL8-EGFP, (D) NtFK9-EGFP, (E) uTob, and (F) NtSS-EGFP by

fluorescence microscopy. Slides shown here are an overlay of the laser scanned image (exc. 488 nm, emm. 510 nm) and the transmitted light image.

[041] Figure 16 is a visualization of plasmolyzed SS-EGFP, uTob, and YK20-EGFP cells by fluorescence microscopy. The (A) SS-EGFP, (B) uTob, (C and D) YK20-EGFP cells were plasmolyzed with 750 mM mannitol. The (E) YK20-EGFP cells were plasmolyzed with 500mM potassium phosphate pH 7. Slides shown here are an overlay of the laser scanned image (exc. 488 nm, emm. 510 nm) and the transmitted light image. Green fluorescence in the cell walls after plasmolysis was not observed for cells expressing YK8-EGFP, YL8-EGFP, and FK9-EGFP.

[042] Figure 17 shows a fractionation of transgenic NtYK20-EGFP culture medium by HIC. Concentrated and desalted NtYK20-EGFP (shown here), NtYK8-EGFP, NtYL8-EGFP, and NtFK9-EGFP culture media were loaded onto a HIC column and eluted stepwise from 2 M NaCl to 1 M NaCl and then to water. The eluate was monitored by in-line fluorescence detection (exc. 488 nm, emm. 510 nm; Hewlett Packard, USA). A single green peak eluted in 1 M NaCl.

[043] Figure 18 shows purification of the YK20-EGFP, YK8-EGFP, YL8-EGFP, and FK9-EGFP fusion glycoproteins by C4 reversed phase HPLC. (A) YK20-EGFP (300 µg), (B) YK8-EGFP (200 µg), (C) YL8-EGFP (200 µg), and (D) FK9-EGFP (45 µg) were injected onto an analytical C4 reversed phase HPLC column. The fusion glycoproteins eluted between 28 and 32 min (58% and 68 % end buffer).

[044] Figure 19 shows the isolation of glycomodules YK20, YK8, YL8, and FK9 by C4 reversed phase HPLC. Glycomodules YK20, YK8, YL8, and FK9 eluted between 30-40% end buffer, prior to their respective undigested fusion glycoproteins.

[045] Figure 20 shows a C4 reversed phase HPLC purification of dYK20-EGFP. Deglycosylated YK20-EGFP (200 µg) was purified on a C4 analytical

column. The dYK20-EGFP fusion protein showed a slightly increased retention time (~69% end buffer) compared to the respective fusion glycoprotein.

[046] Figure 21 demonstrates an examination of acid hydrolysates of (A) YK20, (B) YK8, (C) YL8 glycomodules and (D) FK9-EGFP fusion glycoprotein for IDT. Acid hydrolysates (10 µg each) were fractionated by SEC. The YK20, YK8, and YL8 glycomodules contained IDT whereas FK9-EGFP did not.

[047] Figure 22 shows the detection of Hyp-PS in base hydrolysates of YL8-EGFP fractionated by Superdex-75 gel filtration chromatography. A HMW Hyp-O-glycoside peak was observed in the hydrolysate of YL8-EGFP eluting between 18 and 38 min. This confirmed the presence of Hyp-PS attachment to YL8-EGFP.

[048] Figure 23 shows the purification of the pI 4.6 extensin peroxidase by anion exchange HPLC. Heme absorbance was monitored at 405 nm, and a single peak eluted at ~50% B (250 mM NaCl).

[049] Figure 24 shows *in vitro* crosslinking of P1 extensin (40 µg) by the pI 4.6 extensin peroxidase. Crosslinking reactions at time zero (A) and after 15 minutes of incubation (B) were separated by Superose-6 gel filtration chromatography. Peaks eluting after 22 min are a result of the crosslinking buffer and stopping reagent. P1, monomic P1 extensin; P1XL, crosslinked P1 extensin.

[050] Figure 25 shows *in vitro* crosslinking of YK20 (40 µg) by the pI 4.6 extensin peroxidase. Crosslinking reactions at time zero (A) and after 15 minutes of incubation (B) were separated by Superose-6 gel filtration chromatography. YK20, monomeric YK20; YK20XL, crosslinked YK20. The pI 4.6 extensin peroxidase crosslinks YK20 *in vitro*.

[051] Figure 26 shows *in vitro* crosslinking reactions of FK9 (40 µg). Crosslinking reactions at time zero (A) and after 15 minutes of incubation (B) were

separated by Superose-6 gel filtration chromatography. Note FK9 did not crosslink *in vitro*. FK9 mono, monomeric.

[052] Figure 27 shows *in vitro* crosslinking reactions using BSA (40 µg) as a control substrate. Crosslinking reactions were incubated 0 min (A) and 15 min (B) then separated by Superose-6 gel filtration chromatography. BSA did not crosslink.

[053] Figure 28 shows *in vitro* crosslinking reactions of YK20 (40 µg) lacking the extensin peroxidase enzyme. Crosslinking reactions lacking the addition of peroxidase were incubated 0 min (A) and 15 min (B) then separated by Superose-6 gel filtration chromatography. YK20 does not crosslink in this time frame without the addition of extensin peroxidase.

[054] Figure 29 shows *in vitro* crosslinking reactions of YK20 (40 µg) lacking the H_2O_2 co-substrate. Crosslinking reactions lacking the addition of hydrogen peroxide were incubated 0 min (A) and 15 min (B) then separated by Superose-6 gel filtration chromatography. YK20 does not crosslink with out the addition of H_2O_2.

[055] Figure 30 shows *in vitro* crosslinking reactions lacking extensin substrate. Crosslinking reactions were incubated 0 min (A) and 15 min (B) then separated by Superose-6 gel filtration chromatography. Note that peaks eluting after ~22min are a result of the buffer and stopping reagents.

[056] Figure 31 shows crosslinking rates of P1, YK20, YK8, YL8, and FK9. Crosslinking rates were calculated as described earlier (Everdeen, D.S. et al.1988) using the first order rate equation $A = Ao * e^{-kt}$ where A is the monomer remaining (µg) after t (sec) of incubation; Ao is the amount of monomer (µg) at time zero; k is the first order rate constant.

[057] Figure 32 shows *in vitro* crosslinking of YK20 by a crude cationic void peroxidase preparation. Reactions were incubated 0 min (A) and 30 min (B), then separated by Superose-6 gel filtration(0.5 ml/min; OD 220 nm). Some crosslinking was observed for YK20 (shown here), P1, and YL8.

[058] Figure 33 shows size exclusion chromatography of (A) YK20 acid hydrolysate and (B) YK20XL acid hydrolysate. A putative crosslinking amino acid (unknown) eluting at 10.4 min was isolated from the hydrolysate of YK20XL.

[059] Figure 34 shows a 1-D ^{1}HNMR spectrum of the "unknown." The spectrum was obtained by Dr. Li Tan of Ohio University. (A) N-linked proton resonances, (B)-(E) aromatic proton resonances.

[060] Figure 35 shows a predicted 1-D ^{1}HNMR spectrum of diIDT.

[061] Figure 36 diagramatically illustrates the *pBI121* plasmid with signal sequence-synthetic gene-EGFP. The method of gene construction was adapted from Shpak et al. (*Proc. Natl. Acad. Sci.*, **1999**, *96*, 14736-14741) Overlapping oligonucleotide pairs were annealed and polymerized. The synthetic gene was inserted into *pUC18* as a BamHI-EcoRI fragment, sequenced and then inserted between the signal sequence and EGFP in *pUC-SS-EGFP* as a XmaI-NcoI fragment. Finally, the signal sequence-synthetic gene-EGFP unit was placed in the plant transformation vector *pBI121* as a BamHI-SacI fragment.

[062] Figure 37 is a photomicrograph of cell lines expressing EGFP fusion proteins. Panels A. $[VYK]_6$ and B. $[VFL]_6$ are transformed cell lines that show EGFP fluorescence when viewed with a confocal laser scanning fluorescence mircroscope, excitation 488nm and emission 510mn.

[063] Figure 38 shows results of cross-link assays on a Superose 6 gel filtration column. Superose 6 profiles of P1 extensin at 220nm : (a) before and (b)

after cross-linking for 15 minutes. Superose 6 profiles of VYK module at 220nm : (c) before and (d) after cross-linking for 19 hours. As cross-linking occurs, the monomer peak decreases and a peak representing a larger cross-linked oligomer appears.

[064] Figure 39 is a flowchart of oligonucleotide construction.

[065] Figure 40 shows size exclusion chromatography before (top frame) and after (bottom frame) 30 minutes of crosslinking of $(AP)_4(YK)_{20}$ catalyzed by extensin peroxidase.

[066] Figure 41 shows chromatographs of the cross-linking reaction of $(AP)_4(YK)_{20}$EGFP.

[067] Figure 42 shows a flow chart for gene construction.

DESCRIPTION OF THE EMBODIMENTS

[068] Reference will now be made in detail to the present embodiments (exemplary embodiments) of the invention.

[069] The following abbreviations are used throughout this application: 1-D ^{1}HNMR: 1-D proton nuclear magnetic resonance; ABTS: 2,2′-azino-bis (3-ethylbenzthiazoline-6-sulfonic acid); ACS: American chemical society; BAW: Butanol:acetic acid:water (12:3:5) by volume; bp: base pair; BSA: Bovine serum albumin; BY-2: Bright yellow-2 *Nicotiana tabacum*; CaMV 35S: Cauliflower mosaic virus 35S promoter; CHO: carbohydrate; Da: Daltons; ddH_2O: distilled deionized water; DEAE: diethylaminoethyl; diIDT: diisodityrosine; *E. Coli*: *Escherichia coli*; E1: crude peroxidase preparation from untransformed tobacco culture medium; E2: crude peroxidase preparation from salt-eluates of untransformed tobacco cells; %GC: percentage of guanine and cytosine base pairs; EGFP: enhanced green fluorescent protein; emm: emission wavelength; exc: excitation wavelength; Fry's diIDT: authentic diisodityrosine standard from Dr. Stephen C. Fry; HF: anhydrous

hydrogen fluoride; HIC: Hydrophobic interaction chromatography; HMW: High molecular weight; HPLC: High performance liquid chromatography; HRGP: Hydroxyproline-rich glycoprotein; Hyp: Hydroxyproline; Hyp-O-Aran: O-linked hydroxyproline oligoarabinosides; Hyp-PS: O-linked hydroxyproline arabinogalactan polysaccharide; IDT: isodityrosine; LB: Luria broth; MALDI-TOF MS: Matrix assisted laser desorption ionization-time of flight mass spectrometry; MCS: Multiple cloning sequence; MS: Mass spectrometry; MW: Molecular weight; MWCO: Molecular weight cut-off; My diIDT: Diisodityrosine standard isolated by the inventor; NG-Hyp: Non-glycosylated Hydroxyproline; NOS pro: nopaline synthesis promoter; NOS ter: nopaline synthesis terminator; NPT II: Neomycin phosphotransferase II gene; Nt: Nicotiana tabacum; NT-1: Nicotiana tabacum 1 culture medium; P1: Precursor 1 extensin; P1XL: crosslinked P1 extensin; P2: Precursor 2 extensin; P3: Precursor 3 extensin; PCR: Polymerase chain reaction; PCV: Packed cell volume; PITC: phenylisothiocyanate; PTC-aa: phenylthiocarbamyl-amino acid derivative; Rf: relative to the solvent front; RSH: Arabidopsis mutant "Root-Shoot-Hypocotyl" defective; SEC: Size exclusion chromatography; Ser-O-gal: O-linked monogalactosyl-serine; SH: Schenk and Hildebrandt; SH*: SH medium lacking kinentin; SStob: tobacco signal sequence; TaOpt: Optimum annealing temperature; TFA: Trifluoroacetic acid; Tm: Melting temperature; uTob: untransformed tobacco BY-2 cells; YK20: YK20 glycomodules; and YK20XL: crosslinked YK20 glycomodule.

[070] Earlier work using a synthetic gene approach allowed us to elucidate the codes that drive Hyp-O-glycosylation in plants (Shpak et al. 11272-78;Shpak, Leykam, and Kieliszewski 14736-41;Tan, Leykam, and Kieliszewski 1362-69). Here we have extended this work to the extensins, not only as another test of the Hyp contiguity hypothesis for glycosylation, but also to examine codes for other

posttranslational modifications, including intra- and intermolecular extensin crosslinking, which is a long-standing and formidable problem relevant to plant growth, development, and disease resistance. Thus, we designed a series of crosslinking and non-crosslinking P3-type extensin analogs for their subsequent expression and isolation.

[071] P3-type extensins play a crucial role in the cell wall, judging from their ubiquity and their highly conserved repetitive sequences that include palindromic (bolded) and IDT motifs (underlined): Ser-**Hyp-Hyp-Hyp-Hyp-Ser-Hyp-Ser-Hyp-Hyp-Hyp-Hyp**-Tyr-Tyr-Tyr-Lys. Such supersymmetry combined with potential intra- and intermolecular crosslink sites involving IDT may enhance self-assembly of tightly packed networks (Knupp and Squire 558-77). Likewise, the related RSH extensin required for cell plate orientation during cytokinesis (Hall and Cannon 1161-72) contains 14 IDT motifs and a repetitive symmetry that, although not palindromic, should favor close packing.

[072] Two methods were used to make the synthetic genes. One involved polymerization of duplex DNA oligomer sets (the *FK* set in Figure 1A) (McGrath et al. 186-92; Shpak, Leykam, and Kieliszewksi 14736-41), the result being small genes containing at most only nine internal repeats. A second approach (Lewis et al. 400-06) was used to make the *YK20, YK8*, and *YL8* genes and allowed precise control of the sizes. This strategy involved compatible but nonregenerable restriction sites to generate a gene with double the number of repeats of the starting gene (Fig. 7). Judging from extensin genes, it was estimated that 20 repeats represented a 'full length' extensin, containing about 320 amino acids (Zhou, Rumeau, and Showalter 5-17), although there is considerable variation in extensin lengths. Therefore, one construction, *YK20*, was made containing 20 palindromic P3 crosslinking modules.

Shorter variants, namely YK8, YL8 and FK9, were also made to determine if the number of repeats influenced crosslinking rates and to determine the amino acid requirements for crosslinking. This method is completely amenable to any insert sequence (HRGP or other), for example, by using *BbsI* and *BsmFI* restricton enzymes that have separate recognition and restriction sites. This method also allows the design of synthetic genes having more than one consensus sequence. For instance, repeat units can be combined to control crosslink densities and/or to create modules such as AGP/extensin chimeras.

[073] Transformation of tobacco BY2 cells with the synthetic genes yielded several cell lines for each construction. Cell lines releasing the most fusion protein into the medium were chosen for isolation and biochemical characterization of the transgene products. The yields were on the level of those previously reported (Shpak et al. 11272-78; Tan L 1362-69; Zhao et al. 431-44). Transformed tobacco cells were maintained in SH medium lacking kinetin as this seemed to improve secretion and overall fusion protein yields as compared with SH medium (not empirically determined).

[074] A combination of HIC, gel filtration, and reversed phase chromatographies (Fig. 18) gave pure fusion proteins judging by N-terminal sequence analyses that yielded a single sequence for each protein (Table 5). The amino acid compositions of the isolated YK20, YK8, YL8, and FK9 glycoproteins were also consistent with those predicted by the genes (Table 4) although values for Hyp/Pro and Ser were somewhat lower than predicted. Virtually every Pro residue was hydroxylated and the modules containing Tyr showed conversion of Tyr to IDT. The PTC-isodityrosine was extremely hydrophobic and eluted with and was indistinguishable from the reagent peaks at the end of the gradient during amino acid

analysis. Therefore, IDT was quantified from acid hydrolysates fractionated on a gel filtration column (Schnabelrauch et al. 477-89) (Figure 21) compared to an authentic IDT standard (a gift from Dr. Derek Lamport). Putative IDT motifs of extensins are widespread and quite variable throughout plant species (Table 15). Comparison of the amounts of IDT in YK8 and YL8, show that not all -Y-Y-Y-X- motifs form IDT to the same extent. In other words, the flanking X residue (K or L) appears to influence IDT formation. These crosslinking motif subtleties may represent another level of control over wall crosslinking and assembly.

Table 15. Putative IDT motifs found in extensin genes of various plants. A short-nearly exact protein blast of http://www.ncbi.nlm.nih.gov/BLAST/ with the query "SPPPPYYYK" revealed extensin genes having putative IDT motifs in at least 23 different plant species.

Putative IDT Motif	Plant species
YYYK	*Brassica napus, Phaseolus vulgaris, Solanum tuberosum, Vigna unguiculata, Glycine max, Manihot esculenta, Lycospersicon esculenta, Cicer arietinum, Bromheadia finlaysonia, Petroselinum crispum, Medicago trucatula, Pisum sativum, Vicia faba*
YYYH	*Brassica napus, Phaseolus vulgaris, Glycine max, Manihot esculenta, Lycospersicon esculenta, Cicer arietinum, Arabidopsis thaliana*
YYYS	*Brassica napus, Lycospersicon esculenta, Petroselinum crispum, Arabidopsis thaliana, Nicotiana sylvestris, Oryza sativa*
YYYQ	*Brassica napus, Phaseolus vulgaris, Petroselinum crispum, Pisum sativum, Lupinus angustifolius*
YYYN	*Brassica napus, Phaseolus vulgaris,Vigna unguiculata, Arabidopsis thaliana*
YYYY	*Phaseolus vulgaris, Oryza sativa, Lupinus angustifolius*
YYYT	*Solanum tuberosum, Petroselinum crispum,*
YYYV	*Phaseolus vulgaris*
YYYI	*Bromheadia finlaysonia*
YYYL	*Arabidopsis thaliana*
YHY	*Brassica napus, Phaseolus vulgaris, Glycine max, Lycospersicon esculenta, Cicer arientinum, Arabidopsis thaliana, Pisum sativum, Nicotiana tabacum*
YHYV	*Phaseolus vulgaris, Glycine max, Cicer arietinum,*
YHYS	*Pisum sativum, Catharanthus roseus*
YHYT	*Glycine max, Cicer arietinum*
YHYY	*Catharanthus roseus, Nicotiana sylvestris,*
YHYQ	*Cicer arietinum*
YHYH	*Arabidopsis thaliana*
YHYE	*Daucus carrota*
YVYK	*Brassica napus, Solanum tuberosum, Vigna unguiculata, Glycine max, Lycospersicon esculenta, Arabidopsis thaliana, Lupinu angustifolius, Adiantum capillus-veneris, Catharanthus roseus*
YVYS	*Vigna unguiculata, Arabidopsis thaliana, Pisum sativum, Nicotiana sylvestris*
YVYQ	*Brassica napus*
YVYG	*Vigna unguiculata*

Putative IDT Motif	Plant species
YVY	*Arabidopsis thaliana*
YVYN	*Arabidopsis thaliana*
YVYH	*Arabidopsis thaliana*
YVYA	*Nicotiana tabacum*
YKYK	*Phaseolus vulgaris, Solanum tuberosum, Vigna unguiculata, Glycine max, Pisum sativum, Adiantum capillus-veneris, Catharanthus roseus, Nicotiana sylvestris, Daucus carota*
YKYS	*Phaseolus vulgaris, Pisum sativum, Catharanthus roseus*
YKYP	*Phaseolus vulgaris, Glycine max, Pisum sativum, Vicia faba*
YKYN	*Phaseolus vulgaris, Pisum sativum*
YKYY	*Lycospersicon esculenta*
YKYQ	*Pisum sativum*
YIYK	*Phaseolus vulgaris, Vigna unguiculata, Glycine Max, Cicer arietinum, Arabidopsis thaliana, Lupinus angustifolius, Adiantum capillus-veneris, Catharanthus roseus*
YIYA	*Phaseolus vulgaris, Glycine max, Manihot esculenta, Cicer arietinum, Arabidopsis thaliana, Pisum sativum*
YIYS	*Phaseolus vulgaris, Glycine max, Bromheadia finlaysonia, Arabidopsis thaliana*
YIYG	*Lycospersicon esculenta*
YIYN	*Arabidopsis thaliana*
YLYK	*Arabidopsis thaliana, Lupinus angustifolius, Adiantum capillus-veneris, Nicotiana tabacum*
YLYS	*Cicer arietinum, Arabidopsis thaliana, Pisum sativum, Lupinus angustifolius*
YLYN	*Vigna unguiculata*
YLYT	*Lycospersicon esculenta*
YLYA	*Nicotiana tabacum*
YSYS	*Phaseoulus vulgaris, Vicia faba*
YSYA	*Phaseoulus vulgaris*
YSYD	*Lycospersicon esculenta*
YSYN	*Oryza sativa*
YSYT	*Daucus carota*
YPYL	*Vigna unguiculata*
YPYT	*Arabidopsis thaliana*
YPYS	*Arabidopsis thaliana*
YDYT	*Lycospersicon esculenta*
YDYN	*Arabidopsis thaliana*
YEYK	*Brassica napus, Vigna unguiculata, Arabidopsis thaliana*
YEYS	*Arabidopsis thaliana, Oryza sativa*
YGYT	*Zea diploperennis*
YGYG	*Zea mays*
YQYK	*Vigna unguiculata, Adiantum capillus-veneris*
YQYS	*Nicotiana tabacum*
YAYK	*Lupinus angustifolius*
YFYS	*Brassica napus, Adiantum capillus-veneris*
YMYK	*Brassica napus*
YNYS	*Arabidopsis thaliana*
YTYS	*Lycospersicon esculenta, Daucus carota*

[075] Because endogenous P3 extensins are normally insoluble *in muro*, their precursor forms have not been isolated. Thus, P3 glycosylation profiles can only be inferred from P3 extensin genes (Zhou, Rumeau, and Showalter 5-17) and from Hyp-glycoside profiles of P3 glycopeptides enzymically released from the walls (Lamport 79-115). The Hyp-contiguity hypothesis (Kieliszewski et al. 2541-49; Kieliszewski and Lamport 157-72; Shpak et al. 11272-78; Shpak, Leykam, and Kieliszewksi 14736-41) predicts that P3 extensins, with their abundant contiguous Hyp and nonclustered single Hyp residues, should be extensively arabinosylated mainly with tetra- and triarabinooligosaccharides but lack arabinogalactan polysaccharides (Tables 2 and 3). The monsaccharide compositions and Hyp-glycoside profiles of YK20-EGFP, YK8-EGFP, YL8-EGFP, and FK9-EGFP were consistent with earlier P3 glycopeptide profiles (Lamport 1155-63) and with the predictions for Hyp-glycosylation, the exception being a small amount of Hyp-polysaccharide in YL8-EGFP (Table 3). Judging by earlier Hyp-glycosylation profiles of HRGP analogs having only tandem repeats of the pentapeptide (Ser-Hyp-Hyp-Hyp-Hyp)$_n$ or the dipeptide (Ser-Hyp)$_n$ (Shpak et al. 11272-78;Shpak, Leykam, and Kieliszewksi 14736-41), it was infered that the arabinogalactan polysaccharide probably occurs on 2 out of the 8 lone Hyp residues occupying the center (underlined) of the palindromic repeats: Ser-Hyp-Hyp-Hyp-Hyp-Ser-<u>Hyp</u>-Ser-Hyp-Hyp-Hyp-Hyp-Tyr-Tyr-Tyr-Leu.

[076] As essentially all of the Pro residues in the isolated modules were hydroxylated (Table 4), it was determined that YK20 contained about 180 Hyp residues, YK8 and YL8 contained about 72, and FK9 about 81. Of the total Hyp residues in each glycoprotein, only 3-8 percent (i.e. 3 to 7 Hyp residues/glycoprotein) were not glycosylated, which indicated that often all Hyp residues in the 16-residue

repeats were arabinosylated, including the non-contiguous Hyp. This is consistent with earlier work which demonstrated that although contiguous Hyp residues are preferred arabinosylation sites, non-clustered, lone Hyp residues can be arabinosylated or remain non-glycosylated (Kieliszewski et al. 2541-49).

[077] In addition to arabinosides, extensins generally contain monogalactosylated Ser (Lamport 1155-63;Lamport, Katona, and Roerig 125-31;Smith, Muldoon, and Lamport 1233-39) and this seems to be true for YK20, YK8, FK9, and probably for YL8, although at least some of the Gal in YL8 occurs in the occasional arabinogalactan adduct present. The monosaccharide composition of YK20, YK8, and FK9 shows they contained about 1 mole of Gal for every 10-11 moles of Ara, which suggested monogalactose occurred on about 72-90% of the Ser residues in the P3-type module (Tables 7 and 9). This also is in agreement with native P3 glycopeptides isolated and characterized earlier from tomato cell walls (Lamport 27-31).

***In vitro* crosslinking of the P3 extensin glycomodules**

[078] It was discovered that YK20 was an excellent crosslinking substrate for the well-characterized tomato pI 4.6 extensin peroxidase (Schnabelrauch et al. 477-89), with rates only somewhat lower than those for tomato extensin P1, which is presumably a natural substrate of the enzyme (Figs. 25 and 31). Earlier it was proposed that valine, as in the putative P1 crosslinking motif Val-Tyr-Lys, was a requirement for intermolecular crosslinking of extensins by the pI 4.6 extensin peroxidase (Kieliszewski and Lamport 157-72). However, YK20 contained no valine.

[079] Without wishing to be bound by any particular theory, the underlying principle driving both P3 and P1 self-assembly/molecular recognition for subsequent crosslinking may involve the inherent hydrophobicity of the putative crosslinking sites

(Brady, Sadler, and Fry 323-27;Whitesides, Mathias, and Seto 1312-19), Val-Tyr-Lys in P1 and Tyr-Tyr-Tyr in the P3 motif. Polymers containing separate hydrophilic and hydrophobic regions, a characteristic of the extensins, self-assemble in aqueous solution largely due to the hydrophobic effect which drives non-polar regions away from water and toward each other (Whitesides, Mathias, and Seto 1312-19). Thus, it is believed that P1, YK20, YK8, and YL8 crosslinking is facilitated by the alignment of the hydrophobic crosslinking motifs with one another, which might be further favored by the concurrent alignment of the highly regular and hydrophilic Ser-Hyp_4 glycomodules (Brady, Sadler, and Fry 323-27; Kieliszewski and Lamport 157-72). The rapid rate of P3 crosslinking may partially explain why no one has isolated an endogenous P3-type extensin monomeric precursor to the wall network: it is very rapidly insolubilized *in muro.* Consistent with this likelihood was the incorporation of YK20-EGFP into the tobacco cell wall, presumably through crosslinking as the walls retained green fluorescence after plasmolysis in mannitol or in salt at concentrations that normally remove non-covelently bound material from the wall (Fig. 16). Since both YK8-EGFP and YL8-EGFP are able to undergo *in vitro* crosslinking, it had been expected to see them in the wall as well. This could be due to their size, which has an impact on crosslinking rates (Fig. 31). It is possible that such a small amount of YK8-EGFP and YL8-EGFP is actually crosslinked into the wall that visualization was not possible by fluorescence microscopy, as the relatively low pH of the cell wall could have significantly decreased EGFP fluorescence (CLONTECH).

[080] Two other parameters appeared to influence the crosslinking rates of the P3 analogs. The first is molecular length of the substrate. The rates of YK20 crosslinking, with 320 amino acids per molecule, were much greater than those of YK8, which possessed only 128 amino acids/molecule although both had the same

number of potential crosslink sites per milligram (~ 0.17 mole/mg) of material and similar glycosylation profiles (Tables 2 and 3). Secondly, the presence of lysine in the Tyr-Tyr-Tyr-Lys motif favored crosslinking judging by the rates for YK8, which were double those of YL8. It is possible that the lysine residue facilitates enzyme recognition of the crosslink site or assists catalysis.

Identification of the P3 intermolecular crosslink

[081] The question of the crosslink identity arises and the enzyme or enzymes involved. Given the number of peroxidases in the extracellular matrix, including those that crosslink extensin *in vitro* (Jackson et al. 1065-76); (Price et al. 41389-99), it is likely that multiple extensin peroxidases exist having different affinities for different types of extensins and catalyzing a range of crosslinks. This might explain why peroxidases from tomato crosslink extensins at different rates: the basic peroxidase characterized earlier by Everdeen and colleagues (Everdeen et al. 616-21)) has a specific activity that is 2 to 3 orders of magnitude lower for P1 than the acidic pI 4.6 extensin peroxidase characterized later (Fig. 32) (Schnabelrauch et al. 477-89). All of the P3 analogs that contained tyrosine could be crosslinked by extensin peroxidase while FK9, which lacked tyrosine but had lysine, did not crosslink at all (Figs. 26 and 31). Thus, in the modules tested here, Tyr, but not Lys, is directly involved in crosslinking. These results also indicated that the Val-Tyr-Lys motif is not required for all extensin crosslinking by the pI 4.6 extensin peroxidase (Schnabelrauch et al. 477-89) and suggest that different intramolecular crosslinks occur in extensin as the motif present in P1 (putatively Val-Tyr-Lys) differs substantially from the P3 motif (Tyr-Tyr-Tyr).

[082] The exclusive involvement of Tyr in crosslinking of the P3 analogs described here raises questions about the involvement of the crosslink amino acid

diisodityrosine (Brady, Sadler, and Fry 323-27). Characterized hydrolysates of crosslinked YK20 identified diisodityrosine as the crosslink amino acid produced during crosslinking (Fig. 33), judging by its atomic mass determined by MS (Fig. 34) and other properties (Brady and Fry 87-92) including NMR spectra of the isolated amino acid (Table 14 and Figs. 34 and 35). Thus, these results lend strong support to the suggestion of Brady *et al.* (Brady, Sadler, and Fry 323-27) that diisodityrosine crosslinks are a mechanism for extensin insolubilization *in muro.*

[083] Finally, this work demonstrates the general utility of a synthetic gene approach for elucidating the codes that direct the post-translational modifications of HRGPs, including glycosylation (Shpak et al. 11272-78;Shpak, Leykam, and Kieliszewksi 14736-41;Tan, Leykam, and Kieliszewski 1362-69) and crosslinking. The ramifications range beyond the roles of extensin in the wall to the module-by-module design of new glycoprotein-based biopolymers that combine the special properties conferred by glycosylation and crosslinking.

EXAMPLES

Example 1: Techical Basis for Crosslinking

Materials and Methods

Construction of P3-type synthetic genes

Oligonucleotide design

[084] The oligonucleotides coding for synthetic P3 extensin genes (Fig. 5) were designed taking into consideration known codon biases of native tobacco extensins (Showalter and Rumeau 247-81). Appropriate restriction sites for cloning and/or polymerization were designed into the oligonucleotides as described earlier (Lewis et al. 400-06;Shpak, Leykam, and Kieliszewski 14736-41). The oligonucleotide sequences were analyzed for secondary structure formation (i.e.

hairpin and dimer formation), annealing temperature optimum (TaOpt), melting temperature (Tm), and false priming sites using Primer Premier software (Biosoft International, Palo Alto, CA) (Shpak, Leykam, and Kieliszewski 14736-41). The oligonucleotide sets were ordered from Integrated DNA Technologies (IDT, Coralville, IA).

Construction of the *FK9* synthetic gene

Annealing of the *FK oligonucleotide pair*

[085] Oligonucleotides for the *sense* and *antisense FK internal repeats* (*FK-s* and *FK-as* respectively; Fig. 5a) were each dissolved to 0.5 mM in sterile distilled, deionized water (ddH_2O). The *FK-s* and *FK-as* oligonucleotides (1.5 nmol each) were combined in 30 µl (total volume) of 1x T4 DNA ligase buffer (Promega, Madison, WI) and placed in a heating block at 97° C for 5 min, slowly cooled to 60-65° C over a period of 30 min, then cooled further to 55-60° C over 30 min. The heating block was stabilized at the optimum annealing temperature (58° C, calculated using Primer Premier software) for 1 h. The annealed *FK internal repeat pairs* were stored at -20° C.

Annealing of the *5' linker oligonucleotide pairs*

[086] The *5' linker pair* contained unique restriction sites for subcloning and was prepared by the annealing of oligonucleotides (*5'-SP3-s*) and (*5'-SP3-as*) as previously described (Shpak, Leykam, and Kieliszewski 14736-41). Briefly, 4 nmol each of (*5'-SP3-s*) and (*5'-SP3-as*) were combined in 1x T4 DNA ligase buffer to a final volume of 80 µl. The reaction was heated to 95° C for 5 min, cooled over 1.15 h to 47° C, then held at 47° C for 1.3 h. The annealed *5' linker pairs* were cooled to room temperature then stored at -20° C.

Ligation of the *5' linker* to the annealed *FK internal repeat pair*

[087] The annealed *FK internal repeat pairs* (0.5 nmol) were combined with the annealed *5' linker pairs* (0.25 nmol) in 1x T4 DNA ligase buffer in a sterile screw cap microtube. The sample was heated to 68° C for 25 min, then 45° C for 5 min. After cooling to room temperature, 3 U of T4 DNA ligase (Promega) was added to the sample (40 µl final vol.). The reaction was incubated at ambient temperature for 3.6 h, and then cooled to 4° C for 4.2 h. An additional 1.5 µl (0.075 nmol) of the annealed *5' linker pairs* was added and the reaction mixture was incubated further at 4° C overnight (~14 h). The extent of polymerization was assayed by agarose gel electrophoresis. Excess *5' linker pairs* were removed by Sephacryl S-400 microspin column purification according to the manufacturer's instructions (Amersham Pharmacia Biotech, Piscataway, NJ). Column eluates yielded *5' linker-FK internal repeats*.

[088] To maximize the number of contiguous *FK internal repeat pairs* ligated to each *5' linker pair*, an additional 0.4 nmol of *FK internal repeat pairs* was added to the purified *5' linker-FK internal repeats* along with ligase buffer. To re-initiate elongation, the reaction was heated to just under the Tm (70° C) for 10 min, followed by 60° C for 10 min, then 50° C for 3 h, and finally cooled to room temperature. DNA Ligase (3U, Promega #M1801) was added in a final volume of 50 µl. The reaction was incubated 30 min at ambient temperature then held at 4° C overnight.

Ligation of the annealed *3' linker pair* to *5' linker-FK internal repeats*

[089] The annealed *3' linker pair* (Fig. 5a) was prepared as described for the *5'* linker pair. The *5' linker-FK internal repeats* (1.17 nmol) was heated to 75° C for 10 min. The annealed *3' linker pair* (0.35 nmol) was added to the tube while still

at 75° C then returned to the heat block for an additional 10 min. The sample was slowly cooled, then held at 50° C for 3 h. DNA ligase (3 U) was added at room temperature where it was incubated 3 h before being transferred to the refrigerator (4° C). Unincorporated *3' linker pairs* were removed by Sephacryl S-400 microspin columns (Amersham Pharmacia). The purified *5' linker-FK internal repeats-3' linker* synthetic gene was eluted from the column and stored at -20° C.

Creation of plasmid *pUC18-FK9*

[090] The *5' linker-FK internal repeats-3' linker* synthetic gene was subcloned into a *pUC18 plasmid* vector as a *BamHI-EcoRI* fragment (Fig. 6) (Shpak, Leykam, and Kieliszewski 14736-41). Plasmids were transformed into competent *E. coli* (XL1-Blue, Stratagene La Jolla, CA) and selected for ampicillin resistance. Positive transformants were cultured in 3.5 ml LB medium liquid cultures supplemented 50 μg/ml with ampicillin (220 rpm, 37° C), from which *pUC18-FK internal repeats* plasmids were isolated using the Wizard Plus Minipreps, DNA Purification System (Promega, Madison, WI). Plasmids were screened for insert size by *XmaI-NcoI* digestion followed by 1% (w/v) agarose gel electrophoresis. A plasmid having nine contiguous *FK internal repeats* was selected (*FK9*), and sequenced.

Construction of *YK*, *YL*, and *FL* synthetic genes

[091] A second method was used to build synthetic genes more similar in size to native extensins. This method was adapted from an existing method (Lewis et al. 400-06) and was used to build *YK*, *YL*, and *FL* gene constructs from new sets of oligonucleotide pairs (Fig. 5B, C, and D). Restriction sites for *BbsI, BsmFI, XmaI, NcoI*, and *SacI* were engineered at the 5' and 3' ends of these new oligonucleotides for subcloning and polymerization (Fig. 5B, C, and D).

[092] Dried oligonucleotides were dissolved in ddH_2O to a concentration of 1 mg/ml. Corresponding *sense (ss)* and *antisense (as) oligonucleotide pairs* (34 pmol each) were combined, and diluted to 20 μl with ddH_2O. Oligonucleotide pairs were heat denatured at 94° C for 7 min, then annealed to their respective complimentary 24 base pair annealing sites (60° C for 5 min, followed by 45° C for 10 min; Fig. 5B, C, and D underlined). Samples were cooled to room temperature and primer extension was performed using DNA Polymerase I (Klenow) large fragment (Promega) according to manufacturers instructions (room temperature for 15 min, then heat stopped at 75° C for 10 min). Reactions were loaded onto Sephacryl S-200 microspin purification columns (Amersham Pharmacia Biotech) to change the buffer and remove unincorporated nucleotides. Column eluates yielded duplex DNAs designated *YK, YL*, and *FL* (Fig. 7).

Creation of plasmids *pUC18-YK*, *pUC18-YL*, and *pUC18-FL*

[093] The *YK, YL, and FL* duplex DNAs were each digested with *XmaI-SacI,* followed by Sephacyl S-400 microspin column purification (Amersham Pharmacia Biotech) to remove the small restriction fragment ends (retained on the columns) from duplex DNA monomers. The *YK, YL,* and *FL XmaI-SacI* fragments were subcloned to *pUC18* to yield plasmids *pUC18-YK, pUC18-YL,* and *pUC18-FL* (Fig. 8). Plasmids were sequenced with an *M13 forward primer.*

Creation of plasmids *pUC18-YK20*, *pUC18-YK8*, *pUC18-YL20*, *pUC18-YL8*, and *pUC18-FL8*

[094] Non-regenerable, complimentary sticky ends were produced by separate *BsmFI-ScaI* and *BbsI-ScaI* restrictions (New England Biolabs, Beverly, MA) of plasmid preparations harboring one repeat of the synthetic genes (Fig. 7). Ligation of the insert containing *BsmFI-ScaI* fragment with the insert containing *BbsI-*

ScaI fragment produced plasmids possessing 2 repeats designated *pUC18-YK2* and *pUC18-YL2* respectively. This process was repeated to achieve plasmids of 8 and 20 internal repeats (*pUC18-YK20, pUC18-YK8,* and *pUC18-YL8*).

Creation of plasmids *pUC18-SStob-YK20-EGFP, pUC18-SStob-YK8-EGFP, pUC18-SStob-YL20-EGFP, pUC18-SStob-YL8-EGFP, pUC18-SStob-FL8-EGFP and pUC18-SStob-FK9-EGFP*

[095] Synthetic genes YK20, YK8, YL20, YL8, FL8 and FK9 were inserted between a tobacco extensin signal sequence (SStob) and the enhanced green fluorescent protein gene (EGFP) as XmaI-NcoI fragments in the modified pUC plasmid (Fig. 9; modified pUC plasmid courtesy of Dr. Elena Shpak) (De Loose et al. 95-100;Shpak, Leykam, and Kieliszewski 14736-41;Tan L 1362-69). Plasmids modified pUC-SStob-YK20-EGFP, modified pUC-SStob-YK8-EGFP, modified pUC-SStob-YL20-EGFP, modified pUC-SStob-YL8-EGFP, modified pUC-SStob-FL8-EGFP, and modified pUC-SStob-FK9-EGFP were thus created. These plasmids were again sequenced from the 5' end using the M13 forward primer and from the 3' end using the 566asEGFP primer (Integrated DNA Technologies).

Creation of plasmids *pBI121-SStob-YK20-EGFP, pBI121-SStob-YK8-EGFP, pBI121-SStob-YL20-EGFP, pBI121-SStob-YL8-EGFP, pBI121-SStob-FL8-EGFP* and *pBI121-SStob-FK9-EGFP*

[096] The synthetic genes were subcloned into the binary plant transformation vector *pBI121* as *BamHI-SacI* fragments replacing the β-glucuronidase reporter gene to form plasmids *pBI121-SStob-YK20-EGFP, pBI121-SStob-YK8-EGFP, pBI121-SStob-YL20-EGFP, pBI121-SStob-YL8-EGFP, pBI121-SStob-FL8-EGFP,* and *pBI121-SStob-FK9-EGFP* (Fig. 10) which were then transformed to competent *E. coli.* Positive transformants were selected on solid LB

plates supplemented with kanamycin (30 μg/ml). Plasmids were prepared from 3.5 ml liquid LB culture supplemented with 30 μg/ml kanamycin. The control plasmid *pBI121-SStob-EGFP,* created by Dr. Elena Shpak, was also freshly prepared. Plasmids were screened to confirm insert size by both *BamHI-SacI* digestion and *XmaI-NcoI* digestions followed by agarose gel electrophoresis (0.7% w/v). Plasmids that yielded bands of predicted size in both digestions were selected for transformation to *Agrobacterium tumefaciens.*

DNA sequencing of synthetic genes

DNA sequencing of synthetic genes using the *M13 forward primer*

[097] We verified insert sequences of pUC18-YK, pUC18-YL, pUC18-FL, pUC18-FK9, modified pUC-SStob-YK20-EGFP, modified pUC-SStob-YK8-EGFP, modified pUC-SStob-YL20-EGFP, modified pUC-SStob-YL8-EGFP, modified pUC-SStob-FL8-EGFP and modified pUC-SStob-FK9-EGFP by Big Dye Terminator DNA sequencing (Applied Biosystems, Foster City, CA) using the M13 forward primer (Fig. 11). Cycle sequencing reactions were prepared according to manufacturer's instructions and cycled using either a Stratagene-Robocyler Gradient 40-Thermal Cycler or an Applied Biosystems GeneAmp PCR system 2400. The tubes were placed in the thermocycler and incubated as in Table 1.

Table 1. Thermocycler program for DNA sequencing using the *M13 forward* primer.

Step	Hold 1	Hold 2	Hold 3	Hold 4	Cycles
1	98° C / 5 min	-	-	-	1
2	98° C / 30 sec	50° C / 15 sec	60° C / 4 min	-	25
3	-	-	-	4° C / 99:99 min	-

[098] Sequencing reactions were transferred to a sterile eppendorf tube and adjusted to 20 μl total volume with ddH_2O. Sequencing products were ethanol precipitated by combining 20 μl of sequencing product with 16 μl of sterile water, and

64 μl of 95% non-denatured, ACS grade ethanol. The precipitation was incubated at room temperature for 15 min then centrifuged at room temperature for 20 min at 12,000 x *g*. The supernatants were removed and discarded. The pellets were washed with 250 μl of 70% ethanol then centrifuged 10 min at 12,000 x *g*. Sequencing products were dried in a heat block at 90° C for 1 min then delivered for sequencing to Dr. Morgan Vis of the Automatic DNA Sequencing Facility of the Department of Environment and Plant Biology, Ohio University.

DNA sequencing of synthetic genes using the *EGFP566as primer*

[099] The plasmids pUC18-SStob-YK20-EGFP, pUC18-SStob-YK8-EGFP pUC18-SStob-YL20-EGFP, pUC18-SStob-YL8-EGFP, pUC18-SStob-FL8-EGFP, and pUC18-SStob-FK9-EGFP were sequenced from the 3' end using the EGFP566as primer (Fig. 11; Integrated DNA Technologies) to verify in-frame ligation of synthetic genes between SStob and EGFP. Sequencing reactions using the 566asEGFP primer were performed as described for the M13 forward primer except the annealing temperature (Table 1-Step 2, Hold 2) was lowered to 47° C.

Transformation of *Agrobacterium tumefaciens* and Tobacco cells

Transformation of *Agrobacterium tumefaciens*

[0100] We transformed Agrobacterium tumefaciens (strain LBA4404) with plasmids pBI121-SStob-FK9-EGFP, pBI121-SStob-YK20-EGFP, pBI121-SStob-YK8-EGFP, pBI121-SStob-YL20-EGFP pBI121-SStob-YL8-EGFP, pBI121-SStob-FL8-EGFP, and pBI121-SStob-EGFP by the freeze-thaw method (An et al. 1-19). The plasmids were isolated from LB liquid cultures (3-5 ml) supplemented with kanamycin (30 μg/ml) using Wizard minipreps (Promega). Approximately 35-40 μl of plasmid was added directly to 100 μl of competent, frozen Agrobacterium. The mixture was thawed, then quickly frozen in N_2 (l). The frozen tubes were heated to

37° C for 5 min. LB was then added to a final volume of 1 ml and the tubes were incubated at 28° C for 2-4 h with shaking at ~100 rpm. The cells were then plated onto solid LB plates supplemented with kanamycin (30 μg/ml) and streptomycin (40 μg/ml). The plates were incubated in the dark at 28° C for 2-3 days. Positive transformants were picked and cultured in 3.5-5 ml of LB medium having 30 μg/ml kanamycin and 40 μg/ml streptomycin at 28° C with gentle shaking at (100-175 rpm) for about 16-20 h. The cultures were then pelleted by centrifugation (1000 x g for 5-10 min) and the supernatants removed and discarded. The pellets were resuspended in 1 ml of fresh LB medium.

Transformation of *Nicotiana tabacum* bright yellow-2 (BY-2) suspension cultures

[0101] Tobacco cell-suspension cultures (*Nicotiana tabacum*, BY-2) were transformed by *Agrobacterium* infection as previously detailed (McCormick et al. 81-84). Untransformed tobacco BY-2 cells were cultured (500 ml in a 1 L flask) for 4-5 days on a gyrotary shaker (88-94 rpm) at room temperature in NT-1 medium. The flask of cells was removed from the shaker and the cells were settled. Tobacco cells (10 ml) were transferred under strerile conditions to empty Petri dishes. From section 2.2.1, 100 μl of the transformed *Agrobacterium* suspension was mixed with the cells. The cultures were parafilmed and incubated in the dark for 2-3 days at 28° C. Excess *A. tumefaciens* was removed from the cells by washing 3 times with 10-15 ml of sterile culture medium either NT-1 medium, Schenk and Hildebrandt medium (SH), or modified SH medium (SH*, SH sans kinentin). Washed tobacco cells were then plated on solid medium (phytagel 2 g/L in either NT-1, SH, or SH*) supplemented with kanamycin (200 μg/ml) for selection and timentin (400 μg/ml) to kill the *A. tumefaciens*. Tobacco transformants appeared as small bumps on the

plates after 3-5 weeks. These bumps were chosen for further propagation on kanamycin and timentin supplemented plates (2-4 generations) to ensure complete eradication of the *A. tumefaciens* before the cells were transferred to plates having only kanamycin.

Identification of cell lines expressing synthetic genes and localization of synthetic gene products

Fluorescence microscopy

[0102] Aliquots (5-10 ml) of suspension cultured cells were withdrawn directly from liquid cell cultures (5-20 d). Expression of EGFP in transformed cell lines was viewed using a Zeiss LSM 510 confocal laser scanning microscope set for EGFP visualization (488 nm excitation; 510 nm emission).

Plasmolysis of transformed and untransformed tobacco cells for visualization of EGFP by fluorescence microscopy

[0103] Suspension cultured YK20-EGFP, YK8-EGFP, YL8-EGFP, and FK9-EGFP, and SS-EGFP tobacco cells (5-20 d) were pipeted (10 ml) to 15 ml graduated tubes. The cells were centrifuged and the packed cell volume was recorded. The cells were plasmolyzed by the addition of either of 1 M potassium phosphate buffer pH 7 to a final concentration of 500 mM or 1 M mannitol to a final concentration of 750 mM. The plasmolyzed cells were viewed by fluorescence microscopy.

Verification of transgene insertion by PCR

Isolation of genomic DNA from suspension-cultured tobacco cells

[0104] Genomic DNA was prepared from tobacco suspension cultured cells using the DNAzol ES method (Molecular Research Center, Inc. Cincinnati, OH) according to manufacturer's instructions. Purity and concentration were estimated by their 260/280 nm ratio and 260 nm absorbances respectively.

PCR amplification of transgenes from genomic DNA

[0105] Insertion of our transgenes *SStob-YK20-EGFP* and *SStob-YL8-EGFP* into tobacco was verified by PCR amplification using *SSeq4Sen* (TaOpt = 43° C; Tm = 68° C) and *EGFP566as* (TaOpt = 47° C; Tm = 73.1° C) primers (Fig. 11). Master mixes were created so that each reaction (25 μl) contained 12.5 μl of 2X Promega PCR master mix (Promega), 0.125 nmol of *SSeq4Sen* primer, 0.125 nmol *EGFP566as* primer, 10% (v/v) glycerol, and 0.14 - 0.21 ng of genomic template DNA. Due to a high %GC content in both primer and template DNA, a preincubation step of 98° C for 5 min was employed. The optimized PCR program is shown in Table 2. After completion of the PCR cycles, the reactions were separated by 1% agarose gel electrophoresis, eluted from the gel, and DNA sequenced using the *SSeq4Sen* oligonucleotide primer.

Table 2. Thermocycler program for synthetic gene amplification by PCR. PCR amplicfication reactions were performed using primers SSeq4Sen and EGFP566as on an Applied Biosystems GeneAmp PCR system 2400.

Step	Hold 1	Hold 2	Hold 3	Cycles
1	98° C/5min	-	-	1
2	98° C/30sec	55° C/1min	74° C/3min	25
3	-	-	74° C/5min	1

Isolation and purification of synthetic gene products

Liquid cell culture

[0106] Transformed tobacco cells were cultured in 500 ml of SH* media supplemented with kanamycin (150 μg/ml) in 1L Erlenmeyer flasks at ambient temperatures on a gyrotary shaker set to 88-94 rpm. Cells were subcultured every 14-21 days (Fig. 12).

Culture medium harvest

[0107] Liquid cell cultures (10-21 days) were filtered on a sintered glass funnel (Pyrex 40-60 ASTM). The culture media was collected, concentrated by

rotary evaporation to 50-250 ml, then dialyzed in 3,500 Da molecular weight cut-off (MWCO) dialysis tubing (Spectrum Laboratories, Inc., Rancho Domingo, CA) against ddH_2O at 4° C for 36 h.

Isolation of fusion glycoproteins by hydrophobic interaction chromatography

[0108] Dialyzed medium was re-concentrated by rotary evaporation to 50-250 ml, adjusted to 2 M NaCl (aq) then centrifuged at 20,000 x *g* for 20 min. Supernatants were loaded onto a Butyl-Sepharose 4 Fast Flow hydrophobic interaction chromatography (HIC) column (Fig. 12; 1.6 cm i.d. x 40 cm, Amersham Pharmacia Biotech) equilibrated in 2 M NaCl (aq). Elution proceeded stepwise from 2 M NaCl (aq) to 1 M NaCl (aq), and finally to ddH_2O at flow rates of 1-2 ml/min. Visibly green fractions eluting in 1 M NaCl were collected, dialyzed against ddH_2O (3,500 Da MWCO or 6-8,000 Da MWCO dialysis tubing for 36 h at 4° C; Spectrum Laboratories, Inc.) and then lyophilized.

Fractionation of fusion glycoproteins by gel filtration chromatography

[0109] The HIC fractionated fusion glycoproteins were dissolved ~30 mg/ml in 200 mM sodium phosphate, 0.005% (w/v) sodium azide, pH 7 (Superose buffer) then chromatographed on either a semi-preparative Superose-12 column or Superose-6 gel filtration column (Fig. 12; 16 mm i.d. x 500 mm, Amersham Pharmacia Biotech) equilibrated in Superose buffer, pH 7. Spectrophotometric detection was monitored at 220 nm. Flow rates were between 0.5-1 ml/min. Visibly green fractions corresponding to peaks absorbing at 220 nm were pooled and either stored at -20° C or dialyzed, lyophilized, then stored at -20° C.

Isolation of fusion glycoproteins by reversed phase HPLC

[0110] The fusion glycoproteins isolated after gel filtration chromatography typically were directly injected onto a semi-preparative C4 reversed-phase HPLC

column (Fig. 12; 10 mm i.d. x 250 mm, Vydac, Hesperia, CA) equilibrated in 0.1% (v/v) TFA (aq) and eluted with a linear gradient to 100% end buffer (0.1% (v/v) TFA (aq) in 80% (v/v) acetonitrile) at 2 ml/min over 60 min. UV detection was monitored at 220 nm. The major peak absorbing at 220 nm corresponded to each fusion glycoprotein and was collected and lyophilized. A Hewlett-Packard series 1050 HPLC was used for purification chromatography.

Removal of the EGFP tag by tryptic digestion

[0111] Fusion glycoproteins (5-100 mg) were dissolved in water to 28.6 mg/ml and heated to 100° C for 2-3 min. Once cool, samples were diluted to 5 mg/ml in buffer containing 10 mM $CaCl_2$ (aq) and 2% (w/v) NH_4HCO_3 (aq). Trypsin was added 100:1 (w/w) substrate to enzyme. Reactions were incubated 16-24 h at room temperature with constant stirring. The digestion reactions were lyophilized, then fractionated by Superose-12 gel filtration chromatography and C4 reversed phase chromatography as described above. Peaks corresponding to YK20, YK8, YL8, and FK9 glycomodules (designated YK20, YK8, YL8, and FK9, respectively) were collected and lyophilized. The purified glycomodules were used for amino acid composition analyses and as substrates for *in vitro* crosslinking.

Anhydrous hydrogen fluoride (HF) deglycosylation of fusion glycoproteins

[0112] The dried fusion glycoproteins YK20-EGFP, YK8-EGFP, and YL8-EGFP (5-10 mg) were deglycosylated with anhydrous HF as described earlier (Sanger and Lamport 66-70). Briefly, anhydrous HF containing 10% (v/v) anhydrous methanol at 0° C was added (2-10 mg/ml) to fusion glycoprotein samples. Reactions were incubated on ice for 1 h with constant stirring then quenched by adding them to 6 volumes of ice-cold ddH_2O. The deglycosylated samples were dialyzed at 4° C against ddH_2O, then lyophilized. Dried, deglycosylated fusion glycoproteins

designated dYK20-EGFP, dYK8-EGFP, and dYL8-EGFP were isolated by analytical C4 reversed phase chromatography as described above for semi-preparative C4 HPLC, except the gradient was eluted over 40 min.

Characterization of transgene products

Amino acid composition analyses

[0113] Glycomodules YK20, YK8, and YL8, and crosslinked YK20 (YK20XL) were analyzed for amino acid composition as described by (Bergman, Carlquist, and Jornvall 45-55). Briefly, samples were hydrolyzed at 110° C for 18-24 h in constant boiling 6 N HCl (aq) (Pierce, Rockford, IL) containing 10 mM phenol (Pierce). Hydrolyzed samples were cooled to room temperature then dried under a stream of N_2 (g). Hydrolysates (20 μg) were derivatized with phenylisothiocyanate (PITC) for 15-30 min to form phenylthiocarbamyl-amino acid (PTC-aa) derivatives. The PTC-aa derivatives were separated by reversed phase HPLC on a Prodigy ODS (3) C18 analytical column (4.6 mm i.d. x 150 mm, 3μm particle size, 10 nm pore size, Phenomenex, Torrance, CA) with UV detection at 254 nm. The IDT quantity was determined from the same hydrolysates using a separate assay described earlier (Schnabelrauch et al. 477-89).

[0114] The amino acid composition of FK9 was determined by Dr. Joseph F. Leykam of the Macromolecular Facility at Michigan State University, East Lansing, MI.

Partial N-terminal sequencing

[0115] Partial N-terminal sequences of the fusion glycoprotein FK9-EGFP and anhydrous HF deglycosylated fusion proteins dYK20-EGFP, dYK8-EGFP, and dYL8-EGFP were obtained by Joseph F. Leykam on a 477A Applied Biosystems, Inc. gas-phase sequencer by automated Edman degradation.

Colorimetric hydroxyproline estimation

[0116] Samples (100-1000 μg) were hydrolyzed in 6 N HCl (aq) (not more than 4 mg/ml) at 110° C for 18-24 h. Hydrolysates were dried under N_2 (g) then redissolved in 0.5 ml distilled deionized water. Hydroxyproline estimation was determined by colorimetric assay as described earlier (Lamport and Miller 454-56;Lamport and Northcote 665-66).

Detection and quantification of isodityrosine (IDT)

[0117] Acid hydrolysates of YK20, YK8, YL8, FK9, and YK20XL were fractionated (10 μg each) on a PolyHYDROXYETHYL A column (9.4 mm i.d. x 200 mm, 5 μm particle size, 10 nm pore size, PolyLC Inc., Columbia, MD) eluted isocratically in size exclusion mode (SEC) with 50mM formic acid at 0.8 ml/min. UV detection was monitored at 280 nm for tyrosine absorbance (Schnabelrauch et al. 477-89). IDT was partially identified by co-chromatography with an authentic external IDT standard (courtesy of Dr. Derek T.A. Lamport). A three level calibration was performed using peak areas of known amounts of the external IDT standard, from which an average molar response factor was calculated. This response factor was used to quantify IDT from peak areas of our transgene products.

Hydroxyproline-O-glycoside profiles of the transgene products

[0118] We determined the Hyp-O-glycosylation profiles of YK20-EGFP, YK8-EGFP, YL8-EGFP, and FK9-EGFP as previously described (Lamport and Miller 454-56). The fusion glycoproteins YK20-EGFP (2.20 mg), YK8-EGFP (4.10 mg), YL8-EGFP (2.25 mg), and FK9-EGFP (2.21 mg) were each dissolved to 5 mg/ml in 0.44 N NaOH (aq) and hydrolyzed at 105° C for 18 h. Samples were cooled on ice, neutralized with ice-cold 1 M H_2SO_4 (aq), and then lyophilized. Dried hydrolysates were redissolved in ddH_2O, loaded onto a cation exchange column (6 mm i.d. x 750

mm, Chromobeads C-2, Technicon) equilibrated in water, and eluted with a linear gradient to 0.5 M HCl (aq). An automated, in-line post-column hydroxyproline assay provided the detection of Hyp-glycosides at 560 nm.

Neutral sugar composition analyses

[0119] The neutral sugar compositions of YK20-EGFP, YK8-EGFP, YL8-EGFP, FK9-EGFP, and YK20XL (100 μg each) were determined as alditol acetate derivatives (Albersheim et al. 340-45) using myo-inositol (50 nmol) as an internal standard as previously described (Shpak, Leykam, and Kieliszewski 14736-41). Briefly, samples containing internal standard were hydrolyzed in 200 μl of 2 N trifluoroacetic acid at 121° C for 1 h. Hydrolysates were dried under N_2 (g) at 50° C. Aldehyde groups of the free aldose sugars were reduced to alditols with sodium borohydride ($NaBH_4$ 20 mg/ml in 3 M NH_4OH) at room temperature for 1 h. This reaction was neutralized with glacial acetic acid, and dried thoroughly under N_2 (g) at 40° C then dessicated overnight. Samples were acetylated with acetic anhydride (121° C for 1 h.) then separated by gas chromatography (GC). GC was performed on an HP 5890 Series II GC using an HP-5 column (Crosslinked 5% PH ME Siloxane; 30m (L) X 0.32mm (I.D.) X 0.25μm film thickness; Hewlett Packard, USA) programmed from 130° C to 177° C at a rate of 1.2° C/min then from 177 to 200° C at a rate of 10° C/min.

Uronic acid estimations

[0120] Uronic acids were detected by a manual colorimetric assay (Blumenkrantz and Asboe-Hansen 484-89). The glycomodules YK20, YK8, YL8, FK9, YK20XL (100 μg) were dissolved in concentrated sulfuric acid having 12.5 M sodium tetraborate, and then cooled on ice for 5 min. The samples were hydrolyzed at 100° C for 5 min and again cooled on ice. The m-Hydroxydiphenyl reagent (150

mg m-hydroxydiphenyl in 100 ml 0.5% (w/v) NaOH) was reacted with the samples for 5 min at room temperature and then the 520 nm absorbances were recorded. Uronic acid quantities were estimated from the 520 nm absorbances of external galacturonic acid standards.

Confirmation of Hyp-O-arabinogalactan polysaccharide in YL8-EGFP by gel filtration of base hydrolysates

[0121] To confirm that YL8-EGFP contained O-linked Hyp arabinogalactan polysaccharide (Hyp-PS), 10.9 mg of reverse-phase purified YL8-EGFP was dissolved 5 mg/ml in 0.44N NaOH (aq) and hydrolyzed at 105° C for 18 h. This cleaves the peptide backbone, but leaves the Hyp-O-linked glycosides intact. The hydrolysate was cooled on ice, neutralized with 1 N H_2SO_4 (aq) to pH 8.3, then lyophilized. The dried hydrolysate was dissolved in 1 ml of 20% (v/v) acetonitrile (aq) and then centrifuged 20 min at 10,300 x *g*. Supernatants (500 µl) were separated on a Superdex-75 gel filtration column equilibrated and eluted isocratically (0.4 ml/min) in 20% (v/v) acetonitrile (aq). Fractions were collected (1 min/fraction), lyophilized, and then analyzed for hydroxyproline content (Lamport and Miller 454-56;Lamport and Northcote 665-66).

***In vitro* crosslinking of extensin precursors**

Isolation of the pI 4.6 extensin peroxidase from tomato

[0122] Untransformed Bonnie Best tomato cell suspension cultures (500ml) were grown 10-30 days in SH* medium. The spent culture medium was filtered on sintered glass funnels, immediately shell frozen (350 ml) and lyophilized. Dry culture medium was redissolved in ice-cold ddH_2O to 50 ml and ultra-centrifuged at 250,000 x *g* (60,000 rpm) for 4 h. at 10° C to remove pectin contaminants. Supernatants were adjusted to 95% saturation with ammonium sulfate (s), gently stirred at 4° C for

16-20h, and then centrifuged at 27,000 x *g* for 20 min. The pellets were dissolved in and dialyzed against 4 x 1 L volumes of ice-cold 20 mM sodium acetate buffer pH 6 for 18-24 h. at 4 °C in Spectra-Por 3,500 Da MWCO dialysis tubing (Spectrum Laboratories Inc.). Samples were centrifuged at 12,100 x *g* for 15 min at 4° C. Supernatants were injected onto a preparative DEAE Sepharose Fast Flow (Amersham Pharmacia) anion exchange column (14 mm i.d. x 9.6 cm) equilibrated in 20 mM sodium acetate buffer, pH 6 and eluted at a flow rate of 1 ml/min for 1 h. A linear gradient to 2M NaCl in 20mM sodium acetate buffer, pH 6 was then applied by a gradient maker. Fractions (1ml) were collected and assayed for peroxidase activity by 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) (ABTS) assay as described earlier (Everdeen et al. 616-21). Active fractions were pooled and concentrated by Centricon YM-30 centrifugal filter devices (Millipore, USA). Retentates were pooled, adjusted to 10% (v/v) glycerol, frozen with N_2 (l) in 100-250 µl aliquots, and stored at (-80)°C.

Purification of pI 4.6 extensin peroxidase from tomato by HPLC

[0123] Anionic peroxidase fractions from preparative DEAE anion exchange chromatography were further separated by normal-phase ion exchange HPLC on a DEAE-5PW column (Tosohaas, 7.5 mm i.d. x 750 mm). The column was equilibrated in 20 mM sodium acetate pH 6 and eluted by a linear gradient to 500 mM NaCl (aq) in 20 mM sodium acetate pH 6 over 100 minutes with a flow rate of 1 ml/min. Heme group absorbance, typical of class III plant peroxidases (Welinder et al. 6063-81), was monitored at 405 nm.

Quantification of peroxidase activity by 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) (ABTS) assay

[0124] Peroxidase activity was quantified by ABTS assay as described earlier (Everdeen et al. 616-21;Schnabelrauch et al. 477-89), whereby 12 ng of peroxidase will produce an absorbance change of 1 $A_{405\ nm}$ unit per minute.

***In vitro* crosslinking of P3 glycomodules with tomato extensin peroxidase**

[0125] Tomato P1 extensin was isolated as described earlier for use as a positive crosslinking control (Smith et al. 1021-30;Smith, Muldoon, and Lamport 1233-39). Reversed phase HPLC purified glycomodules YK20, YK8, YL8, and FK9 and native tomato P1 extensin were used as substrates for *in vitro* crosslinking reactions. Extensin substrate stock solutions (6 mg/ml) were prepared by diluting 10 mg/ml substrates in ddH_2O with McIlvaine buffer pH6. Crosslinking reactions were performed at room temperature by combining substrate (3 mg/ml), extensin peroxidase enzyme (1 ng per 60 μg of substrate), and H_2O_2 (60 μM). Reaction time was initiated upon addition of H_2O_2. The stopping reagent (50 mM 2-mercaptoethanol) was added prior to H_2O_2 in time zero reactions to prevent crosslinking from occurring. Reactions were stopped by the addition of 50 mM 2-mercaptoethanol (16.7 mM final concentration) after specified times.

[0126] Detection of *in vitro* crosslinking and rate estimations were provided by analytical Superose-6 gel filtration chromatography eluted isocratically in pH 7 Superose buffer (Amersham Pharmacia; 0.75 ml/min; OD 220nm detection). Crosslinking rates were calculated by the decrease in peak area of the monomer after 1 min of incubation versus an equal amount of the time zero reaction using the first order rate equation (Schnabelrauch et al. 477-89). Rates were expressed as the amount of substrate crosslinked (μg) per sec (Schnabelrauch et al. 477-89).

Detection of extensin peroxidase activity in tobacco cell suspension cultures

[0127] Untransformed tobacco BY-2 suspension cultured cells were grown 10 days in NT-1 culture medium to 17% PCV. The cells (100 ml) were filtered on a sintered glass funnel, washed with 1.25 L of water, and then eluted with 500 ml of 1 M KCl (aq) in 20 mM sodium acetate buffer, pH 6. This salt-eluted fraction was designated (E2). Spent culture medium (350 ml) was lyophilized then redissolved to 50 ml with ice-cold water. This crude medium fraction was designated (E1). Subsequent steps were kept at or below 4° C. The salt-eluate E2 sample was then dialyzed (24-36 h; 4° C; 3,500 Da MWCO tubing) against 20 mM sodium acetate buffer, pH 6. A 250 ml aliquot of the dialyzed E2 was lyophilized, and then redissolved to 1 ml with water. Enzyme activity was estimated by ABTS assay. The E1 and E2 samples were then examined for extensin crosslinking.

***In vitro* crosslinking with crude cationic tomato peroxidases**

[0128] Crude tomato culture medium was prepared as described for pI 4.6 extensin peroxidase preparation. The preparative DEAE void fraction (non-binding) was collected. An aliquot (10-15 ml) was concentrated to 3.2 ml using Centricon YM-30 centrifugal filter devices (Millipore). The concentrated void peroxidases were adjusted to 10% (v/v) glycerol and then stored at -80° C. The void peroxidases were quantified by ABTS assay then tested for *in vitro* crosslinking. Crosslinking reactions were incubated with substrates (3mg/ml), enzyme (1.7 ng/60μg substrate), and H_2O_2 (60 μM) for 0 and 30 min. The reactions were separated by gel filtration on an analytical Superose-6 column (Amersham Pharmacia) eluted at 0.5 ml/min with UV detection at 220 nm.

Analysis of crosslinked extensin products

Isolation of diisodityrosine (di-IDT) from untransformed tomato cell walls

Isolation of cell walls

[0129] Untransformed Bonnie Best tomato cell suspension cultures were filtered on a sintered glass funnel. The cell walls were isolated as described earlier (Lamport 151-218;Lamport and Northcote 52P).

Acid hydrolysis of untransformed tomato cell walls

[0130] In a closed reflux apparatus, 1.55 g of untransformed tomato cell walls were added to 465 ml of 6 N HCl (aq) containing 10mM phenol (Pierce). The hydrolysis vessel was purged of air by vacuum then flushed with N_2 (g). The cell walls were refluxed for 24 h then cooled. The hydrolysate was concentrated by rotary evaporation then dried completely under a stream of N_2 (g).

Strong cation exchange chromatography of cell wall hydrolysates

[0131] The cell wall hydrolysate was redissolved in 100 ml of ddH_20 then chromatographed on Whatman P11 strong cation exchange column eluted by gravity as described earlier (Brady, Sadler, and Fry 323-27). Fractions 0-9 were collected (100 ml each) and concentrated to dryness by rotary evaporation. Fractions 0-9 were each redissolved in 1 ml of ddH_2O. A 1 μl aliquot of fractions 0-9 was spotted on Whatman 3MM chromatography paper and examined under UV light (254 nm). The spots were then exposed to NH_4OH vapor and reexamined under 254 nm light.

Paper chromatography

[0132] Paper chromatography was performed as previously described (Brady, Sadler, and Fry 323-27). An authentic dilDT standard was a gift from Dr. Stephen C. Fry of the University of Edinburgh and was chromatographed to detect dilDT in fraction 7.

Size exclusion chromatography

[0133] Fraction 7 (500 μl) from P11 cation exchange chromatography was filtered with Millipore spin filters (0.45 μm pore size; filter type HV; Nihon Millipore Kogyo K.K. Japan). The filtrate (250 μl per injection) was fractionated on a PolyHYDROXYETHYL A column (9.4 mm i.d. x 200 mm; 5 μm particle size, *20 nm pore size*, PolyLC Inc.) eluted isocratically in size exclusion mode (SEC) with 50mM formic acid at 0.8 ml/min and UV detection at 280 nm (Schnabelrauch et al. 477-89). Six fractions, designated SEC1 F1-6, were collected, checked for blue fluorescence as described in section 2.7.1.3, and then lyophilized.

[0134] SEC1 F4 was redissolved in 0.1 ml of ddH_2O and refractionated by SEC on a PolyHYDROXYETHYL A column (9.4 mm i.d. x 200 mm; 5 μm particle size, *10 nm pore size*, PolyLC Inc.; 10 μl per injection) as described above (Schnabelrauch et al. 477-89). Five fractions, designated SEC2 F1-5, were collected, checked for blue fluorescence, and then lyophilized.

Reverse-phase HPLC purification of diisodityrosine

[0135] Diisodityrosine (my diIDT) was purified from SEC2 F4 by reversed phase chromatography on a Waters Spherisorb ODS-2 C18 column (4.6 mm i.d. x 250 mm; 5μm particle size; Alltech Associates, Inc., Deerfield, IL). Briefly, SEC2 F4 was redissolved in 50 μl of ddH_2O and injected onto the C18 column equilibrated in ddH_2O. The column was washed with water for 15 min then eluted at 0.8 ml/min with a linear gradient to 60% (v/v) acetonitrile (aq) over 25 min. A peak absorbing at 280 nm eluted at 24.8 min was collected. Further fractionation of this peak on the 10nm pore size PolyhydroxyethylA column confirmed its retention time at 10.4 min.

Isolation of YK20 crosslinking amino acids

***In vitro* crosslinking of YK20 for isolation of crosslinking amino acids**

[0136] YK20 (20 mg) was crosslinked overnight under the conditions described in section 2.6.4 with pI 4.6 extensin peroxidase isolated from tomato suspension cultures. The reaction was stopped with 50mM 2-mercaptoethanol then lyophilized. The dried crosslinking reaction was dissolved in a minimal volume of water then centrifuged at 10,300 x *g* for 20 min. The insoluble pellet (YK20XL) was washed with ddH_20 until the conductivity of the supernatants was equal to that of water.

Acid Hydrolysis of YK20XL and YK20 for the isolation of crosslinking amino acids

[0137] The YK20XL and YK20 samples (2.2 mg and 1 mg respectively) were dissolved to 2 mg/ml in 6N HCl with 10mM phenol in sealed glass conical vials and then hydrolyzed for 20 h. at 110°C in a heating block. Hydrolysates were dried under N_2 (g) then resuspended in 100 µl ddH_2O.

Size exclusion chromatography for the purification of the crosslinking amino acids

[0138] Hydrolysates of YK20XL and YK20 were fractionated by SEC on a PolyHYDROXYETHYL A column (9.4 mm id X 200 mm; 5 µm particle size; 10 nm pore size; PolyLC) as previously described. UV detection was monitored at 280 nm for tyrosine (and tyrosine derivative) absorbance. The peaks were collected and lyophilized.

Molecular mass determination of the YK20 crosslinking amino acid by MALDI TOF mass spectrometry

[0139] The unknown crosslinking amino acid candidate was identified by MALDI-TOF mass spectrometry by Dr. Michael Hare at Oregon State University.

Identification of the crosslinking amino acid unknown by 1-D ^{1}HNMR

[0140] The unknown crosslinking amino acid from hydrolysate of YK20XL was characterized by 1-D ^{1}HNMR techniques by Dr. Li Tan of Ohio University.

Results

Synthetic gene construction

[0141] The synthetic genes *SStob-YK20-EGFP, SStob-YK8-EGFP, SStob-YL20-EGFP, SStob-YL8-EGFP, SStob-FK9-EGFP, and SStob-FL8-EGFP* were assembled from oligonucleotide primers and sequenced (Lewis et al. 400-06;Shpak 179;Shpak, Leykam, and Kieliszewski 14736-41;Tan L 1362-69) (Fig. 13). The synthetic gene *SStob-EGFP* was kindly donated by Dr. Elena Shpak. These synthetic genes were inserted into the binary plant transformation vector *pBI121* under the promotion of the *CaMV 35 S* constitutive promoter forming plasmids *pBI121-SStob-YK20-EGFP*, *pBI121-SStob-YK8-EGFP*, *pBI121-SStob-YL20-EGFP*, *pBI121-SStob-YL8-EGFP*, *pBI121-SStob-FK9-EGFP*, *pBI121-SStob-FL8-EGFP*, and *pBI121-SStob-EGFP* (Shpak, Leykam, and Kieliszewski 14736-41).

Transformation of *A. tumefaciens* and tobacco cells with synthetic genes

[0142] Agrobacterium tumefaciens was transformed with plasmids pBI121-SStob-YK8-EGFP, pBI121-SStob-YK20-EGFP, pBI121-SStob-YL8-EGFP, pBI121-SStob-YL20-EGFP, pBI121-SStob-FL8-EGFP, pBI121-SStob-FK9-EGFP, and pBI-SStob-EGFP (McCormick et al. 81-84). Tobacco BY-2 cell suspension cultures were subsequently transformed with these synthetic genes by A. tumefaciens infection

yielding cell lines designated NtYK20-EGFP, NtYK8-EGFP, NtYL20-EGFP, NtYL8-EGFP, NtFK9-EGFP, NtFL8-EGFP, and NtSS-EGFP.

Verification of synthetic gene insertion by PCR

[0143] Insertion of the synthetic genes *SStob-YK20-EGFP* and *SStob-YL8-EGFP* into tobacco cells was verified by PCR of genomic DNA with oligonucleotide primers complimentary to *SStob* and *EGFP* (Fig. 14).

Synthetic gene expression of YK20-EGFP, YK8-EGFP, YL8-EGFP, and FK9-EGFP

[0144] The transformed cell lines NtYK20-EGFP, NtYK8-EGFP, NtYL8-EGFP, NtFK9-EGFP, NtSS-EGFP demonstrated EGFP expression when viewed by fluorescence microscopy (Fig. 15). Untransformed tobacco (uTob) cells were viewed as a control. Although NtSS-EGFP and NtYL20-EGFPdemonstrated EGFP expression when viewed by fluorescence microscopy (not shown), no fusion glycoprotein was able to be isolated from the culture medium of these cell lines. The FL8-EGFP cell line was never observed to exhibit green fluorescence (not shown).

Cellular localization of YK20-EGFP to the cell wall

[0145] The YK20-EGFP, SS-EGFP, and uTob cells were plasmolyzed to determine the cellular localization of their respective transgene products. The fluorescence in YK20-EGFP cells was localized to the cell wall, with a concentration in the crosswall, whereas SS-EGFP was localized to intracellular compartments (Fig. 16). Plasmolysis of YK8-EGFP, YL8-EGFP, or FK9-EGFP cells did not demonstrate visible fluorescence in the cell wall (not shown). The YK20-EGFP fusion proteins remained with the cell wall after plasmolysis with 500 mM potassium phosphate pH 7 (Fig. 16E).

Isolation of the fusion glycoproteins YK20-EGFP, YK8-EGFP, YL8-EGFP and FK9-EGFP

[0146] Transformed YK20-EGFP, YK8-EGFP, YL8-EGFP, and FK9-EGFP cells demonstrated EGFP secretion into the culture medium. Concentrated culture medium was fractionated by HIC (Fig. 17). Fluorescent green fractions were collected, desalted by dialysis, and then lyophilized.

[0147] To further purify the HIC-fractionated YK20-EGFP, YK8-EGFP, YL8-EGFP, and FK9-EGFP fusion glycoproteins, gel filtration chromatography was performed using either a Superose-6 or 12 column. Chromatographs typically showed two fluorescent green peaks absorbing at 220 nm. The larger molecular weight peak corresponded to the P3-EGFP fusion glycoprotein and the smaller for cleaved EGFP (not shown).

[0148] The YK20-EGFP, YK8-EGFP, YL8-EGFP, and FK9-EGFP fusion glycoproteins eluted between 55-65 % end buffer as single peaks upon C4 reversed phase HPLC fractionation (Fig. 18). Typically, the fusion glycoprotein yields ranged from 3-27 mg/l medium for YK20-EGFP, from 4-7 mg/l for YK8-EGFP, and from 6-23 mg/l for YL8-EGFP. Yields for FK9-EGFP were lower at 0.1 to 3.3 mg/l.

Isolation of the glycomodules YK20, YK8, YL8, and FK9

[0149] Fractionation of tryptic digestions of YK20-EGFP, YK8-EGFP, YL8-EGFP and FK9-EGFP by gel filtration followed by reversed phase HPLC yielded pure glycomodules designated YK20, YK8, YL8, and FK9 (Fig. 19). Trypsin was not able to cleave lysine residues within the P3 modules, presumably due to the protection of the peptide backbone by the Hyp-O-oligoarabinosides (Lamport 1155-63;Smith et al. 1021-30;Stafstrom and Staehelin 242-46).

Deglycosylation of the fusion glycoproteins YK20-EGFP, YK8-EGFP, and YL8-EGFP by anhydrous hydrogen fluoride solvolysis

[0150] Deglycosylated fusion proteins were purified by C4 reversed phase HPLC (Fig. 20). Protein weight recoveries were recorded to estimate the mass of carbohydrate removed (Table 3). FK9-EGFP was not deglycosylated by anhydrous HF solvolysis.

Table 3. Estimated percent weight carbohydrate of YK20-EGFP, YK8-EGFP, and YL8-EGFP. Protein weight recoveries were recorded after HF deglycosylation, from which the percent carbohydrate was estimated.

Sample	Protein wt. recovery (%)	Carbohydrate wt. recovery (%)
YK20-EGFP	42	68
YK8-EGFP	49	51
YL8-EGFP	48	52

Amino acid composition analyses of the YK20, YK8, YL8, and FK9 glycomodules

[0151] Amino acid compositions of the purified glycomodules YK20, YK8, YL8 and FK9 were determined (Table 4). All of the samples exhibited nearly complete hydroxylation of proline residues. Isodityrosine (IDT), a diphenyl ether-linked tyrosine dimer (Fig. 1) was observed in the hydrolysates of YK20, YK8, and YL8, but not in FK9 (Fig. 21).

Table 4. Amino acid compositions of the glycomodules YK20, YK8, YL8, and FK9. Actual compositions (Protein) are compared to the deduced compositions (Gene) from synthetic gene translations.

	Amino Acid Composition (mol%)							
	YK20		YK8		YL8		FK9	
Amino Acid	Protein	Gene	Protein	Gene	Protein	Gene	Protein	Gene
Hyp	59.8	0	60	0	56	0	52.5	0
Pro	0	55.4	0	54.1	0	54.1	0	55.3
Ser	18.2	19	18	19.3	20	19.3	17.8	19.5
Tyr	9	18.3	9	17.8	13	17.8	0	0
1/2IDT	8.7	0	8	0	5	0	0	0
Phe	0	0	0	0	0	0	21	17.5
Lys	4.3	6.4	5	6.7	0	0.7	7	6.5
Leu	0	0	0	0	6	6	0.6	0
Thr	0	0.3	0	0.7	0	0.7	0	0
Met	0	0.3	0	0.7	0	0.7	0.9	0.6
Val	0	0.3	0	0.7	0	0.7	0.2	0.6
Total	100	100	100	100	100	100	100	100

Partial N-terminal sequencing of dYK20-EGFP, dYK8-EGFP, dYL8-EGFP and FK9-EGFP by automated Edman degradation

[0152] The partial N-terminal sequences of dYK20-EGFP, dYK8-EGFP, dYL8-EGFP and FK9-EGFP matched those predicted from the gene sequences (Table 5). The two terminal Arg residues of dYK20-EGFP, dYK8-EGFP, and dYL8-EGFP fusion proteins were a result of restriction site design. Originally, YK20-EGFP and YL8-EGFP were sent for sequencing, but the sequences did not exactly match those predicted from their genes. Deglycosylation of these fusion glycoproteins prior to N-terminal sequencing, however, yielded their predicted sequences.

Table 5. Partial N-terminal sequences of dYK20-EGFP, dYK8-EGFP, dYL8-EGFP, and FK9-EGFP.

Fusion protein	Partial N-terminal sequence
dYK20-EGFP	RRPSOOOOYYYKSOOOOSOS
dYK8-EGFP	RROSOOOOYYYKSOOOOSOS
dYL8-EGFP	RRPSOOOOYYYLSOOOOSOS
FK9-EGFP	SOOOOSOSOOOOFFFKSOOOOSOSOOOO

Hyp-O-glycoside profiles of the YK20-EGFP, YK8-EGFP, YL8-EGFP and FK9-EGFP fusion glycoproteins

[0153] All of the fusion glycoproteins showed greater than 82% of the total Hyp residues possessed O-linked tri or tetraoligoarabinosides (Table 6) as predicted by the Hyp Contiguity Hypothesis (Kieliszewski and Lamport 157-72). With the exception of YL8-EGFP, Hyp-PS was not detected in the fusion glycoproteins.

Table 6. Hyp-O-glycoside profiles of the YK20-EFGP, YK8-EFGP, YL8-EGFP and FK9-EGFP fusion glycoproteins. EGFP does not contain Hyp and is not glycosylated (Shpak E. et al. 1999).

	Percentage of total hydroxyproline			
Hyp-Glycoside[a]	YK20-EGFP	YK8-EGFP	YL8-EGFP	FK9-EGFP
Hyp-PS	0	0	3	0
Hyp-Ara_4	56	56	55	42
Hyp-Ara_3	32	31	27	40
Hyp-Ara_2	4	4	6	5
Hyp-Ara_1	5	5	4	5
NG-Hyp	3	4	5	8

[a] Hyp-PS, Hyp polysaccharide; Hyp-Ara_n, Hyp-arabinoside$_{1-4}$; NG-Hyp, non-glycosylated Hyp.

Neutral sugar composition analyses of the fusion glycoproteins YK20-EGFP, YK8-EGFP, YL8-EGFP, and FK9-EGFP

[0154] At least 84% of the neutral sugar carbohydrate attached to our fusion glycoproteins was arabinose (Table 7). Significant amounts of galactose were observed in all of the fusion glycoproteins indicating the presence of monogalactosyl-serine (Ser-O-gal). A lower ratio of arabinose:galactose was observed for YL8-EGFP. Essentially no uronic acids were detected in any of the fusion glycoproteins or glycomodules. Thus, we estimated that YK20-EGFP was 39% wt carbohydrate (CHO), YK8-EGFP was 48% wt. CHO, YL8-EGFP was 45% wt. CHO, and FK9-EGFP was 32% wt. CHO.

Table 7. Neutral sugar compositions of the YK20-EGFP, YK8-EGFP, YL8-EGFP, and FK9-EGFP fusion glycoproteins. EGFP is not glycosylated (Shpak, E. et al 1999).

	Glycosyl Composition (mol%)			
Glycosyl residue	YK20-EGFP	YK8-EGFP	YL8-EGFP	FK9-EGFP
Arabinose	90	91	84	91
Galactose	8	7	12	9
Rhamnose	1	1	2	0
Xylose	0	0	1	0
Glucose	1	1	1	0

Detection of Hyp-PS in the YL8-EGFP fusion glycoprotein

[0155] Base hydrolysis followed by gel filtration chromatography indicated a high molecular wt. (HMW) Hyp-O-glycoside occurred in YL8-EGFP and comprised approximately 5.2% of the total Hyp as compared to 3% of the total Hyp calculated from the Hyp-O-glycoside profile (Fig. 22). This confirmed that YL8-EGFP contains a small amount of Hyp-PS.

Estimated molecular mass of the fusion glycoproteins YK20-EGFP, YK8-EGFP, YL8-EGFP, and FK9-EGFP

[0156] Molecular masses of YK20-EGFP, YK8-EGFP, YL8-EGFP, and FK9-EGFP were estimated from the Hyp-O-glycoside data, the neutral sugar composition data, and the deduced amino acid composition data. EGFP contributes 26,924 Da to each fusion glycoprotein. The weight contributions of the arabinooligosaccharides were estimated from the Hyp-O-glycoside profiles (Table 8). Galactose was estimated from the neutral sugar compositions and the weight contribution of the arabinooligosaccharides (Table 9). The protein weight contributions of the modules were predicted from the deduced amino acid sequence (Table 10; as shown in Fig. 13). The sum of these values yielded predicted molecular weights of each fusion glycoprotein (Table 11).

Table 8. Weight contribution of the arabinooligosaccharides of YK20-EGFP. Weight contributions of the O-linked arabinooligosaccharides were estimated from the Hyp-O-glycoside profile data. Reported here is YK20-EGFP only. The YK8-EGFP, YL8-EGFP, and FK9-EGFP fusion glycoproteins were estimated the same way. MW, molecular weight (Da/mol)

YK20-EGFP	% of Hyp	Number/molecule	MW of Arabinosides	Daltons
Hyp-$(Ara)_4$	56	101	528	53222
Hyp-$(Ara)_3$	32	58	396	22810
Hyp-$(Ara)_2$	4	7	264	1901
Hyp-$(Ara)_1$	5	9	132	1188
NG-Hyp	3	5	0	0
Totals	100	180		79121

Table 9. Weight contribution of the monogalactosides of YK20-EGFP. From table 6 we know that the weight contribution of the oligoarabinosides is 79121. From table 5 we know that the fusion glycoprotein is 90 mol% arabinose thus (599 mol equiv.). Galactose was 8 mol% galactose (53 mol equiv.) The weight contribution of galactose was calculated from these values. The weight contributions of galactosides in YK8-EGFP, YL8-EGFP, and FK9-EGFP fusion glycoproteins were estimated the same way. MW, molecular weight (Da/mol); NG-Ser, non-glycosylated Ser.

YK20-EGFP	Number/Molecule	MW of Gal	Daltons of Gal
Ser-O-Gal	53	162.2	8642
NG-Ser	8	0	0
Total Ser	61		8642

Table 10. Weight contribution of the protein of YK20. The deduced translation of the YK20 gene sequence in Figure 1A was used to estimate the molecular wt. contribution of the polypeptide. The weight contributions of the polypeptide in YK8-EGFP, YL8-EGFP, and FK9-EGFP were estimated the same way.

Amino acid	Number of residues	Residue wt.	Daltons	Module mol%
Ser	61	87.08	5312	18.8
Hyp	180	113.10	20358	55.4
Tyr	60	147.18	8831	18.5
Lys	20	128.17	2563	6.2
Pro	1	97.12	97	0.3
Thr	1	101.11	101	0.3
Arg	2	156.19	312	0.6
Total	325		37575	100.0

Table 11. Estimated molecular weights (Da) of YK20-EGFP, YK8-EGFP, YL8-EGFP, and FK9-EGFP. Summation of the wt contributions of the arabinosides, the galactosides and the polypeptide components yielded the estimated molecular weights of each fusion glycoprotein. Values are reported in Daltons.

Weight Contributors	YK20-EGFP	YK8-EGFP	YL8-EGFP	FK9-EGFP
Arabinosides	79121	31152	30096	32208
Galactosides	8642	2945	5283	3914
EGFP	26924	26924	26924	26924
Module	37575	15387	15267	17237
Total (Daltons)	152262	76408	77570	80283

[0157] These calculations indicate that YK20-EGFP is 152 kDa of which 58% is carbohydrate (CHO), YK8-EGFP is 76.4 kDa (45% wt. CHO), YL8-EGFP is 77.6 kDa (45% wt. CHO), and FK9-EGFP is 80 kDa (45% wt. CHO). These values compare well with the weight recoveries after HF deglycosylation (Table 3) and the neutral sugar compositions (Table 7), with the exception of YK20-EGFP and FK9-EGFP. The weight percent carbohydrate of these fusion glycoproteins appear to be underestimated from the neutral sugar compositions. Only YL8-EGFP has a Gal:Ser ratio >1, as some of the Gal residues are in a Hyp-PS linkage.

Isolation of pI 4.6 extensin peroxidase from tomato

[0158] A single extensin peroxidase peak was retained and then eluted from the DEAE-5PW anion exchange HPLC column (Fig. 23). Thus, preparative DEAE purified enzyme was used for all further crosslinking reactions.

***In vitro* crosslinking of extensin precursors with pI 4.6 extensin peroxidase**

[0159] Extensin peroxidase isolated from tomato cell suspension cultures was used to determine the substrate requirements for intermolecular covalent crosslinking *in vitro*. Native tomato P1 extensin was used as a positive crosslinking control substrate (Fig. 24) (Schnabelrauch et al. 477-89). Native tomato P1, YK20, YK8, and YL8 were able to be covalently crosslinked *in vitro* by the tomato pI 4.6 extensin peroxidase (Fig. 25). The FK9 glycomodule was not a substrate for the pI 4.6 extensin peroxidase nor was BSA (Figs. 26 and 27). It was also shown that both the enzyme and peroxide were required for *in vitro* crosslinking of YK20 (Fig. 28 and 29). Peaks eluting after 22 min are a result of the crosslinking buffers and stopping reagent (Fig. 30).

In vitro crosslinking rate determination for the P1, YK20, YK8, YL8 and FK9 substrates

[0160] _In vitro_ crosslinking reaction rates were determined from freshly prepared substrate stocks of P1, YK20, YK8, and YL8 using a first order rate equation. Crosslinking reactions were performed in triplicate and the mean crosslinking rates (μg crosslinked per sec) +/- the estimated standard deviations were 6.2 +/- 0.68 for P1, 4.9 +/- 0.31 for YK20, 3.4 +/- 0.13 for YK8, 1.4 +/- 0.07 for YL8 (Fig. 31). Since FK9 does not crosslink, the reaction rate is reported as zero.

In vitro crosslinking of extensin precursors with tobacco salt-eluates

[0161] To check for the presence of an extensin peroxidase in tobacco BY-2 cell suspension cultures, _in vitro_ crosslinking assays using crude tobacco culture media (E1) and a salt-eluate of intact tobacco cells (E2) on native tomato P1 extensin and YK20 substrates. The amount of enzyme activity was quantified by ABTS assay for E1 and E2 and equal amounts of both enzymes (~35ng, 7 times the normal amount) were used per standard amount of substrate (60 μg). I found that the E2 was able to catalyze the complete _in vitro_ crosslinking of P1 and YK20 over a reaction time of 18 h, whereas E1 was not.

In vitro crosslinking of YK20 with crude cationic tomato peroxidases

[0162] The preparative DEAE void contained a significant amount of peroxidase activity. This fraction was tested for its ability to crosslink P1, YK20, and YL8 _in vitro_. The crosslinking reactions were the same as the standard reactions except, the enzyme concentration and the incubation times were doubled. A minor amount of crosslinking was observed for all three substrates (Fig. 32).

Amino acid composition of the crosslinked YK20

[0163] Amino acid compositions of the crosslinked YK20 (YK20XL) were compared with the YK20 glycomodule (Table 12). The amino acid compositions showed a decrease in IDT and total Tyr equivalents after crosslinking.

Table 12. Amino acid compositions of YK20 and YK20XL. Losses of IDT and net Tyr equivalents were observed after *in vitro* crosslinking.

	Amino Acid Composition (mol%)		
	YK20		YK20XL
Amino Acid	Protein	Gene	Protein
Hyp	59.8	0	56.8
Pro	0	55.4	2.3
Ser	18.2	19	18.2
Tyr	9	18.3	11.3
1/2IDT	8.7	0	2.6
Phe	0	0	0
Lys	4.3	6.4	6.4
Leu	0	0	0
Thr	0	0.3	2.4
Met	0	0.3	0
Val	0	0.3	0
Total	100	100	100

Neutral sugar composition of the crosslinked YK20

[0164] There were no significant differences in the neutral sugar composition between the soluble YK20XL and the YK20 glycomodule (Table 13). Note Glc, Xyl, and Rha are most likely contaminants. We did not detect uronic acids in either YK20 or YK20XL.

Table 13. Neutral sugar compositions of YK20 and YK20XL. A slight decrease in the mol % of arabinose and increase in galactose were observed upon crosslinking.

	Mol Percent	
Glycosyl Residue	YK20	YK20XL
Ara	91	88
Gal	7	9
Rha	1	0
Xyl	0	1
Glc	1	2

Isolation of a putative crosslinking amino acid from crosslinked YK20

[0165] A putative crosslinking amino acid, designated "unknown", was detected in the acid hydrolysates of YK20XL, but not in the YK20 glycomodule (Fig. 33). The unknown was collected and then fractionated again by size exclusion chromatography.

Identification of crosslinking amino acid (unknown) by paper chromatography

[0166] Paper chromatography of the "unknown" compound from figure 33 and an authentic diisodityrosine standard (Fry's dilDT) demonstrated that they had the same Rf value (Table 14). A putative dilDT standard (my dilDT) freshly prepared from tomato cell wall hydrolysates, also co-chromatographed with the unknown and the authentic standard.

Table 14. Paper chromatography of Fry's dilDT, "unknown", and my dilDT in BAW.

Sample	Rf value	Fluorescence*
Fry's dilDT	0.06	+
Unknown	0.06	+
My dilDT	0.06	+

*Vivid blue fluorescence was observed when viewed under short-wave (254 nm) UV light after exposure with NH_3 vapor.

Identification of crosslinking amino acid (unknown) by size exclusion chromatography

[0167] The Fry's dilDT standard and the unknown were fractionated by size exclusion chromatography. The authentic dilDT standard co-eluted with the unknown (not shown).

Identification of crosslinking amino acid (unknown) MALDI-TOF mass spectrometry

[0168] The unknown was sent to Dr. Michael Hare of Oregon State University for molecular weight determination by MALDI-TOF mass spectrometry.

The theoretical molecular mass of diIDT is [M+H] = 719. The mass spectrum of the unknown peak contained a molecular ion of [M+H] = 719. The same molecular ion was observed in Fry's diIDT both here and as reported earlier (Brady, Sadler, and Fry 323-27) and my diIDT standard. Thus, the putative crosslinking amino acid formed *in vitro* was diIDT.

Identification of crosslinking amino acid (unknown) 1-D ^{1}HNMR

[0169] The 1-D ^{1}HNMR spectrum of the "unknown" sample showed 4 groups of resonances from the aromatic region. The resonances at 7.203 (dd, J = 8.4 Hz, 2.2 Hz) and 6.965 (dd, 8.4 Hz, 2.2 Hz) ppm (peaks B and D in Figure 34) suggested this sample contained 1,4-substituted benzene ring(s). Two mutually coupled single-proton doublet resonances (J = 2.2 Hz) at 6.900 and 6.970 ppm (peaks E and C in Figure 34) suggested the existence of 1,2,3,5-substituted benzene ring(s). The N-linked protons showed at 8.330 ppm (peak A in Figure 34). Furthermore, the ratio of integrated area of signals A:B:C:D:E was 12:4:2:4:2. These results strongly suggested that the "unknown" sample corresponded to diDT, in which all the N atoms were fully protonated.

[0170] By using the ChemDraw program, the predicted data of diDT showed similar spectrum pattern and chemical shifts as those of the "unknown" also indicated that it was diIDT (Fig. 35).

[0171] **APPLICATION OF THE TECHNOLOGY**

[0172] **Example 2: Fiber Formation**

[0173] Crosslinking extensin modules (SOOOOTOVYK; see Tables 16 and 17) (Note: here onward the single letter codes are used to denote amino acids, with O representing Hyp) are placed at the N- and C-termini, and a central stretch composed of rigid SOOOO repeats (bolded below) flanked on either side with elastin

sequences (italicized below; See Table 18). This construction should produce tough, rigid, protease-resistant, elastic fibres of high tensile strength, analogous to fibres found in mussel byssus threads but with enhanced water-holding ability in the glycosylated regions.

[0174] Each of the repetitive modules in Examples 2A and 2B, including the elastin module, has already been successfully expressed in plants. Peroxidase-catalyzed intermolecular crosslinking of the Thr-Hyp-Val-Tyr-Lys module at the ends of the molecules should produce long fibres. (Note: the Lys residues in P1 extensin sequences are resistant to trypsin).

[0175] Example 2A:

SOOOOTOVYKSOOOOTOVYKSOOOO[*VGVPG VGVPG]*$_5$ [**SOOOOSOOOOSOOOO]**$_6$ [*VGVPG VGVPG]*$_5$- SOOOOTOVYKSOOOOTOVYKSOOOO

[0176] The variant below replaces the central rigid, arabinosylated SOOOO repeats with the more flexible SOSOSO repeats (Bolded below), or AOAOAO repeats, that serve as arabinogalactan polysaccharide addition sites. These polysaccharide additions sites greatly increase the amount of carbohydrate on the molecule, impart a negative charge due to the abundance of uronic acids in the polysaccharide, and promote an extended conformation.

[0177] Example 2B:

SOOOOTOVYKSOOOOTOVYKSOOOO[*VGVPG VGVPG]*$_5$
[**SOSOSOSOSOSOSOSOSOSOSOSO]**$_6$ [*VGVPG VGVPG]*$_5$
SOOOOTOVYKSOOOOTOVYKSOOOO

[0178] **Example 3: HRGP-based Emulsifiers**

[0179] Gum arabic glycoprotein is an excellent emulsifier (J-F Xu and M. Kieliszewski, unpublished data). Features of GAGP as well as other HRGPs can be exploited to create novel emulsifiers. Example 3 is very similar to Example 2, except it lacks crosslinking motifs and the elastin repeats of Example 2 have been replaced

with a smaller (12 residues, italicized below), extended (Pro-rich) but flexible, hydrophobic module. Charge repulsions and steric interference produced by the rhamnoglucuronoarabinogalactan polysaccharide sidechains attached to the AOAO repeats should help stabilize an oil in water emulsion by preventing aggregation/flocculation of the emulsion particles. Further stabilization can be achieved by adding crosslinking modules. Variations include glycoproteins having only a single hydrophobic module and a single glycosylated HRGP module or many of each type dispersed regularly throughout the chain.

[0180] Example 3:

VPGVPGVPGVPG [**AOAOAOAOAOAOAOAOAOAO**]$_3$ *VPGVPGVPGVPG*

[0181] **Example 4: HRGP-based Emulsifiers with Crosslinking Motifs**

[0182] This Example exploits variations of the GAGP repeat and the extensin TOVYK intermolecular crosslinking motif (underlined) to create a crosslinked emulsifier. The C-terminus variation of the GAGP repeat (italicized and bolded) has decreased glycosylation, increased flexibility, and increased hydrophobicity, which will allow the molecule to interact with the surface of an oil droplet. Such a molecule might facilitate slow-release drug delivery from an oil-in-water emulsion.

SOOOTLSOSOTOTOOLGPHSOOOTLSOSOTOTOOLGPHTOVYKSOOOTLSOSOTOTOOLGPHSO
OOTLSOSOTOTOOLGPHTOVYKSOOOTLSOSOTOTOOLGPHSOOOTLSOSOTOTOOLGPHTOVY
KSOOOTLSOSOTOTOOLGPH***SOLPTLSOLPTOTOOLLPH SOLPTLSOLPTOTOOLLPH***

[0183] Methods to test the properties of the inventive materials include:

[0184] 1. Crosslinking – cross-linking of monomers to form multimers can be tested several ways:

[0185] a. Size exclusion chromatography is an effective assay as already demonstrated by Schnabelrauch et al. (Schnabelrauch,L.S., Kieliszewski,M.J., Upham,B.L., Alizedeh,H., and Lamport,D.T.A. (1996) *Plant J.*, **9**, 477-489.)

[0186] b. Assay of free Lys in crosslinks putatively involving Lys residues by the use of the highly reactive acrylonitrile for the cyanoethylation of all the freely available $-NH_2$ groups.

[0187] c. Dityrosine or isodityrosine formation for crosslinks involving Tyr residues (Epstein,L. and Lamport,D.T.A. (1984) *Phytochem.*, **23,** 1241-1246).

[0188] d. Surface hydrophobicity measurements using cis-parinaric acid as a fluorescent probe (Liu, M. and Damodaran, S. (1999) *J.Agric.Food Chem.*, **47,** 1514-1519; Kato, A. and Nakai, S. (1980) *Biochim.Biophys.Acta*, **624,** 13-20).

[0189] 2. Mechanical properties of biopolymer films

[0190] a. Compression isotherms via Langmuir film balance (Fauconnier,M.L., Blecker,C., Groyne,J., Razafindralambo,H., Vanzeveren,E., Marlier,M., and Paquot,M. (2000) *J.Agric.Food Chem.*, **48,** 2709-2712).

[0191] 3. Emulsifying properties will be evaluated as follows:

a. At different pH
b. Different ionic strengths (Popineau, Y., Pineau, F., Evon, P., and Berot, S. (1999) *Nahrung*, **43,** 361-367)
c. Stability of the emulsion (Liu, M. and Damodaran, S. (1999) *J.Agric.Food Chem.*, **47,** 1514-1519)
d. Droplet size and surface charge: Emulsions containing the smallest globules tend to be the most stable (200-500 nm), as do emulsions that have a high surface charge. Droplet size will be measured in a Coulter counter as described by Dickenson et al. (Dickenson,E., Rolfe,S.E., and Dalgleish,D.G. (1988) *Food Hydrocolloids*, **2,** 397-405). Zeta potential measurements are typically carried out using a Doppler electrophoresis apparatus or via moving-boundary electrophoresis (Washington, C. (1990) *Int.J.Pharm.*, **66,** 1-21).

[0192] 4. Fiber formation and Fiber properties-

[0193] Extrusion of crosslinked products through an appropriately sized nozzle will yield fibers. After a curing stage, the physical properties of the fibers will be determined in an extensometer to measure their elastic modulus and tensile strength (Cosgrove,D.J. (1993) *New Phytol.*, **124,** 1-23).

[0194] 5. Elastic properties/viscoelasticity can be measured using a rheometer as described by Larre et al (Larre, C., Denery-Papini, S., Popineau, Y., Deshayes, G., Desserme, C., and Lefebvre, J. (2000) *Cereal Chem.(2000)*, **77**, 121-127).

[0195] 6. Lipid encapsulation - lipid exposed at the surface of a microcapsule can be measured via extraction with chloroform as in: Minemoto, Y., Adachi, S., and Matsuno, R. (1999) *Food Sci.Technol.Res.*, **5**, 289-293.

[0196] Table 16. Repetitive Sequences Common in Hydroxyproline-rich Glycoproteins

Characteristic Sequence	HRGP
Ser-Hyp-Hyp-Hyp-Hyp-Thr-Hyp-Val-Tyr-Lys	P1-type extensin
Ser-Hyp-Hyp-Hyp-Hyp-Ser-Hyp-Ser-Hyp-Hyp-Hyp-Hyp-Hyp-Tyr-Tyr-Tyr-Lys	P3 extensin
Pro-Hyp-Val-Tyr-Lys	Proline-rich protein
Ala-Hyp-Ala-Hyp	Arabinogalactan-protein

[0197] Table 17. HRGP repetitive glycomodules, peptide modules and their corresponding properties

Repetitive module	Properties
X-Hyp-Hyp-Hyp-Hyp	Glycomodule common in extensins; this module has a polyproline-II extended conformation; Extensive arabinosylation of Hyp enhances the polyproline-II conformation and rigidifies the module (X = Ser, Ala, Thr) .
X-Hyp-Hyp	Glycomodule common in arabinogalactan- proteins; Less polyproline II conformation than the X-Hyp-Hyp-Hyp-Hyp glycomodule; First Hyp is always arabinosylated, the second is occasionally. (X = Ser, Ala, Thr)
X-Hyp-X-Hyp	Glycomodule which defines the arabinogalactan-proteins (i.e. clustered non-contiguous Hyp); An extended random coil conformation further enhanced by arabinogalactan-polysaccharide addition to each Hyp (X= Ser, Ala).
X-Hyp-Val-Tyr-Lys	Peptide module of extensins and PRPs; adhesion; peroxidase-catalyzed intermolecular cross-linking; Possible reverse β-turns; tandem repeats of this module may increase elasticity. (X= Thr, Glu, Pro, His, Ile).
Tyr-Tyr-Tyr-Lys Tyr-Lys-Tyr-Lys	Peptide crosslinking modules common in extensin; Peroxidase-catalyzed (?) intramolecular isodityrosine cross-linking sequences; probable intermolecular cross-link that enhances hydrophobicity; intramolecular crosslink rigidifies this module.
Lys-Pro	Ionic/covalent cross-linking motif of extensins and PRPs

[0198] Table 18. Non-HRGP peptide modules and corresponding properties

Thr-Val-Gln-Gln-Glu-Leu Pro-Gly-Gln-Gln-Ile-Val	Sequences for transglutaminase crosslinking to Lys residues in HRGPs; formation of N^{ε}-(γ-glutamyl)lysine.
Leu-Cys-Cys-Ser X-Cys-Gly	Inter/intramolecular disulfide bond formation. (X = Gln, Lys, Arg).
Val-Pro-Gly-Val-Gly	Elastomeric module

[0199] **Example 5: Additional Cross-Linking Motifs**

[0200] Summary

[0201] P1 extensin is a hydroxyproline-rich glycoprotein (HRGP) that is an integral part of the primary plant cell wall. It consists mainly of two repetitive motifs:

SPPPPTP**VYK** and SPPPPVKPYHPTP**VYK**. Earlier work demonstrated the requirement of a Val-Tyr-Lys motif for *in vitro* intermolecular cross-linking of P1 by extensin peroxidase. Tyrosine and lysine contain reactive groups, which are proposed to play a role in the cross-linking of extensin. Here we have tested the role of Val-Tyr-Lys motif in cross-linking by designing synthetic genes and expressing them in tobacco cells. These genes encode the P1 motif: SOOOOTOVYK (O is Hyp) and variants that differ in their cross-linking motifs with substitutions of Tyr → Phe and Lys → Leu. Figure 36 shows the plasmid construct.

[0202] The genes were expressed in *Nicotiana tabacum* BY2 suspension cultured cells as enhanced green fluorescent protein (EGFP) fusion proteins and were targeted through the ER and Golgi to the cell wall via a tobacco extensin signal sequence. (Figure 37 shows photomicrographs of cells.) The transgene products were isolated from cell culture medium by a combination of hydrophobic interaction and reverse phase C4 column chromatography. The fusion proteins underwent hydroxylation of proline and extensive arabinosylation of Hyp residues similar to endogenous P1. The VYK construct $[SPPPPTPVYK]_6$ (one of the cell lines expressing this glycoprotein is designated VYK C1) was cross-linked *in vitro* by tomato extensin peroxidase, while the VFL construct $[SPPPPTPVFL]_6$ (one of the cell lines expressing this glycoprotein is designated VFL-F) showed no cross-linking activity. Another cell line (designated VYL-6.1) expressing 6 repeats of the DNA sequence [SPPPPTPVYL] has not been tested for crosslinking. (Figure 38 shows results of crosslinking studies.)

[0203] Expermental Methods

[0204] Isolation of the fusion proteins

[0205] Soluble P1 extensin fusion proteins containing six repeats were isolated from the culture medium. The media was concentrated by rotary evaporation, dialyzed against water, and then freeze-dried. The dry sample was redissolved in 2M sodium chloride and injected onto a hydrophobic interaction column. A step gradient of decreasing sodium chloride was used to elute the column. Fractions were monitored for fluorescence using a Hewlett-Packard 1100 Series flow-through fluorometer. The fluorescent fractions were then purified using gel filtration and reverse-phase chromatography.

[0206] *In Vitro* Cross-linking Assays of P1 Extensin Gene Products

[0207] The isolated fusion proteins were treated with trypsin to remove EGFP and were then tested as substrates for the pI 4.6 peroxidase isolated from tomato. The cross-linking of the monomers was measured by a Superose 6 gel filtration assay of cross-linked product. Native P1 extensin served as the positive control and the $(SPPPPTPVFL)_6$ served as the negative control.

[0208] Conclusion

[0209] Three fusion proteins were expressed in tobacco cell cultures. These proteins were isolated from the media using hydrophobic interaction, gel filtration, and reverse-phase chromatography. Neutral sugar analysis has shown that the main sugar is arabinose with trace amounts of galactose, which is consistent with native P1 extensin.

[0210] The module was cleaved from EGFP by tryptic digestion and the module was then tested for cross-linking activity. Native P1 tomato extensin was used as a control. It appears that the glycoprotein $(SOOOOVYK)_6$ does cross-link, although to a much lesser extent than P1. The glycoprotein $(SOOOOVFL)_6$ did not crosslink. The glycoprotein $(SOOOOVYL)_6$ has not been tested for crosslinking.

[0211] **Example 6: Cross-Linking HRGPs with Arabinogalactan Glycomodules**

[0212] Overview

[0213] The range of crosslinking HRGPs was expanded to include some containing arabinogalactan glycomodules. Thus, we combined the Ser-Hyp_4-Ser-Hyp-Ser-Hyp_4-Tyr-Tyr-Tyr-Lys extensin-type arabinosylated repeats (abbreviated YK below) that can crosslink with $(Ala\text{-}Hyp)_n$ arabinogalactan-type repeats (designated AlaPro below). We constructed genes encoding *(AlaPro)$_4$(YK)$_{20}$*, *(AlaPro)$_4$(YK)$_8$*, and *(AlaPro)$_4$(YK)$_4$* genes, expressed them as EGFP fusion proteins, and tested whether or not the resulting glycoproteins could be cross-linked by tomato extensin peroxidase.

[0214] Methods

[0215] 1). Design of oligonucleotides for constructsA.

[0216] Oligonucleotide set for constructing *(AlaPro)$_4$(YK)$_{20/8/4}$* genes—oligo#1

```
     R    A    P    A    P    A    P    A
5`-C CGG  GCT  CCA  GCA  CCT  GCC  CCA  GCC-3`
       3`- CGA  GGT  CGT  GGA  CGG  GGT  CGG  GGT A-5`
    XmaI
```

The 5' end of the sense oligonucleotide (encoding the sequence APAPAPA) above contained a XmaI cut site and the 5' end of the antisense oligonucleotide contained a BbsI cut site. This allowed insertion of the construct into preexisting plasmids pUC-$(YK)_{20}$, pUC-$(YK)_8$, and pUC-$(YK)_4$.

[0217] 2). Construction of pUC-$SS^{tob}(AlaPro)_4(YK)_{20/8/4}$EGFP, pUC-$SS^{tob}(YK)_{20/8/4}(AlaPro)_3$- EGFP and pBI-$SS^{tob}(AlaPro)_4(YK)_{20/8/4}$EGFP, pBI-$SS^{tob}(YK)_{20/8/4}(AlaPro)_4$EGFP plasmids.

[0218] The oligonucleotide sets were ligated into *pUC-$(YK)_{20}$*, *pUC-$(YK)_8$*, and *pUC-$(YK)_4$*, respectively (See Figure 39). The resulting *$(AlaPro)_4(YK)_{20/8/4}$* and *$(YK)_{20/8/4}(AlaPro)_3$* fragments were ligated into *pUC-$SS^{tob}(AP)_{51}EGFP$* and replace the $(AP)_{51}$ gene. The *$SS^{tob}(AlaPro)_4(YK)_{20/8/4}EGFP$* and *$SS^{tob}(YK)_{20/8/4}(AlaPro)_3$* fragments were finally ligated into the plant transformation vector pBI121 vector to give plasmids: *pBI-$SS^{tob}(AlaPro)_4(YK)_{20/8/4}EGFP$* and *pBI-$SS^{tob}(YK)_{20/8/4}(AlaPro)_4EGFP$*.

[0219] 3). Transformation of above gene cassettes into tobacoo BY2 suspension cultured cells.

[0220] Above-modified pBI plasmids were transformed into Agrobacterium; transformants were selected by kanamycin and streptomycin resistance. The agrobacteria were then used to infect suspension cultured tobacco BY2 cells. The transformed tobacco cells were selected for kanamycin resistance and cultured in SH media for 20 days at room temperature.

[0221] 4). Fusion protein isolation and purification.

[0222] Culture medium from the transformed tobacco cells was collected and concentrated under vacuum at 28°C. The concentrated media was adjusted with NaCl to 2M and loaded on a Butyl-sepharose HIC column. The column was eluted with a gradient starting with 2M NaCl and ending with water. Green fractions were collected and freeze-dried. The crude fusion proteins were then purified on a Superose-12 preparative gel filtration column. Green fractions were either repurified on a C-4 reverse phased semi-preparative column in a gradient from 0.1% TFA (aq) to 100% of 80% acetonitrile in 0.1% TFA (aq) in 100 minutes, or concentrated and desalted on centricon spin columns of cutting size of 5 kDa.

[0223] 5). Removal of EGFP from the isolated fusion proteins.

[0224] EGFP was removed from fusion proteins $(AlaPro)_4(YK)_{20/8/4}EGFP$ by tryptic digestion in a solution of 2% ammonium bicarbonate, 5 mM $CaCl_2$. The glycomodules were isolated from the solution by size exclusion chromatography on the Superose-12 column, and further purified on C-4 reverse phased semipreparative column using a 60 min gradient from starting with 100% 0.1% TFA (aq) and ending with 80% acetonitrile in 0.1% TFA (aq).

[0225] 6). Hyp-glycoside profiles of fusion proteins.

[0226] 7.2 mg of $(AlaPro)_4(YK)_{20}EGFP$ was hydrolyzed in 0.44 N NaOH at 105°C for 18 hr. The hydrolysate was neutralized and freeze-dried.

[0227] 7). Sugar analysis of fusion proteins.

[0228] One-hundred μg of fusion proteins were dissolved in 200 μl of 2N TFA and heated at 121°C for 2 hrs. The hydrolysate was then reduced and acetylated as described before.

[0229] 8). Cross-linking reaction of glycomodules and fusion proteins.

[0230] Cross-linking reactions were carried out in citrate buffer with H_2O_2 and pI 4.6 extensin peroxidase as described before. The reaction mixtures were then size-fractionated on an analytical Superose-6 column, which was monitored for UV absorbance at 220 nm.

[0231] Results

[0232] 1). Sequences of each construct

[0233] The pUC-derived plasmids were sequenced using a M13 forward primer (see construct sequences in A below).

A. $(AP)_4(YK)_{20}EGFP$

A P A P A P A P S P P P P Y Y
GCT CCA GCA CCT GCC CCA GCC [CCA TCA CCA CCA CCA CCT TAC TAC

Y K S P P P P S P

TAC AAG TCT CCT CCT CCC CCA TCA CCA]$_{20}$ EGFP

B. (AP)$_4$(YK)$_8$EGFP

A P A P A P A P S P P P P Y Y
GCT CCA GCA CCT GCC CCA GCC [CCA TCA CCA CCA CCA CCT TAC TAC

Y K S P P P P S P
TAC AAG TCT CCT CCT CCC CCA TCA CCA]$_8$ EGFP

C. (AP)$_4$(YK)$_4$EGFP

A P A P A P A P S P P P P Y Y
GCT CCA GCA CCT GCC CCA GCC [CCA TCA CCA CCA CCA CCT TAC TAC

Y K S P P P P S P
TAC AAG TCT CCT CCT CCC CCA TCA CCA]$_4$ EGFP

[0234] 2). Monosaccharide composition

[0235] Neutral sugar analyses and uronic acid assays showed that (AP)$_4$(YK)$_{20}$ was mainly arabinosylated, but also contained Rha, Ara, Gal, and GlcUA.

Table 19. Monosaccharide composition of (AP)$_4$(YK)$_{20}$EGFP

Monosaccharide	(AP)$_4$(YK)$_{20}$EGFP
	Molar percentage
Rha	1.5
Ara	80.2
Gal	11.7
GlcUA	6.5

[0236] 3). Hyp-glycoside profile.

[0237] The Hyp-glycoside profile of (AlaPro)$_4$(YK)$_{20}$EGFP showed that the glycoprotein was mainly glycosylated with arabinosides. The molar percentage of Hyp-polysaccharide was in consistent with the molar percentage of Hyp flanked by

Ala residues to the total Hyp, based on the gene sequence, implying that the Ala-Hyp repeat motifs were the sites of arabinogalactan polysaccharide addition.

Table 20. Hy-glycoside profile of $(AlaPro)_4(YK)_{20}EGFP$

Hyp Type	Molar %	Hyp Type	Molar %
Hyp-polysaccharide	4.86	Hyp-Ara_2	9.71
Hyp-Ara_4	45.67	Hyp-Ara	7.68
Hyp-Ara_3	29.49	Hyp	2.60

[0238] 4). Cross-linking.

[0239] Figure 40 shows size exclusion chromatography before (top frame) and after (bottom frame) 30 minutes of crosslinking of $(AP)_4(YK)_{20}$ catalyzed by extensin peroxidase. Monomeric $(AP)_4(YK)_{20}$ eluted at 37 minutes in the top panel. As monomeric $(AP)_4(YK)_{20}$ was crosslinked, it increased in size. After 30 minutes $(AP)_4(YK)_{20}$ was polymerized and voided the column (18 min).

[0240] Figure 41 shows chromatographs of the cross-linking reaction of $(AP)_4(YK)_{20}EGFP$. The upper and lower figures showed the size shift of $(AP)_4(YK)_{20}EGFP$ after crosslinked for 0 minutes and 30 minutes.

[0241] $(AP)_4(YK)_8$, and $(AP)_4(YK)_4$ were also crosslinked by extensin peroxidase, although the rates of crosslinking were slower.

[0242] **Example 7: Construction of Genes Encoding Cross-linkable Human Elastin – Arabinogalactan Protein Fusion Proteins**

[0243] Methods

[0244] I. Design of oligonucleotides encoding the human elastin motifs and repetitive Ala-Pro.

[0245] A set of oligonucleotides was designed based on the codons favored by tobacco cells. The oligonucleotide set encoded four repeats of human elastin

motif Val-Pro-Gly-Val-Gly (Reiersen, et al. 1998) which were evenly spaced by an AGP motif (three repeats of Ala-Pro) (Tan, et al. 2003).

[0246] The two sticky ends were designed to anneal to the vector fragments of *pUC-(YK)$_2$* or *pUC-(YK)$_4$* as BbsI and BsmF1 fragments (Held, 2004). Here, *YK* stands for the gene encoding the sequence: Ser-Pro-Pro-Pro-Pro-Ser-Pro-Ser-Pro-Pro-Pro-Pro-Tyr-Tyr-Tyr-Lys.

Oligonucleotide set:

P S G V G V P G V G A P A P
5`- CCA TCA GGA GTA GGT [GTT CCA GGA GTT GGC]$_2$[GCT CCA GCA CCT
3`-GT CCT CAT CCA [CAA GGT CCT CAA CCG]$_2$[CGA GGT CGT GGA
Sticky end Elastin Motif AGP motif

A P A G V G V P G V G
GCC CCA GCC] GGT GTT GGA [GTA CCT GGT GTT GGT]$_2$-3`
CGG GGT CGG] CCA CAA CCT [CAT GGA CCA CAA CCA]$_2$ GGT A-5`
Elastin Motif Sticky end

[0247] II. Construction of plasmids *pUC-(YK$_2$-E$_2$AE$_2$)$_4$*, *pUC-(YK$_4$-E$_2$AE$_2$)$_4$*, *pUC-(YK$_2$-E$_2$AE$_4$AE$_2$)$_2$*, and *pUC-(YK$_4$-E2AE4AE2)$_2$*. (Note: abbreviations *E$_2$AE$_2$* are E: elastin motif; A: AGP motif)

[0248] 1. Construction of plasmid *pUC-E$_2$AE$_2$* and *pUC-E$_2$AE$_4$AE$_2$*

[0249] Plasmid *pUC-YK$_8$* was digested by BbsI and BsmF1 to remove the *YK$_8$* gene (Held, 2004).

[0250] Fifty ng of above set of oligonucleotides were annealed to each other in 1X ligase buffer and ligated to the above *pUC* vector. The formed plasmid was named *pUC-E$_2$AE$_2$*. The sequence of *E$_2$AE$_2$* was verified by DNA sequencing.

[0251] Plasmid *pUC-E_2AE_2* was digested by two sets of restriction enzymes BbsI/ScaI and BsmFI/ScaI, respectively. The 1.1 kb BsmFI-ScaI fragment and 1.8 kb BbsI-ScaI fragments were ligated to each other to form double repeats of the gene cassette *E_2AE_2* (Figure 1) (Held, et al. 2004). The corresponding plasmid was named *pUC-$E_2AE_4AE_2$*. The sequence of *$E_2AE_4AE_2$* was confirmed by DNA sequencing.

[0252] Figure 42 shows a flow chart for gene construction. X represented for the desired gene. In this project, X can be *E_2AE_2*, or *YK_2*, or *YK_4*. The ligation could be between two same genes, or between two different genes.

[0253] 2. Construction of plasmids *pUC-YK_2-E_2AE_2*, *pUC-YK_4-E_2AE_2*, *pUC-YK_2-$E_2AE_4AE_2$*, and *pUC-YK_4-$E_2AE_4AE_2$*

[0254] a). By using the same strategy shown in Figure 42, the BsmFI-ScaI fragment of *pUC-YK_2* (Held 2004), containing the *YK_2* gene was ligated with the BbsI-ScaI fragment of *pUC-E_2AE_2* and *pUC-$E_2AE_4AE_2$*, respectively, which led to the formation of *pUC-YK_2-E_2AE_2* and *pUC-YK_2-$E_2AE_4AE_2$*.

[0255] b). As shown in Figure 42, the BsmFI-ScaI fragment of *pUC-YK_4* (Held, 2004), containing the *YK_4* gene was ligated with the BbsI-ScaI fragment of *pUC-E_2AE_2* and *pUC-$E_2AE_4AE_2$*, respectively, which lead to the formation of *pUC-YK_4-E_2AE_2* and *pUC-YK_4-$E_2AE_4AE_2$*.

[0256] The purpose of constructing genes encoding different sized cross-linking motifs and elastin/AGP motifs was to test the differences of physical properties between them. All the genes were confirmed by DNA sequencing.

[0257] 3. Construction of plasmids pUC-$(YK_2$-$E_2AE_2)_4$, pUC-$(YK_4$-$E_2AE_2)_4$, pUC-$(YK_2$-$E_2AE_4AE_2)_2$, and pUC-$(YK_4$-E2AE4AE2$)_2$.

[0258] a). By using the method shown in Figure 42, *pUC-YK_2-E_2AE_2*, *pUC-YK_2-$E_2AE_4AE_2$*, *pUC-YK_4-E_2AE_2*, and *pUC-YK_4-$E_2AE_4AE_2$* were dimerized,

respectively, with themselves. The desired gene sizes of *$(YK_2$-$E_2AE_4AE_2)_2$* and *$(YK_4$-E2AE4AE2)$_2$* were about 600 and 800 bps, respectively, which were long enough for expression by tobacco cells. The corresponding plasmids were named *pUC-$(YK_2$-$E_2AE_4AE_2)_2$*, and *pUC-$(YK_4$-E2AE4AE2)$_2$*.

[0259] b). The obtained *pUC-$(YK_2$-$E_2AE_2)_2$* and *pUC-$(YK_4$-$E_2AE_2)_2$* were further dimerized by the same method shown above, respectively. The formed *pUC-$(YK_2$-$E_2AE_2)_4$*, *pUC-$(YK_4$-$E_2AE_2)_4$* had the gene cassettes of sizes about 800 and 1200 bps, respectively.

[0260] All the sequences were confirmed by DNA sequencing.

[0261] III. Construction of plasmids pBI-SS^{tob}-$(YK_2$-$E_2AE_2)_4$-EGFP, pBI-SS^{tob} -$(YK_4$-$E_2AE_2)_4$-EGFP, pBI-SS^{tob} -$(YK_2$-$E_2AE_4AE_2)_2$-EGFP, and pBI-SS^{tob} -$(YK_4$-E2AE4AE2$)_2$-EGFP.

[0262] 1. The constructed gene fragments in *pUC-$(YK_2$-$E_2AE_2)_4$*, *pUC-$(YK_4$-$E_2AE_2)_4$*, *pUC-$(YK_2$-$E_2AE_4AE_2)_2$*, and *pUC-$(YK_4$-E2AE4AE2$)_2$* were removed from the plasmids by the restriction enzymes XmaI and NcoI, and ligated into *pUC-SS^{tob}-$(AP)_{11}$-EGFP* by replacing the *$(AP)_{11}$* fragment (flanked by XmaI and NcoI restriction sites), respectively (Shpak, et al., 1999, 2001, Tan, et al 2003, Tan 2003). The four new plasmids were named *pUC-SS^{tob}-$(YK_2$-$E_2AE_2)_4$-EGFP*, *pUC-SS^{tob} -$(YK_4$-$E_2AE_2)_4$-EGFP*, *pUC-SS^{tob} -$(YK_2$-$E_2AE_4AE_2)_2$-EGFP*, and *pUC-SS^{tob} -$(YK_4$-E2AE4AE2$)_2$-EGFP*, respectively.

[0263] 2. The gene cassettes in above four new *pUC* plasmids were further removed from original plasmids by BamHI and SacI, and ligated into *pBI 121* between the BamHI and SacI restriction sites (Shpak, et al., 1999, 2001, Tan, et al 2003, Tan 2003). Then, we obtained the plant transformation plasmids: *pBI-SS^{tob}-*

$(YK_2$-$E_2AE_2)_4$-EGFP, *pBI-SS^{tob} -$(YK_4$-$E_2AE_2)_4$-EGFP*, *pBI-SS^{tob} -$(YK_2$-$E_2AE_4AE_2)_2$-EGFP*, and *pBI-SS^{tob} -$(YK_4$-E2AE4AE2$)_2$-EGFP*.

[0264] All gene sequences were confirmed by DNA sequencing with a primer derived from tobacco extensin signal sequence.

[0265] IV. Transfer the genes into tobacco cells.

[0266] 1. Each 100 ng of above-modified *pBI* plasmid was transformed into Agrobacterium by the Freeze-thaw method (An, et al. 1988). The positive colonies of agrobacterium were selected by Kanamycin/Streptomycin resistance.

[0267] 2. Overnight-growed transformed agrobacteria were used to co-culture with 4-day cultured tobacco cells (in SH media) at 28°C for 2 days.

[0268] 3. The infected tobacco cells were washed 4 times with SH media to remove the excess agrobacteria. The tobacco cells were spread on SH solid plates with Kanamycin (100 μg/ml) and Timentin (200 μg/ml).

[0269] The following documents, some of which have been cited and referred to herein, are considered part of this disclosure and are incorporated herein by reference in their entirety.

[0270] Albersheim, P. et al. "A method for the analysis of sugars in plant cell-wall polysaccharides by gas liquid chromatography." Carbohyd.Res. 5 (1967): 340-45.

[0271] An, G. et al. "Binary Vectors." Dordrecht, Netherlands: Martinus Nijhoff, 1988. 1-19.

[0272] Benhamou, N. et al. "Differential accumulation of hydroxyproline-rich glycoproteins in bean root nodule cells infected with a wild-type strain or C4-

dicarboxylic acid mutant of *Rhizobium leguminosarum* bv. *phaseoli*." Planta 184 (1991): 457-67.

[0273] Bergman, T., M. Carlquist, and H. Jornvall. "Amino Acid Analysis by High Performance Liquid Chromatography of Phenylthiocarbamyl Derivatives." Ed. B. Wittmann-Liebold. Berlin: Springer Verlag, 1986. 45-55.

[0274] Blumenkrantz, N. and G. Asboe-Hansen. "New methods for quantitative determination of uronic acids." Anal.Biochem. 54 (1973): 484-89.

[0275] Brady, J. D. and S. C. Fry. "Formation of di-isodityrosine and loss of isodityrosine in the cell walls of tomato cell-suspension cultures treated with fungal elicitors or H2O2." Plant Physiol. 115.1 (1997): 87-92.

[0276] Brady, J. D., I. H. Sadler, and S. C. Fry. "Di-isodityrosine, a novel tetrameric derivative of tyrosine in plant cell wall proteins: a new potential cross-link." Biochem.J. 315 (1996): 323-27.

[0277] Brady, J. D., I. H. Sadler, and S. C. Fry. "Pulcherosine, an oxidatively coupled trimer of tyrosine in plant cell walls: its role in cross-link formation." Phytochemistry **47** (1998): 349-53.

[0278] Brownleader, M. D. et al. "Purification and partial characterization of tomato extensin peroxidase." Plant Physiol. 109 (1995): 1115-23.

[0279] Carpita, N. C. and D. M Gibeaut. "Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth." Plant J. 3 (1993): 1-30.

[0280] Cassab, G. I. "Arabinogalactan proteins during the development of soybean (*Glycine max*) root nodules." Planta 168 (1986): 441-46.

[0281] Chen, J. and J. E. Varner. "An extracellular matrix protein in plants: characterization of a genomic clone for carrot extensin." EMBO J. 4 (1985): 2145-51.

[0282] Corbin, D. R., N. Sauer, and C. J Lamb. "Differential regulation of a hydroxyproline-rich glycoprotein gene family in wounded and infected plants." Mol.Cell.Biol. 7 (1987): 4337-44.

[0283] De Loose, M. et al. "The extensin signal peptide allows secretion of a heterologous protein from protoplasts." Gene 99 (1991): 95-100.

[0284] Epstein, L. and D. T. A. Lamport. "An intramolecular linkage involving isodityrosine in extensin." Phytochemistry 23 (1984): 1241-46.

[0285] Everdeen, D. S. et al. "Enzymic crosslinkage of monomeric extensin precursors *in vitro*." Plant Physiol. 87 (1988): 616-21.

[0286] Fong, C. et al. "A gymnosperm extensin contains the serine-tetrahydroxyproline motif." Plant Physiol. 99 (1992): 548-52.

[0287] Franssen, H. J. et al. "Characterization of cDNA for nodulin-75 of soybean: A gene product involved in early stages of root nodule development." Proc.Natl.Acad.Sci. 84 (1987): 4495-99.

[0288] Frueauf, J. B. et al. "Peptides isolated from cell walls of *Medicago truncatula* nodules and uninfected root." Phytochemistry 55.5 (2000): 429-38.

[0289] Fry, S. C. "Isodityrosine, a new cross-linking amino acid from plant cell-wall glycoprotein." Biochem.J. 204 (1982): 449-55.

[0290] Fry, S. C. "Formation of isodityrosine by peroxidase isozymes." J.Exp.Bot. 190 (1987): 853-62.

[0291] Fujimoto, D. Horiuchi K. and Hirama M. Biochem.Biophys, Res.Commun. 99 (1981): 637-43.

[0292] Fujimoto, D. Horiuchi K. and Hirama M. Biochem.Biophys, Res.Commun. 99 (1981): 637-43.

[0293] Goodenough, U. W. et al. "Crystals of the *Chlamydomonas reinhardtii* cell wall: polymerization, depolymerization and purification of glycoprotein monomers." J.Cell Biol. 103 (1986): 405-17.

[0294] Hall, Q. and M. C. Cannon. "The cell wall hydroxyproline-rich glycoprotein RSH is essential for normal embryo development in Arabidopsis." Plant Cell 14 (2002): 1161-72.

[0295] Han, Y. et al. "Hydroxyproline-rich glycoproteins expressed during stress responses in cassava." Euphytica 120 (2001): 59-70.

[0296] Heckman, J. W., B. T. Terhune, and D. T. A. Lamport. "Characterization of native and modified extensin monomers and oligomers by electron microscopy and gel filtration." Plant Physiol. 86 (1988): 848-56.

[0297] Held MA, Tan L, Kamyab A, Hare M, Shpak E, and Kieliszewski MJ (2004) Di-isodityrosine is the intermolecular crosslink of isodityrosine-rich extensin analogs crosslinked in vitro. *J. Biol. Chem.* 279: 55474-55482.

[0298] Hirsinger, Cathy et al. "The tobacco extensin gene Ext 1.4 is expressed in cells submitted to mechanical constraints and in cells proliferating under hormone control." J.Exp.Bot. 50 (1999): 343-55.

[0299] Jackson, P. A. P. et al. "Rapid deposition of extensin during the elicitation of grapevine callus cultures is specifically catalyzed by a 40-kilodalton peroxidase." Plant Physiol. 127 (2001): 1065-76.

[0300] Keegstra, K. et al. "The structure of plant cell walls. III. A model of the walls of suspension-cultured sycamore cells based on the interconnections of the macromolecular components." Plant Physiol. 51 (1973): 188-96.

[0301] Kerr, T. and I. W. Bailey. "The cambium and its derivative tissues. X Structure, optical properties and chemical composition of the so-called middle lamella." Journal of the Arnold Arboretum 15 (1934): 327-49.

[0302] Kieliszewski, M. J. "The latest hype on Hyp-O-glycosylation codes." Phytochemistry 57.3 (2001): 319-23.

[0303] Kieliszewski, M. J. et al. "A histidine-rich extensin from *Zea mays* is an arabinogalactan protein." Plant Physiol. 99 (1992): 538-47.

[0304] Kieliszewski, M. J. and D. T. A. Lamport. "Extensin: repetitive motifs, functional sites, posttranslational codes and phylogeny." Plant J. 5 (1994): 157-72.

[0305] Kieliszewski, M. J. et al. "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 (1995): 2541-49.

[0306] Knupp, C. and J. M. Squire. "Molecular packing in network-forming collagens." The Scientific World 3 (2003): 558-77.

[0307] Lamport, D. T. A. "Oxygen fixation into hydroxyproline of plant cell wall protein." J.Biol.Chem. 238 (1963): 1438-40.

[0308] Lamport, D. T. A. "The protein component of primary cell walls." Adv.Bot.Res. 2 (1965): 151-218.

[0309] Lamport, D. T. A. "Hydroxyproline-O-glycosidic linkage of the plant cell wall glycoprotein extensin." Nature 216 (1967): 1322-24.

[0310] Lamport, D. T. A. "The isolation and partial characterization of hydroxyproline-rich glycopeptides obtained by enzymic degradation of primary cell walls." Biochemistry 8 (1969): 1155-63.

[0311] Lamport, D. T. A. "Is the primary cell wall a protein-glycan network?" 1st International Protoplast Colloquium, Versailles.I.N.R.A. (1973): 27-31.

[0312] Lamport, D. T. A. "Structure, Biosynthesis and Significance of Cell Wall Glycoproteins." Ed. F. A. Loewus and V. C. Runeckles. New York: Plenum Publishing Corp., 1977. 79-115.

[0313] Lamport, D. T. A. and L. Epstein. "A new model for the primary cell wall: a concatenated extensin- cellulose network." Proc.Ann.Plant Biochem.& Physiol.Symp.Columbia-Missouri. 2 (1983): 73-83.

[0314] Lamport, D. T. A., L. Katona, and S. Roerig. "Galactosyl serine in extensin." Biochem.J. 133 (1973): 125-31.

[0315] Lamport, D. T. A. and D. H. Miller. "Hydroxyproline arabinosides in the plant kingdom." Plant Physiol. 48 (1971): 454-56.

[0316] Lamport, D. T. A. and D. H. Northcote. "Hydroxyproline in primary cell walls of higher plants." Nature 188 (1960): 665-66.

[0317] Lamport, D. T. A. and D. H. Northcote. "The use of tissue cultures for the study of plant cell walls." Biochem.J. 76 (1960): 52P.

[0318] Lewis, R. V. et al. "Expression and purification of a spider silk protein: a new strategy for producing repetitive proteins." Protein Expression Purif. 7 (1996): 400-06.

[0319] Li, B and Daggett, V (2002) Molecular basis for the extensibility of elastin. *J. Muscle Res. and Cell Motility.* 23: 561-573

[0320] McCormick, S. et al. "Leaf disc transformation of cultivated tomato (*L. esculentum*) using *Agrobacterium tumefaciens*." Plant Cell Reports 5 (1986): 81-84.

[0321] McGrath, K. P. et al. "Chemical and biosynthetic approaches to the production of novel polypeptide materials." Biotechnol.Prog. 6 (1990): 186-92.

[0322] Merkouropoulos, G. and A. H. Shirsat. "The unusual Arabidopsis extensin gene atExt1 is expressed throughout plant development and is induced by a variety of biotic and abiotic stresses." Planta 217 (2003): 356-66.

[0323] Merkouropoulos, Georgios, David C. Barnett, and Anil H. Shirsat. "The Arabidopsis extensin gene is developmentally regulated, is induced by wounding, methyl jasmonate, abscisic, and salicylic acid and codes for a protein with unusual motifs." Planta 208.2 (1999): 212-19.

[0324] Mort, A. J. "Partial Characterization of Extensin by Selective Degradation of Cell Walls." Diss. Michigan State University, 1978.

[0325] Mort, A. J. and D. T. A. Lamport. "Anhydrous hydrogen fluoride deglycosylates glycoproteins." Analyt.Biochem. 82 (1977): 289-309.

[0326] Price, N. J. et al. "A Biochemical and Molecular Charaterization of LEP1, an Extensin Peroxidase from Lupin." J.Biol.Chem. 278.42 (2003): 41389-99.

[0327] Qi, X. et al. "Solubilization and partial characterization of extensin fragments from cell walls of cotton suspension cultures. Evidence for a covalent cross-link between extensin and pectin." Plant Physiol. 108 (1995): 1691-701.

[0328] Roberts, K. "Crystalline glycoprotein cell walls of algae: Their structure, composition and assembly." Phil.Trans.R.Soc.Lond.B 268 (1974): 129-46.

[0329] Sanger, M. P. and D. T. A. Lamport. "A micro-apparatus for liquid hydrogen fluoride solvolysis: sugar and amino sugar composition of *Erysiphe graminis* and *Triticum aestivum* cell walls." Analyt.Biochem. 128 (1983): 66-70.

[0330] Schnabelrauch, L. S. et al. "Isolation of pI 4.6 extensin peroxidase from tomato cell suspension cultures and identification of Val-Tyr-Lys as putative intermolecular cross-link site." Plant J. 9 (1996): 477-89.

[0331] Shirsat, A. H. et al. "The Brassica napus extA extensin gene is expressed in regions of the plant subject to tensile stresses." Planta 199 (1996): 618-24.

[0332] Showalter, A. M. and D. Rumeau. "Molecular Biology of Plant Cell Wall Hydroxyproline-Rich Glycoproteins." Ed. W. S. Adair and R. P. Mecham. New York: Academic Press, 1990. 247-81.

[0333] Showalter, A. M. and J. E. Varner. "Molecular Details of Plant Cell Wall Hydroxyproline-Rich Glycoproteins." Ed. C. J. Arntzen and Ryan C. New York: Alan R Liss, 1987. 375-92.

[0334] Showalter, A. M. and J. E. Varner. "Plant Hydroxyproline-Rich Glycoproteins." Ed. J. Preiss. New York: Academic Press, 1989. 485-520.

[0335] Showalter, A. M. et al. "Tomato extensin and extensin-like cDNAs: structure and expression in response to wounding." Plant Mol.Biol. 16 (1991): 547-65.

[0336] Shpak, E. et al. "Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in *Nicotiana tabacum*." J.Biol.Chem. 276 (2001): 11272-78.

[0337] Shpak, E., J. Leykam, and M. J. Kieliszewksi. "Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-O-glycosylation codes." Proc.Natl.Acad, Sci. 96 (1999): 14736-41.

[0338] Shpak, Elena. "Synthetic Genes for the Elucidation of Hydroxyproline O-Glycosylation Codes." Diss. University of Ohio, 2000.

[0339] Shpak, Elena, Joseph F. Leykam, and Marcia J. Kieliszewski. "Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-O-glycosylation codes." Proc.Natl.Acad.Sci.U.S.A. 96.26 (1999): 14736-41.

[0340] Smith, J. J., E. P. Muldoon, and D. T. A. Lamport. "Isolation of extensin precursors by direct elution of intact tomato cell suspension cultures." Phytochemistry 23 (1984): 1233-39.

[0341] Smith, J. J. et al. "Tomato extensin precursors P1 and P2 are highly periodic structures." Phytochemistry 25 (1986): 1021-30.

[0342] Stafstrom, J. P. and L. A. Staehelin. "The role of carbohydrate in maintaining extensin in an extended conformation." Plant Physiol. 81 (1986): 242-46.

[0343] Stephen C.Fry. The Growing Plant Cell Wall: Chemical and Metabolic Analysis. Ed. M Wilkens. Essex: Longman Scientific & Technical; John Wiley and Sons, Inc., 1988.

[0344] Tan, L. (2003) Dissertation: O-glycosylation motifs in arabinogalactan-proteins.

[0345] Tan L, Leykam JF Kieliszewski MJ. "Glycosylation motifs that direct arabinogalactan addition to arabinogalactan-proteins." Plant Physiol. 132.3 (2003): 1362-69.

[0346] Tan, L., J. Leykam, and M. J. Kieliszewski. "Glycosylation motifs that direct arabinogalactan addition to arabinogalactan-proteins." Plant Physiol. 132 (2003): 1362-69.

[0347] van Holst, G.-J. and J. E. Varner. "Reinforced polyproline II conformation in a hydroxyproline-rich glycoprotein from carrot root." Plant Physiol. 74 (1984): 247-51.

[0348] Welinder, K. G. et al. "Structural diversity and transcription of class III peroxidases from Arabidopsis thaliana." Eur.J.Biochem. 269 (2002): 6063-81.

[0349] Whitesides, G. M., J. P. Mathias, and C. T. Seto. "Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures." Science 254 (1991): 1312-19.

[0350] Yoshiba, Yoshu et al. "Characterization of four extensin genes in Arabidopsis thaliana by differential gene expression under stress and non-stress conditions." DNA Research 8.3 (2001): 115-22.

[0351] Zhao, Z. D. et al. "Tomato LeAGP-1 arabinogalactan-protein purified from transgenic tobacco corroborates the Hyp contiguity hypothesis." Plant J. 31.4 (2002): 431-44.

[0352] Zhou, J., D. Rumeau, and A. M. Showalter. "Isolation and characterization of two wound-regulated tomato extensin genes." Plant Mol.Biol. 20 (1992): 5-17.

[0353] While particular embodiments of the subject invention have been described, it will be obvious to those skilled in the art that various changes and modifications of the subject invention can be made without departing from the spirit and scope of the invention. In addition, while the present invention has been described in connection with certain specific embodiments thereof, it is to be understood that this is by way of illustration and not by way of limitation and the scope of the invention is defined by the appended claims which should be construed as broadly as the prior art will permit.

[0354] The disclosure of all patents, patent applications (and any patents which issue thereon, as well as any corresponding published foreign patent applications), and publications mentioned throughout this description are hereby incorporated by reference herein. It is expressly not admitted, however, that any of

the documents incorporated by reference herein teach or disclose the present invention.

[0355] It should be understood that every maximum numerical limitation given throughout this specification will include every lower numerical limitation, as if such lower numerical limitations were expressly written herein. Every minimum numerical limitation given throughout this specification will include every higher numerical limitation, as if such higher numerical limitations were expressly written herein. Every numerical range given throughout this specification will include every narrower numerical range that falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein.

[0356] Except where otherwise indicated, all numbers expressing quantities of ingredients, reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about." Accordingly, unless indicated to the contrary, the numerical parameters set forth in the following specification and attached claims are approximations that may vary depending upon the desired properties sought to be obtained by the present invention. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should be construed in light of the number of significant digits and ordinary rounding approaches.

[0357] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of the ordinary skill in the art to which this invention belongs. The terminology used in the description of the invention herein is for describing particular embodiments only and is not intended to be limiting of the invention. As used in the description of the invention and the

appended claims, the singular forms "a," "an," and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety.

[0358] The specification is most thoroughly understood in light of the teachings of the references cited within the specification, all of which are hereby incorporated by reference in their entirety. The embodiments within the specification provide an illustration of embodiments of the invention and should not be construed to limit the scope of the invention. The skilled artisan recognizes that many other embodiments are encompassed by the claimed invention and that it is intended that the specification and examples be considered as exemplary only, with a true scope and spirit of the invention being indicated by the following claims.

WHAT IS CLAIMED IS:

1. A non-naturally occurring protein comprising at least one modification, wherein the at least one modification comprises the addition of the amino acid sequence Tyr-X-Tyr, or a substitution or deletion to result in the presence of Tyr-X-Tyr in the sequence, wherein X is chosen from any amino acid, and wherein the protein, in the presence of oxidizing conditions, forms at least one of intermolecular crosslinks and intramolecular crosslinks, and wherein the .

2. The non-naturally occurring protein according to claim 1, wherein X is chosen from Tyr, Lys, and Val.

3. The non-naturally occurring protein according to claim 1, wherein the amino acid squence comprises Tyr-X-Tyr-Lys.

4. The non-naturally occurring protein according to claim 1, wherein the protein comprises two or more Tyr-X-Tyr motifs.

5. The non-naturally occurring protein according to claim 1, intermolecularly crosslinked to a non-naturally occurring protein comprising the amino acid sequence Tyr-X-Tyr.

6. The non-naturally occurring protein according to claim 1, intramolecularly crosslinked to itself.

7. The non-naturally occurring protein according to claim 1, further comprising the amino acid sequence X-Hyp_n, wherein X is any amino acid, and n is from 1 to about 1000.

8. The non-naturally occurring protein according to claim 7, wherein X is chosen from Ser, Ala, Val, and Thr.

9. The non-naturally occurring protein according to claim 1, further comprising the amino acid sequence X-Hyp-X-Hyp, wherein X is any amino acid.

10. The non-naturally occurring protein according to claim 9, wherein X is chosen from Ser, Ala, Val, and Thr.

11. A non-naturally occurring protein comprising the amino acid sequence Val-Tyr-Lys.

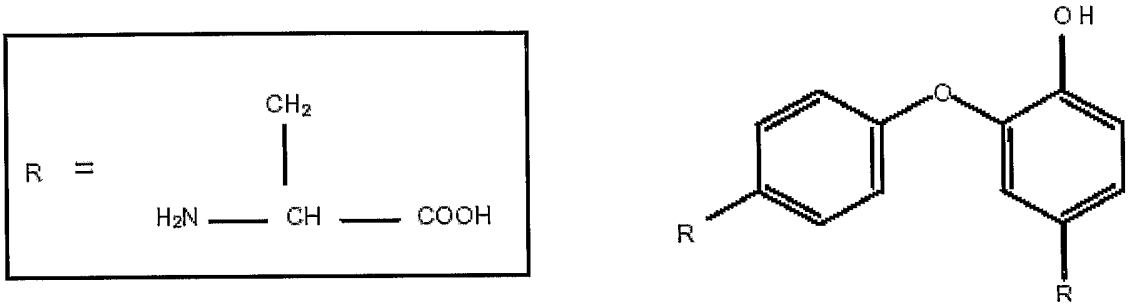


Figure 1

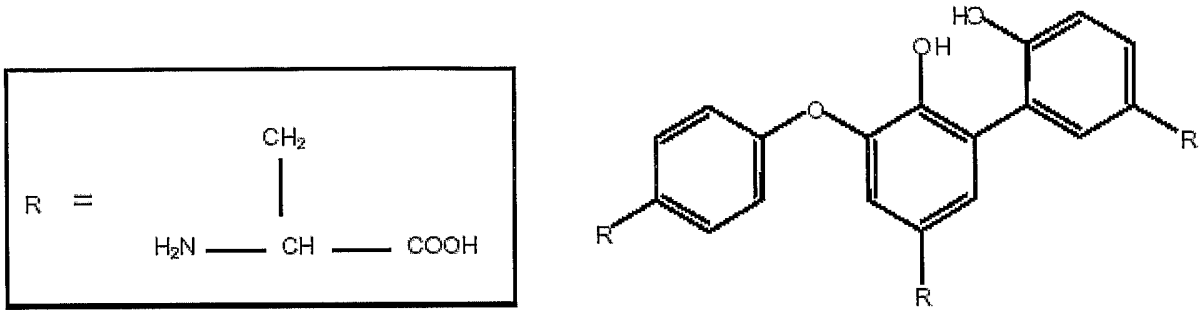


Figure 2

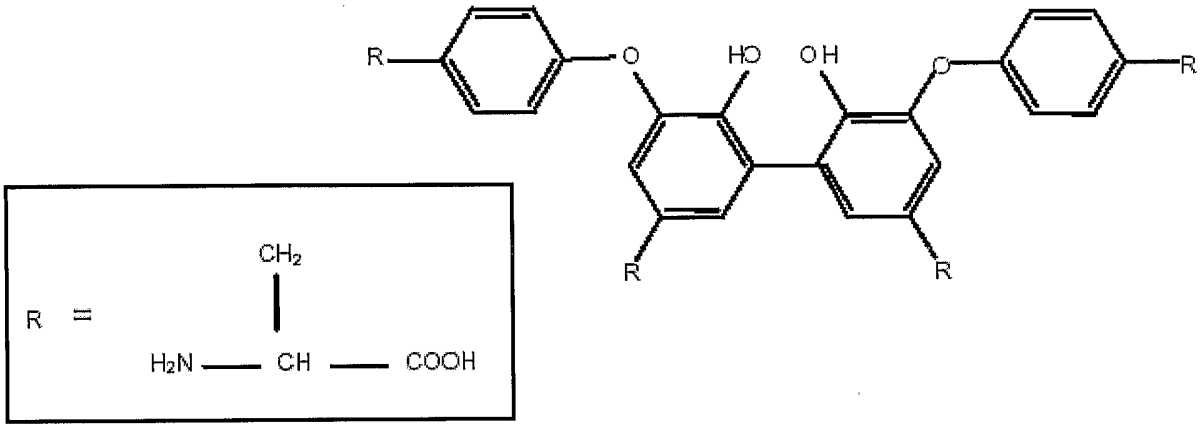


Figure 3

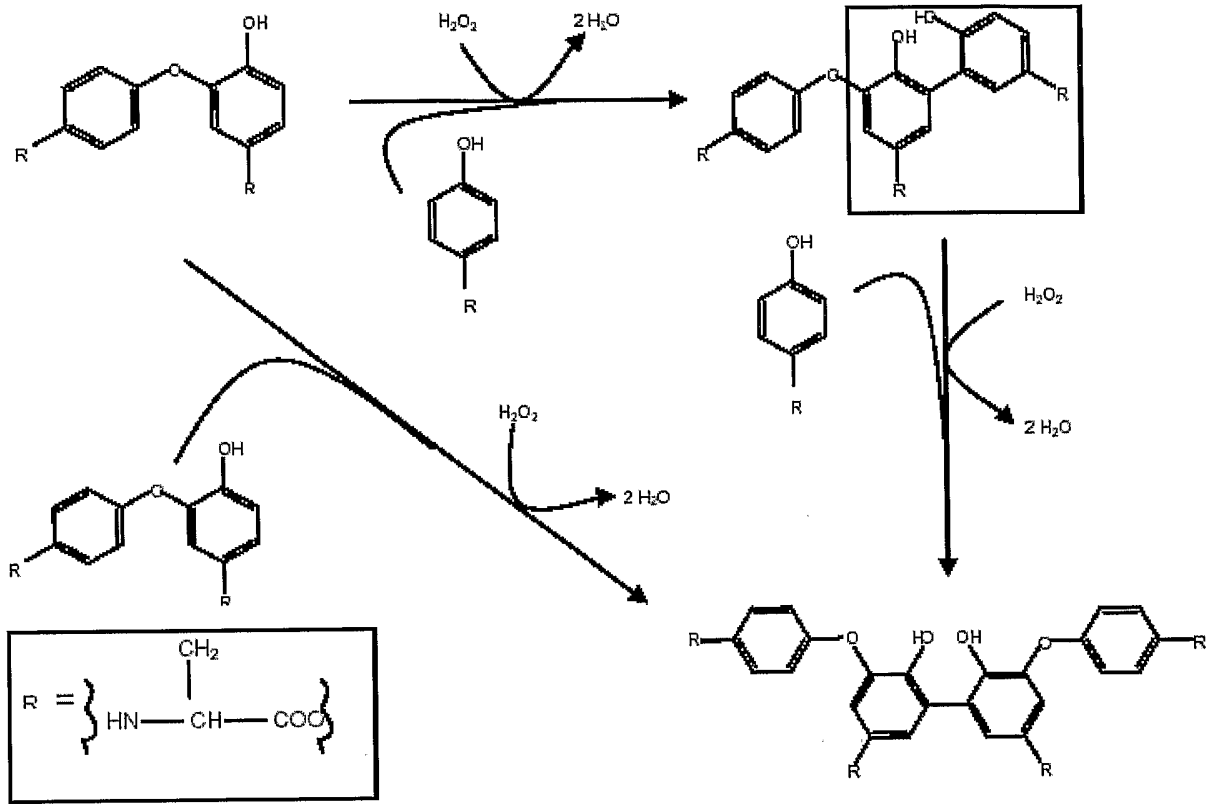

OH
R
H_2O_2
2 H_2O
HO
OH
R
R
OH
R
H_2O_2
2 H_2O
H_2O_2
2 H_2O
OH
R
R = CH2 HN CH CO
HO
OH
R
R
R
R

Figure 4

(A) *FK oligonucleotide pairs*

5' Linker pair (5'SP3-s and 5'SP3-as)

BamHI XmaI
5'-GCT GCC GGA TCC TCA ACC CGG GCC TCA CCA-3'
3'-CGA CGG CCT AGG AGT TGG GCC CCG AGT GGT GGT GGT GGA-5'
A A G S S T R A S P

FK Internal repeat pair (FK-s and FK-as)

5'-CCA CCA CCT AGT CCG TCA CCC CCT CCC CCA TTT TTC TTT AAG AGC CCA
TCA GGC AGT GGG GGA GGG GGT AAA AAG AAA TTC TCG GGT GGT GGT GGA-5'
P P P S P S P P P P F F F K S P

3'-Linker pair (3'SP4-s and 3'SP4-as)

NcoI SacI EcoRI
5'-CCA CCA CCT TCC ATG GCA TAA TAG AGC TCG AAT TCC TGG CCC ACC-3'
3'- AGG TAC CGT ATT ATC TCG AGC TTA AGG ACC GGG TGG-5'
P P P S M A - -

(B) *YK oligonucleotide pair*

YK-Sense oligonucleotide (YK-ss)

XmaI BbsI
5'-GGG ACC CGG GGA AGA CGA CCA TCA CCA CCA CCA CCT TAC TAC TAC AAG TCT CCT-3'
G T R G R R P S P P P P Y Y Y K S P

YK-Antisense oligonucleotide (YK-as)

NcoI BsmFI SacI
3'-GGT GGA ATG ATG ATG TTC AGA GGA GGA GGG GGT AGT GGT AGT TGG TAC CGC AGG GCC ACT ATC CTC GAG TTA-5'
P P Y Y Y K S P P P P S P S T M A S R - - E L N

(C) *YL oligonucleotide pair*

YL-Sense oligonucleotide (YL-ss)

XmaI BbsI
5'-GGG ACC CGG GGA AGA CGA CCA TCA CCA CCA CCA CCT TAC TAC TAC CTA TCT CCT-3'
G T R G R R P S P P P P Y Y Y L S P

YL-Antisense oligonucleotide (YL-as)

NcoI BsmFI SacI
3'-GGT GGA ATG ATG ATG GAT AGA GGA GGA GGG GGT AGT GGT AGT TGG TAC CGC AGG GCC ACT ATC CTC GAG TTA-5'
P P Y Y Y L S P P P P S P S T M A S R - - E L N

(D) *FL oligonucleotide pair*

FL-Sense oligonucleotide (FL-ss)

XmaI BbsI
5'-GGG ACC CGG GGA AGA CGA CCA TCA CCA CCA CCA CCT TTT TTC TTT CTA TCT CCT-3'
G T R G R R P S P P P P F F F L S P

FL-Antisense oligonucleotide (FL-as)

NcoI BsmFI SacI
3'-GGT GGA AAA AAG AAA GAT AGA GGA GGA GGG GGT AGT GGT AGT TGG TAC CGC AGG GCC ACT ATC CTC GAG TTA-5'
P P F F F L S P P P P S P S T M A S R - - E L N

Figure 5

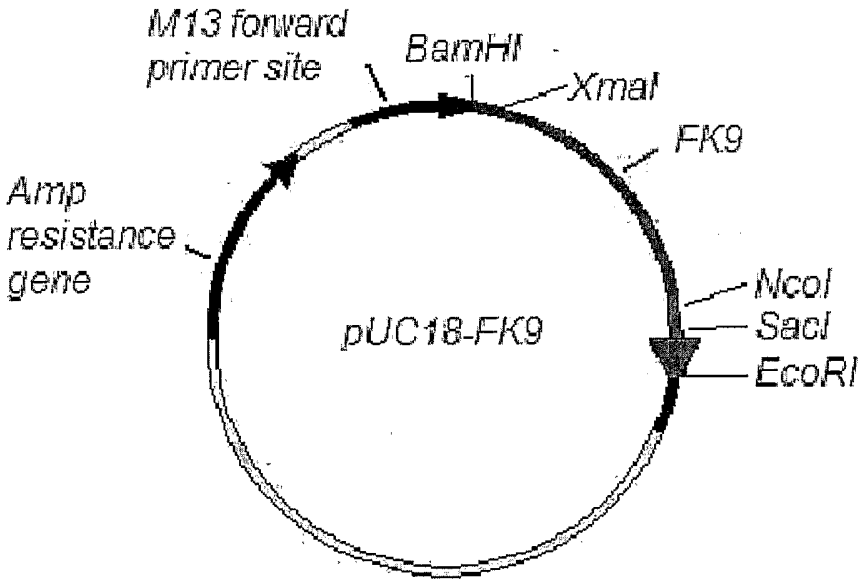

M13 forward
primer site
BamHI
XmaI
FK9
Amp
resistance
gene
pUC18-FK9
NcoI
SacI
EcoRI

Figure 6

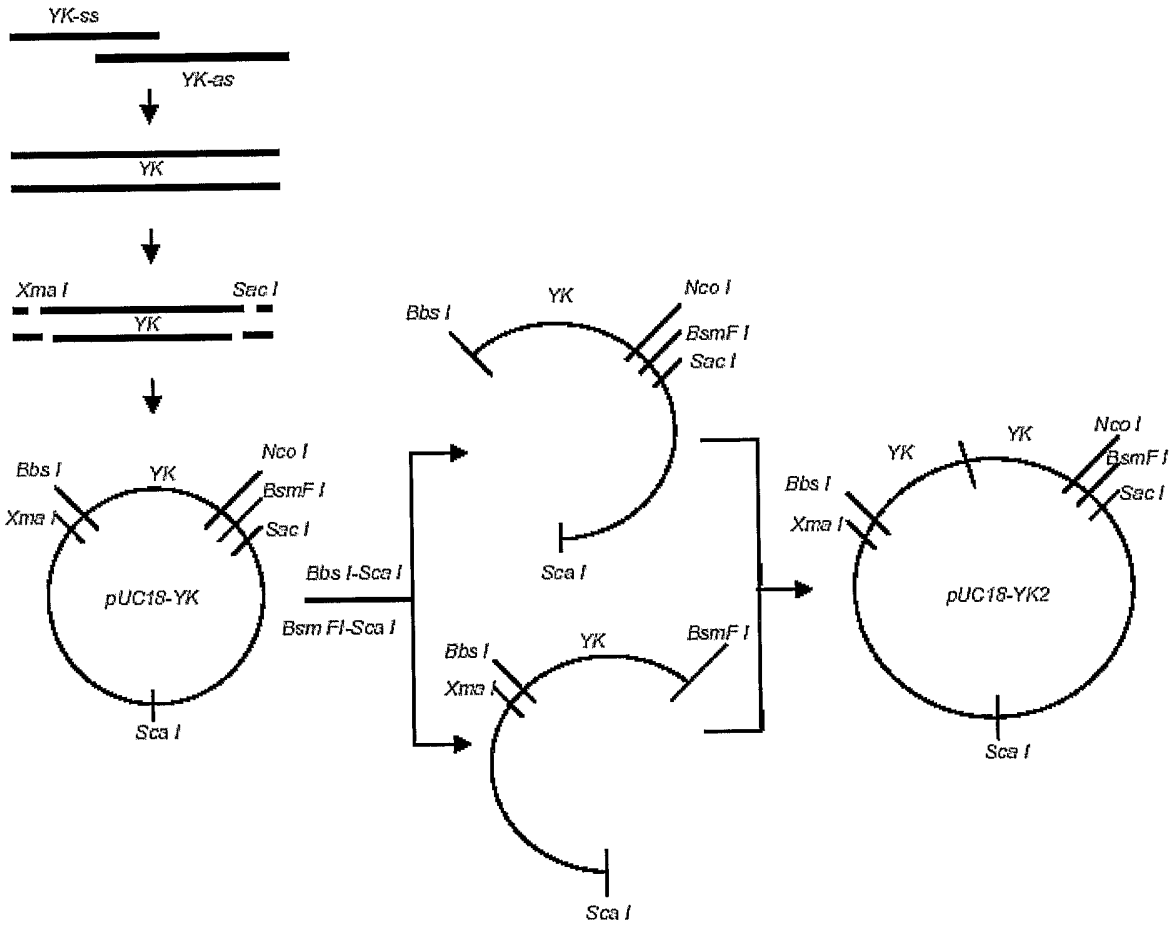

YK-ss
YK-as
YK
Xma I
Sac I
YK
Bbs I
YK
Nco I
BsmF I
Sac I
Xma I
pUC18-YK
Sca I
Bbs I-Sca I
Bsm FI-Sca I
Bbs I
YK
Nco I
BsmF I
Sac I
Sca I
Bbs I
Xma I
YK
BsmF I
Sca I
YK
YK
Nco I
BsmF I
Sac I
Bbs I
Xma I
pUC18-YK2
Sca I

Figure 7

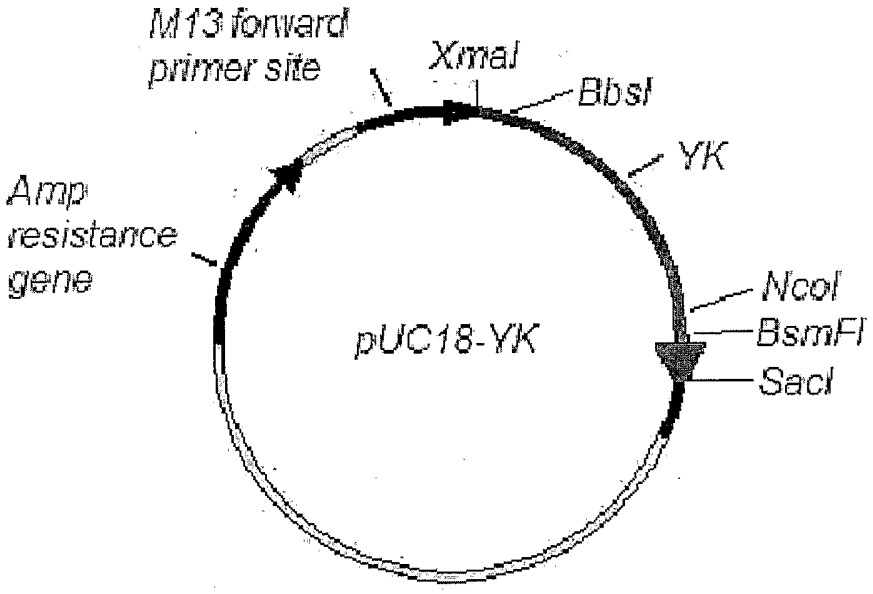

M13 forward primer site
XmaI
BbsI
YK
Amp resistance gene
pUC18-YK
NcoI
BsmFI
SacI

Figure 8

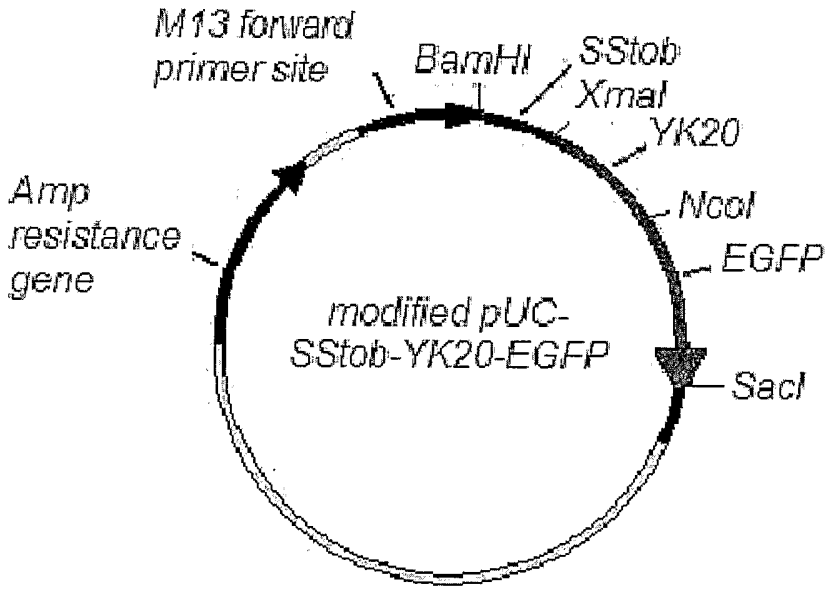

M13 forward primer site
BamHI
SStob
XmaI
YK20
NcoI
EGFP
SacI
Amp resistance gene
modified pUC-SStob-YK20-EGFP

Figure 9

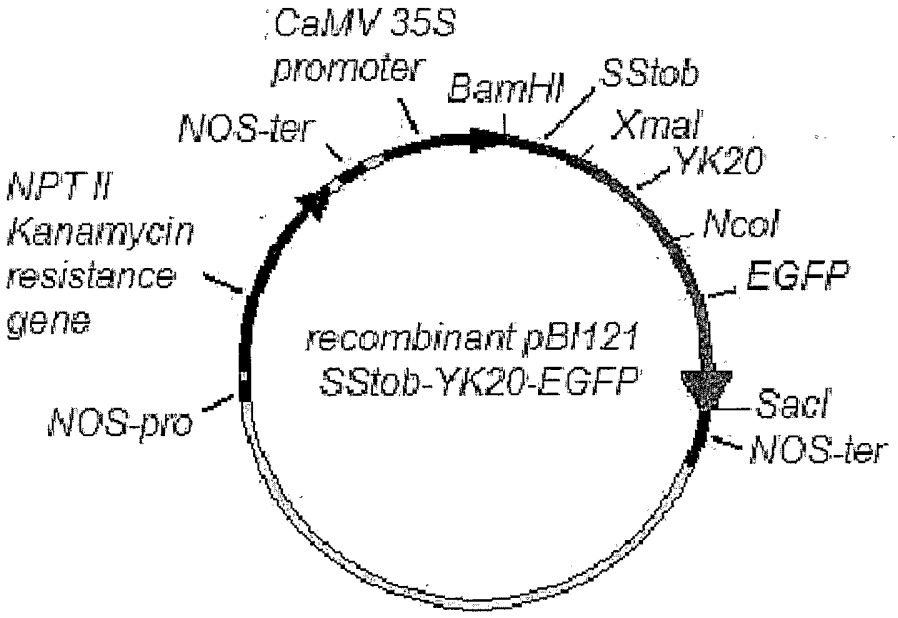

CaMV 35S
promoter
BamHI
SStob
XmaI
YK20
NcoI
EGFP
SacI
NOS-ter
NOS-pro
NPT II
Kanamycin
resistance
gene
NOS-ter
recombinant pBI121
SStob-YK20-EGFP

Figure 10

(a) *M13 forward* oligonucleotide primer

5'-GTA AAA CGA CGG CCA GTG-3'

(b) *Tobacco signal sequence* (*SSeq4Sen)* oligonucleotide primer

5'-CGG CAA TGG GAA AAA TGG CTT CT-3'

(c) *EGFP* (*EGFP566as*) oligonucleotide primer

5'-GAA GAT GGT GCG CTC CTG GAC GT-3'

Figure 11

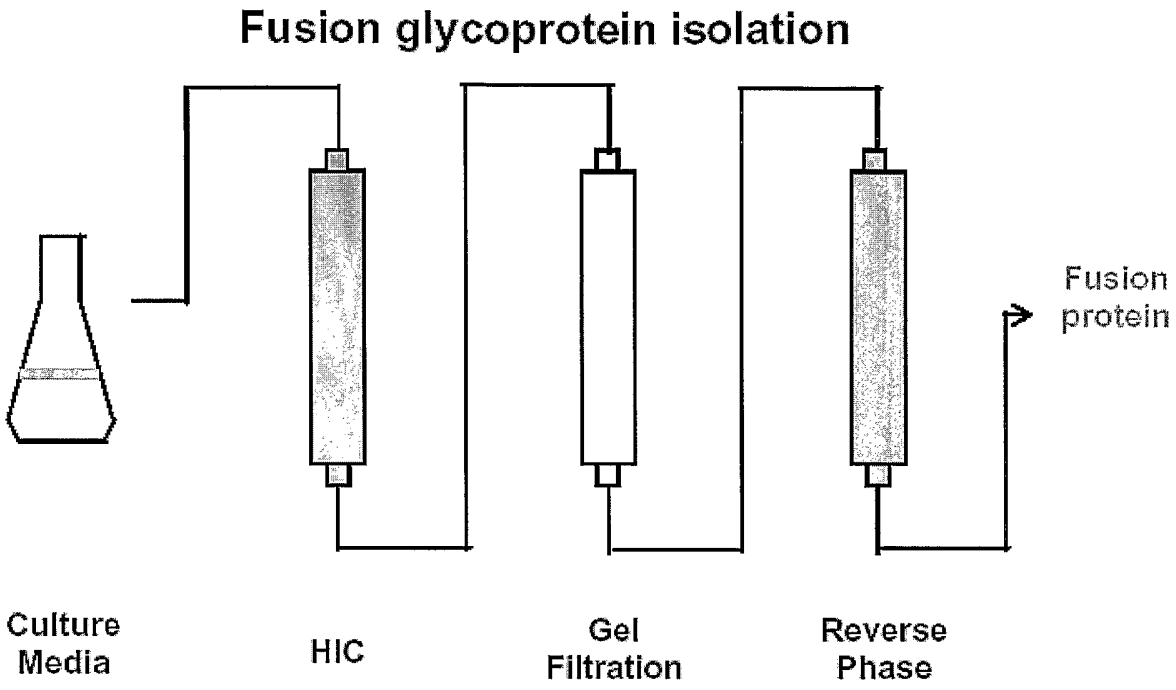

Fusion glycoprotein isolation
Fusion protein
Culture Media
HIC
Gel Filtration
Reverse Phase

Figure 12

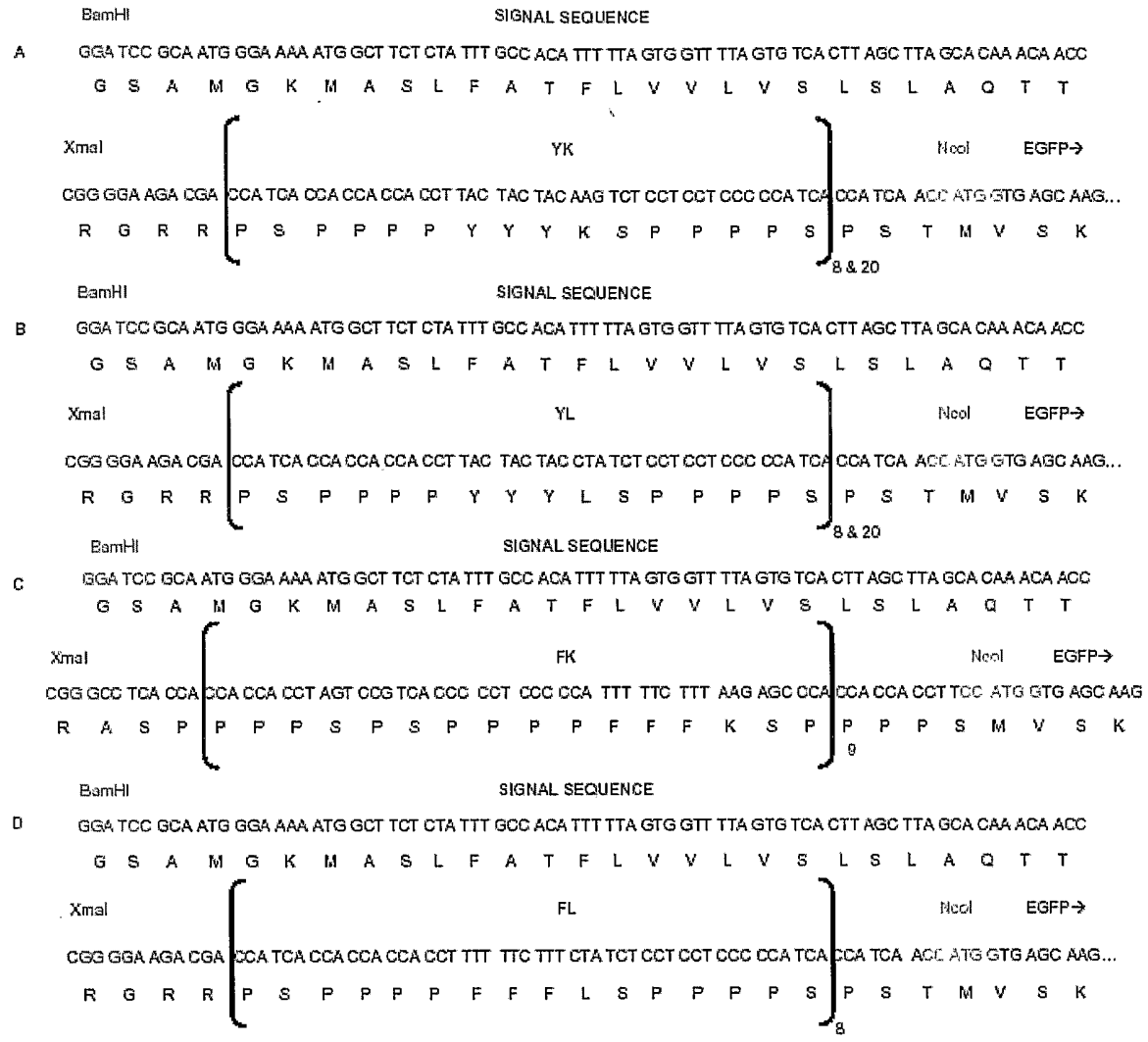

BamHI
SIGNAL SEQUENCE
A
GGA TCC GCA ATG GGA AAA ATG GCT TCT CTA TTT GCC ACA TTT TTA GTG GTT TTA GTG TCA CTT AGC TTA GCA CAA ACA ACC
G S A M G K M A S L F A T F L V V L V S L S L A Q T T
XmaI
YK
NcoI
EGFP→
CGG GGA AGA CGA CCA TCA CCA CCA CCA CCT TAC TAC TAC AAG TCT CCT CCT CCC CCA TCA CCA TCA ACC ATG GTG AGC AAG...
R G R R P S P P P P Y Y Y K S P P P P S P S T M V S K
8 & 20
BamHI
SIGNAL SEQUENCE
B
GGA TCC GCA ATG GGA AAA ATG GCT TCT CTA TTT GCC ACA TTT TTA GTG GTT TTA GTG TCA CTT AGC TTA GCA CAA ACA ACC
G S A M G K M A S L F A T F L V V L V S L S L A Q T T
XmaI
YL
NcoI
EGFP→
CGG GGA AGA CGA CCA TCA CCA CCA CCA CCT TAC TAC TAC CTA TCT CCT CCT CCC CCA TCA CCA TCA ACC ATG GTG AGC AAG...
R G R R P S P P P P Y Y Y L S P P P P S P S T M V S K
8 & 20
BamHI
SIGNAL SEQUENCE
C
GGA TCC GCA ATG GGA AAA ATG GCT TCT CTA TTT GCC ACA TTT TTA GTG GTT TTA GTG TCA CTT AGC TTA GCA CAA ACA ACC
G S A M G K M A S L F A T F L V V L V S L S L A Q T T
XmaI
FK
NcoI
EGFP→
CGG GCC TCA CCA CCA CCA CCT AGT CCG TCA CCC CCT CCC CCA TTT TTC TTT AAG AGC CCA CCA CCA CCT TCC ATG GTG AGC AAG
R A S P P P P P S P S P P P P F F F K S P P P P S M V S K
9
BamHI
SIGNAL SEQUENCE
D
GGA TCC GCA ATG GGA AAA ATG GCT TCT CTA TTT GCC ACA TTT TTA GTG GTT TTA GTG TCA CTT AGC TTA GCA CAA ACA ACC
G S A M G K M A S L F A T F L V V L V S L S L A Q T T
XmaI
FL
NcoI
EGFP→
CGG GGA AGA CGA CCA TCA CCA CCA CCA CCT TTT TTC TTT CTA TCT CCT CCT CCC CCA TCA CCA TCA ACC ATG GTG AGC AAG...
R G R R P S P P P P F F F L S P P P P S P S T M V S K
8

Figure 13

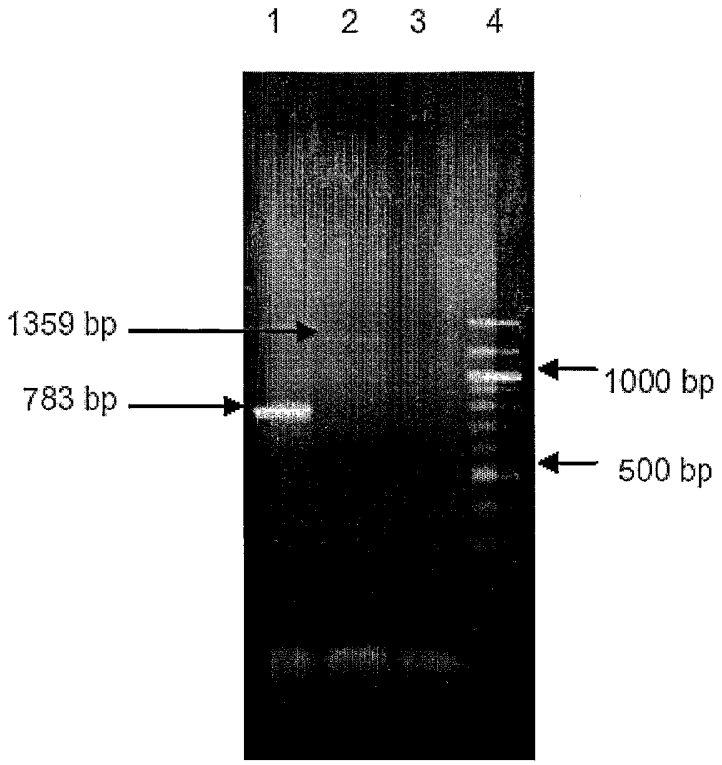


Figure 14

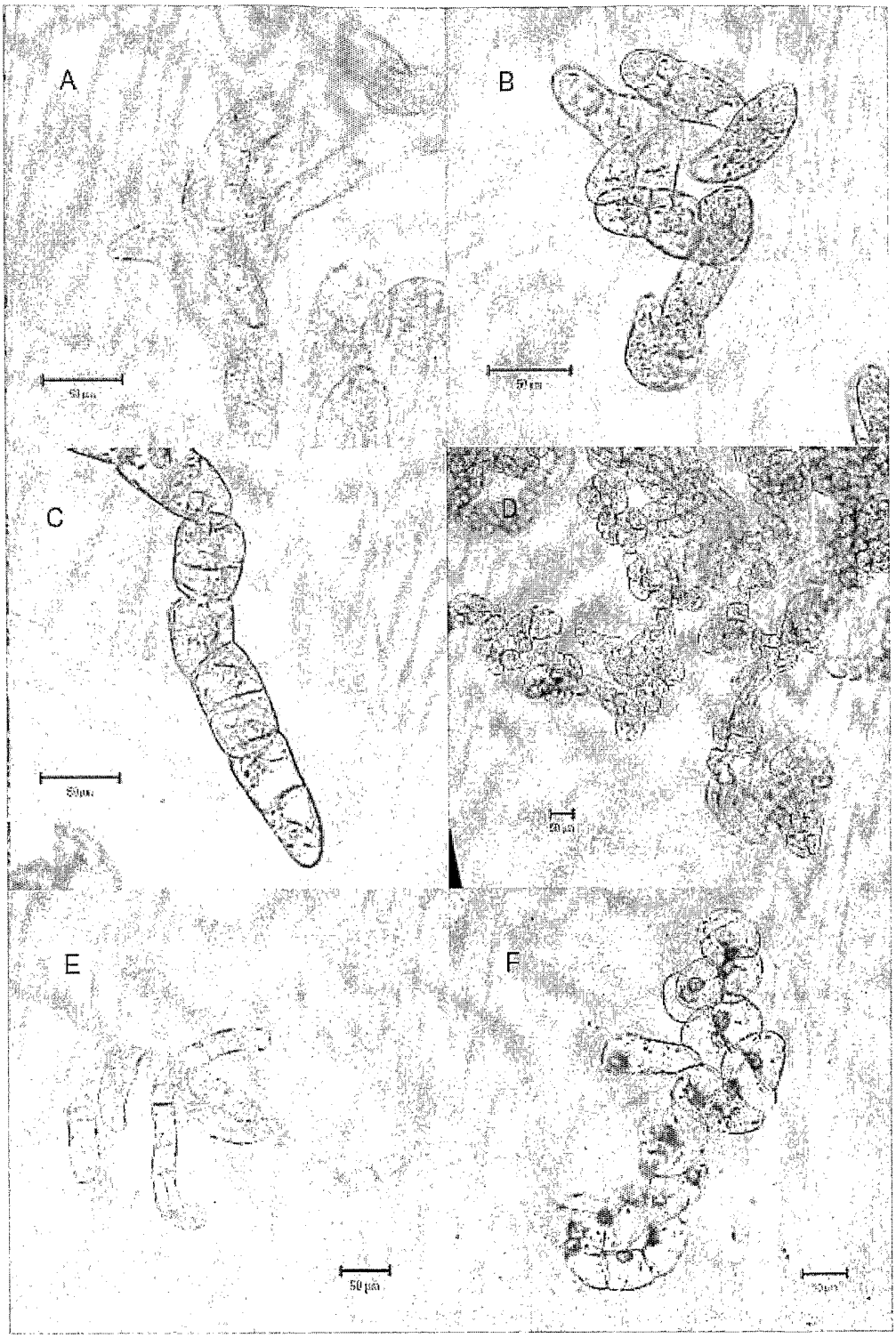


Figure 15

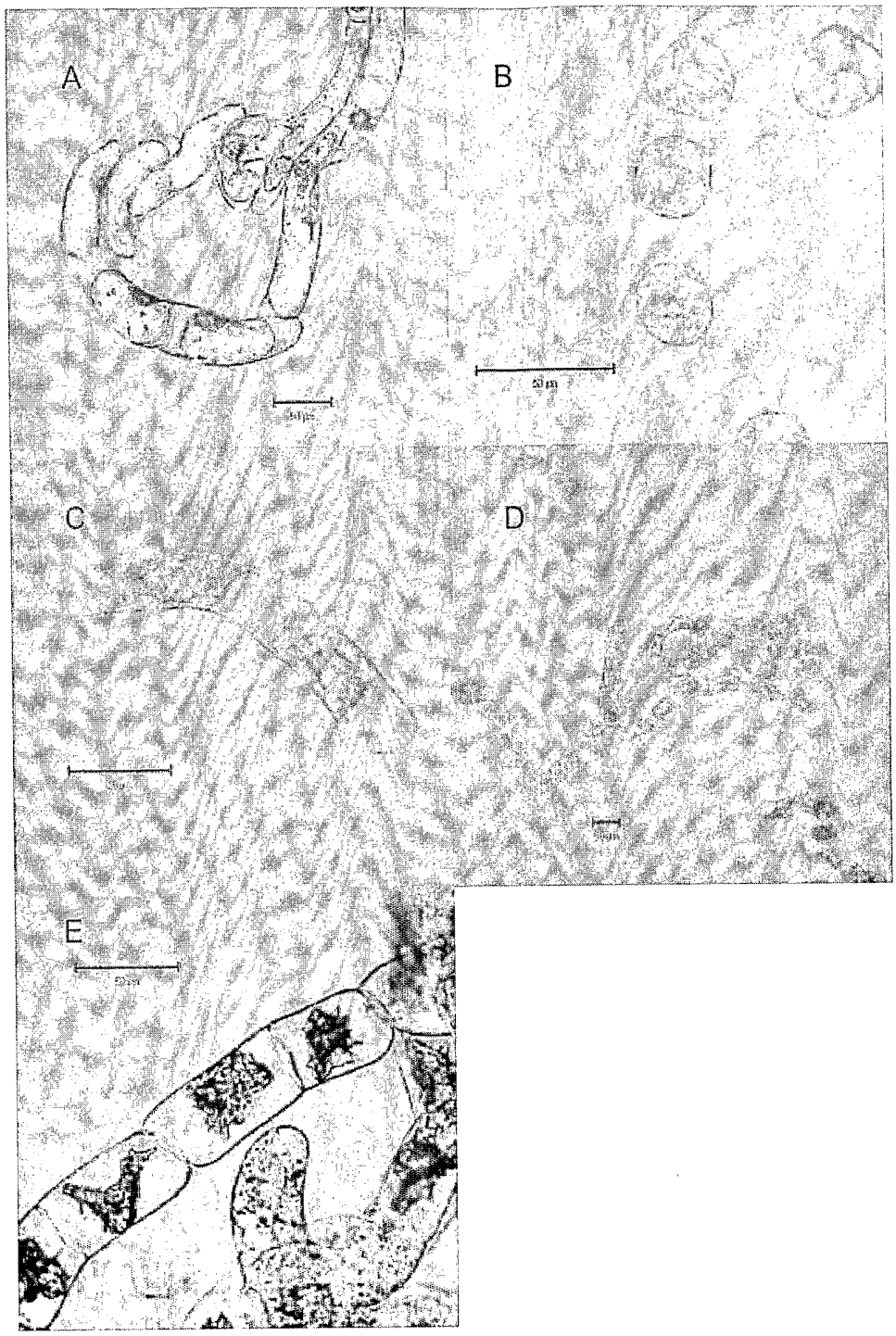


Figure 16

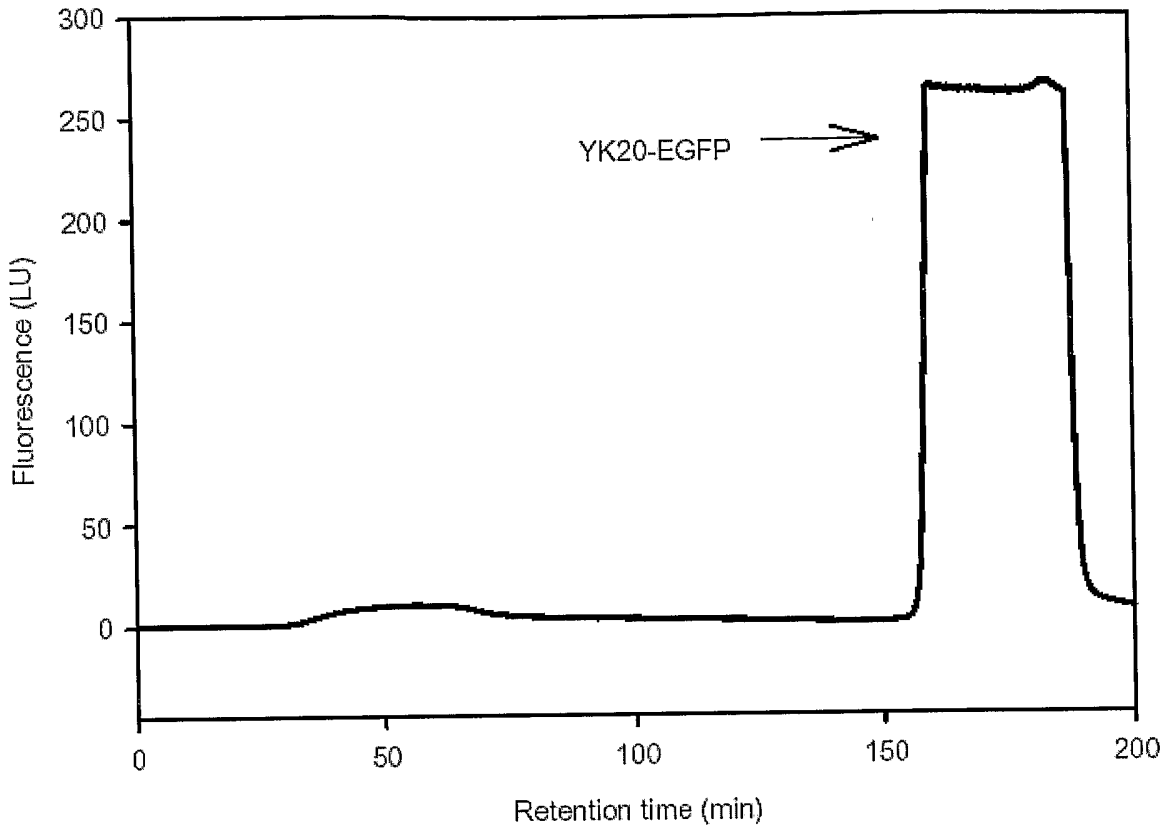

300
250
200
150
100
50
0
Fluorescence (LU)
YK20-EGFP
0
50
100
150
200
Retention time (min)

Figure 17

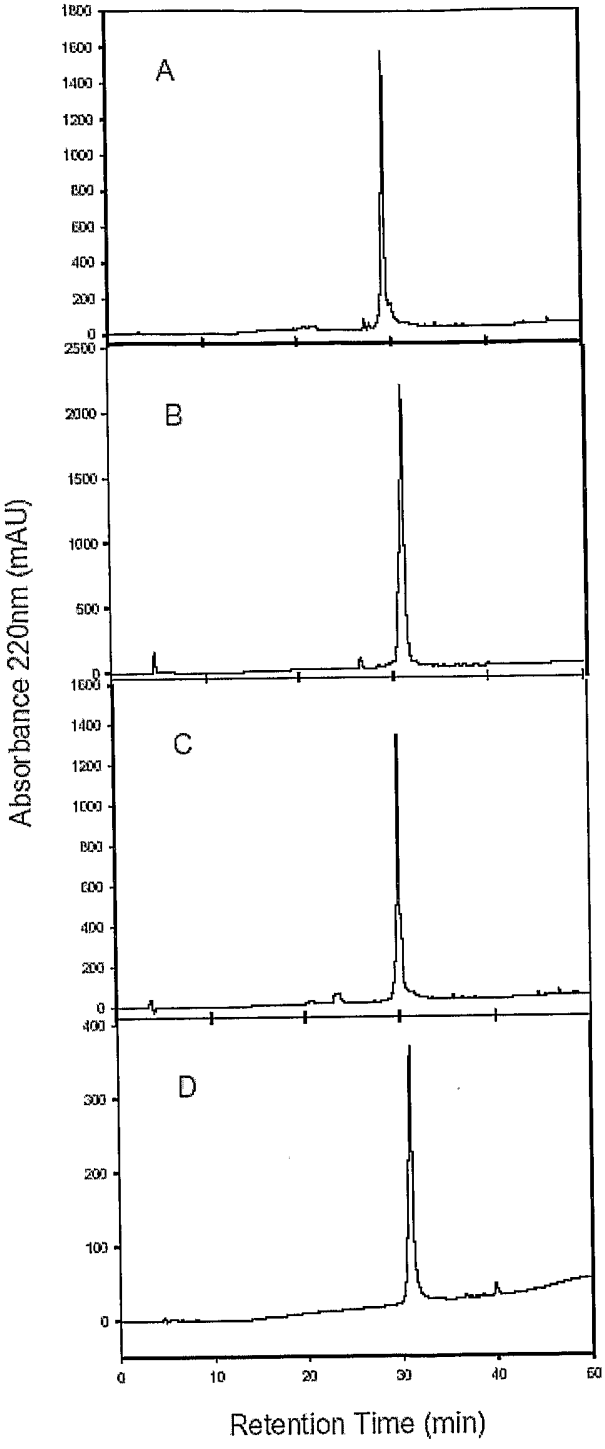


Figure 18

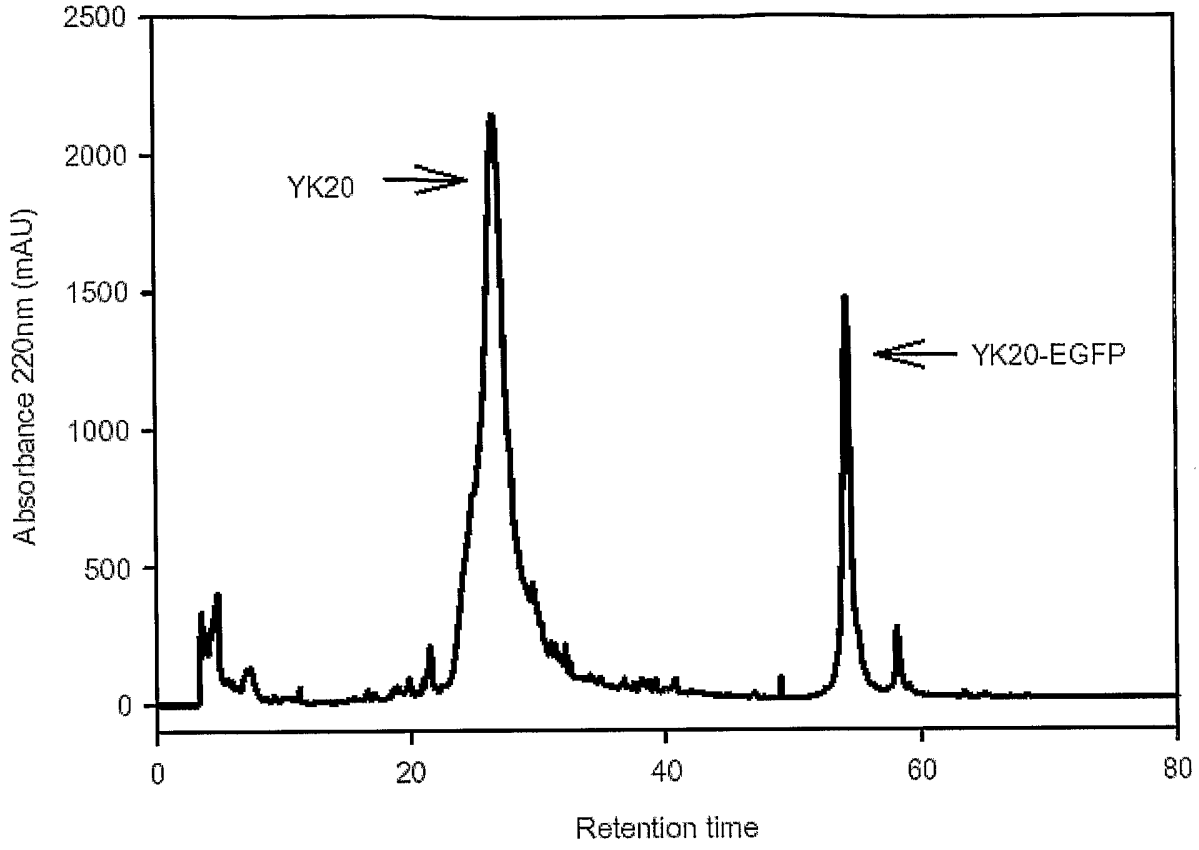

2500
2000
1500
1000
500
0
Absorbance 220nm (mAU)
YK20
YK20-EGFP
0
20
40
60
80
Retention time

Figure 19

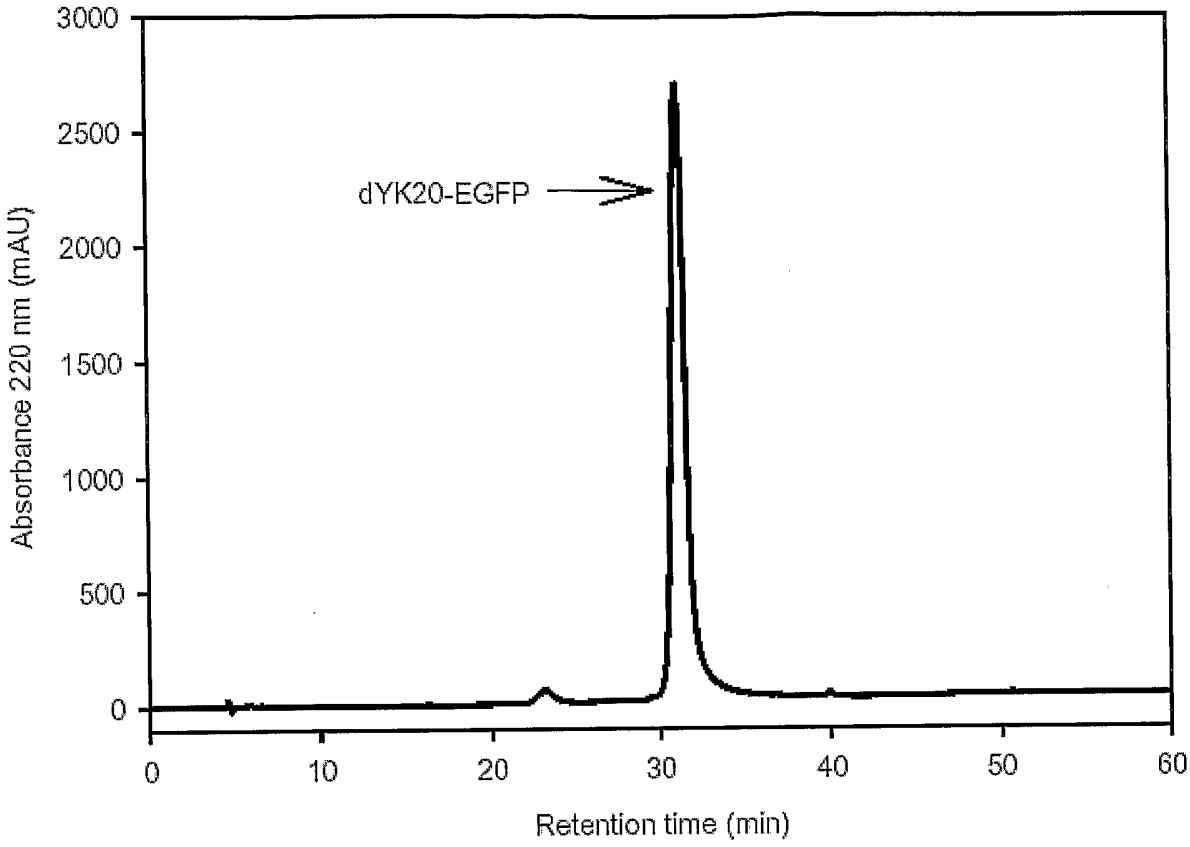

3000
2500
2000
1500
1000
500
0
Absorbance 220 nm (mAU)
dYK20-EGFP
0
10
20
30
40
50
60
Retention time (min)

Figure 20

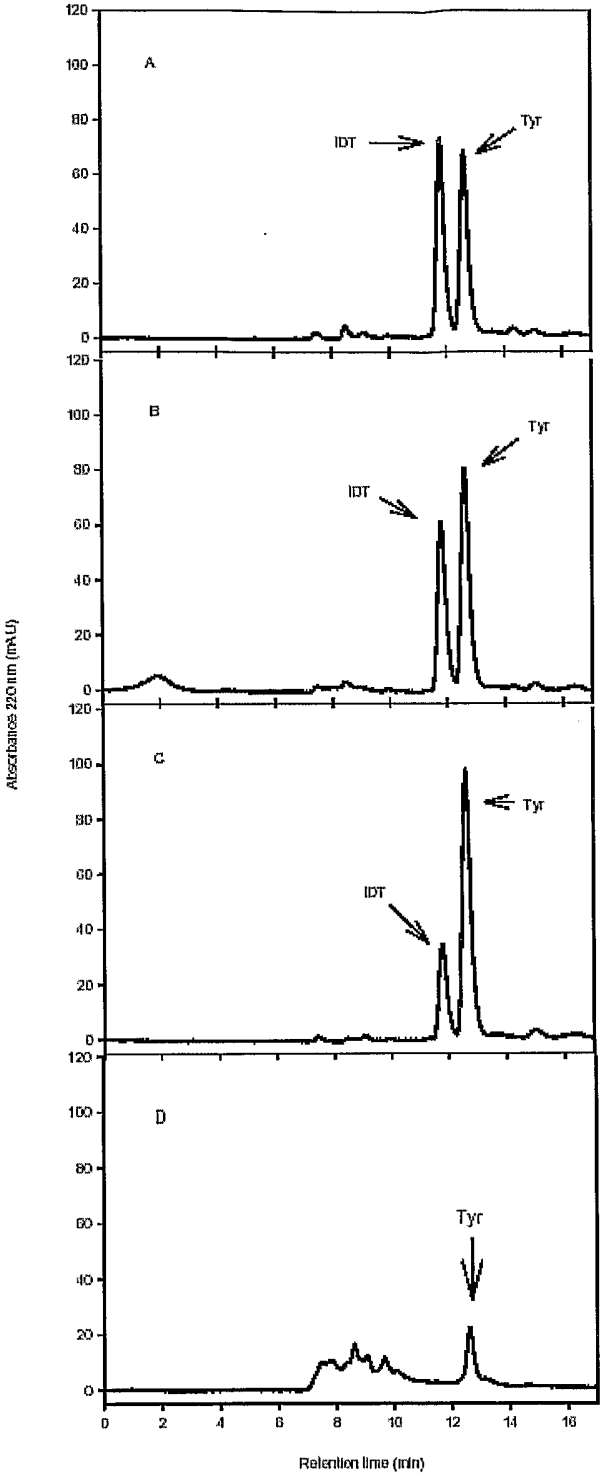


Figure 21

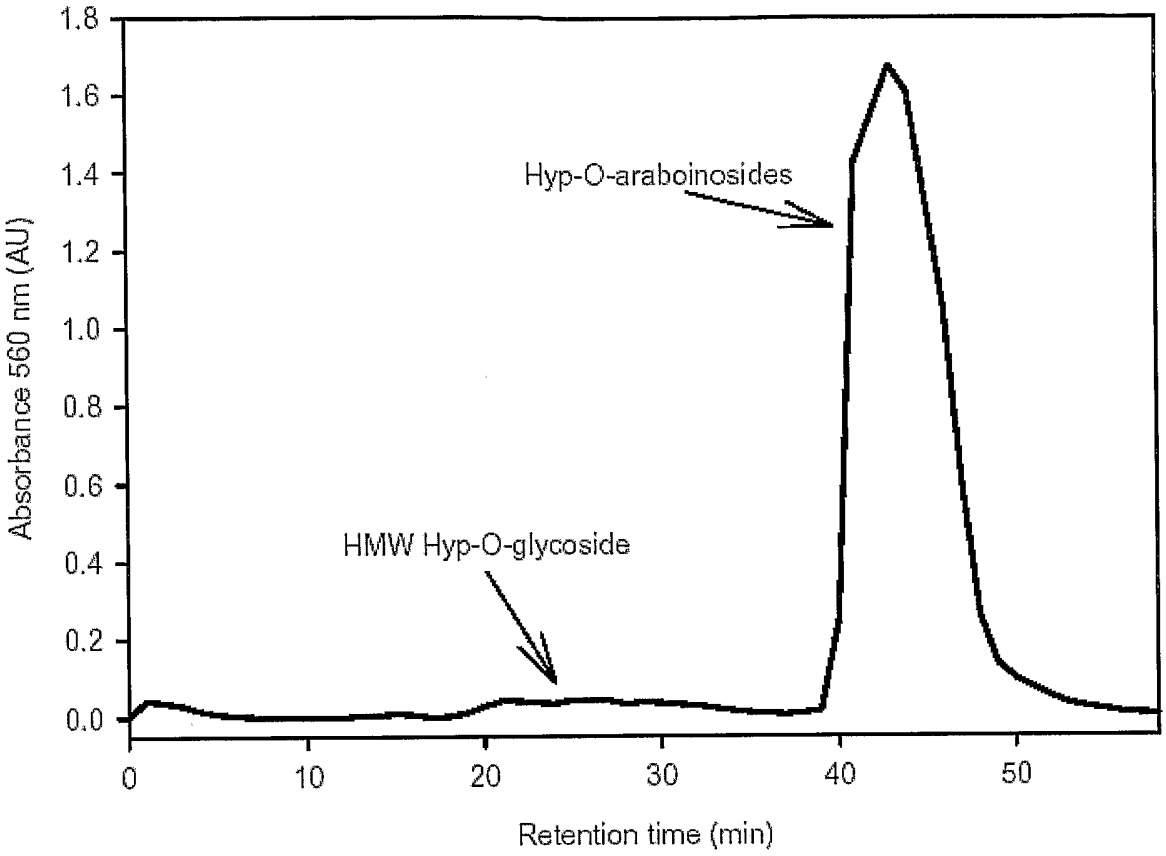

Absorbance 560 nm (AU)
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Hyp-O-araboinosides
HMW Hyp-O-glycoside
0
10
20
30
40
50
Retention time (min)

Figure 22

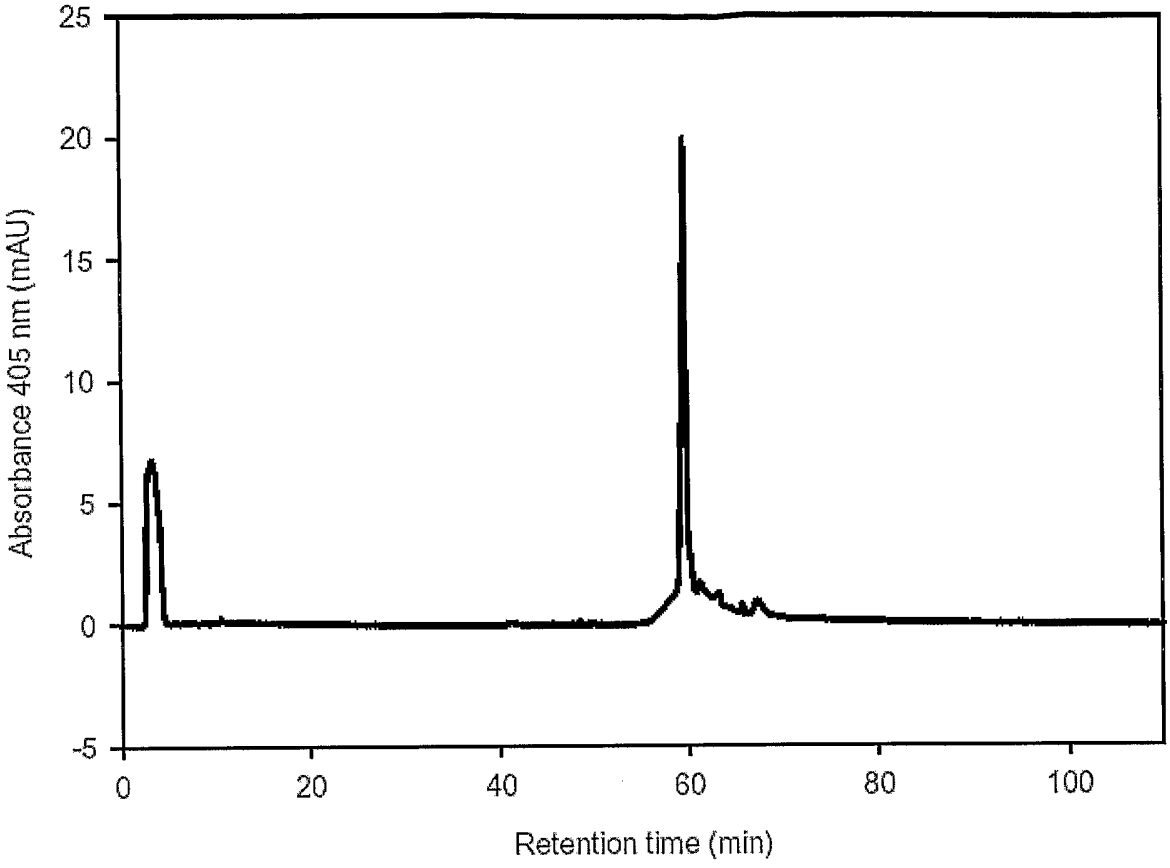

25
20
15
10
5
0
-5
Absorbance 405 nm (mAU)
0
20
40
60
80
100
Retention time (min)

Figure 23

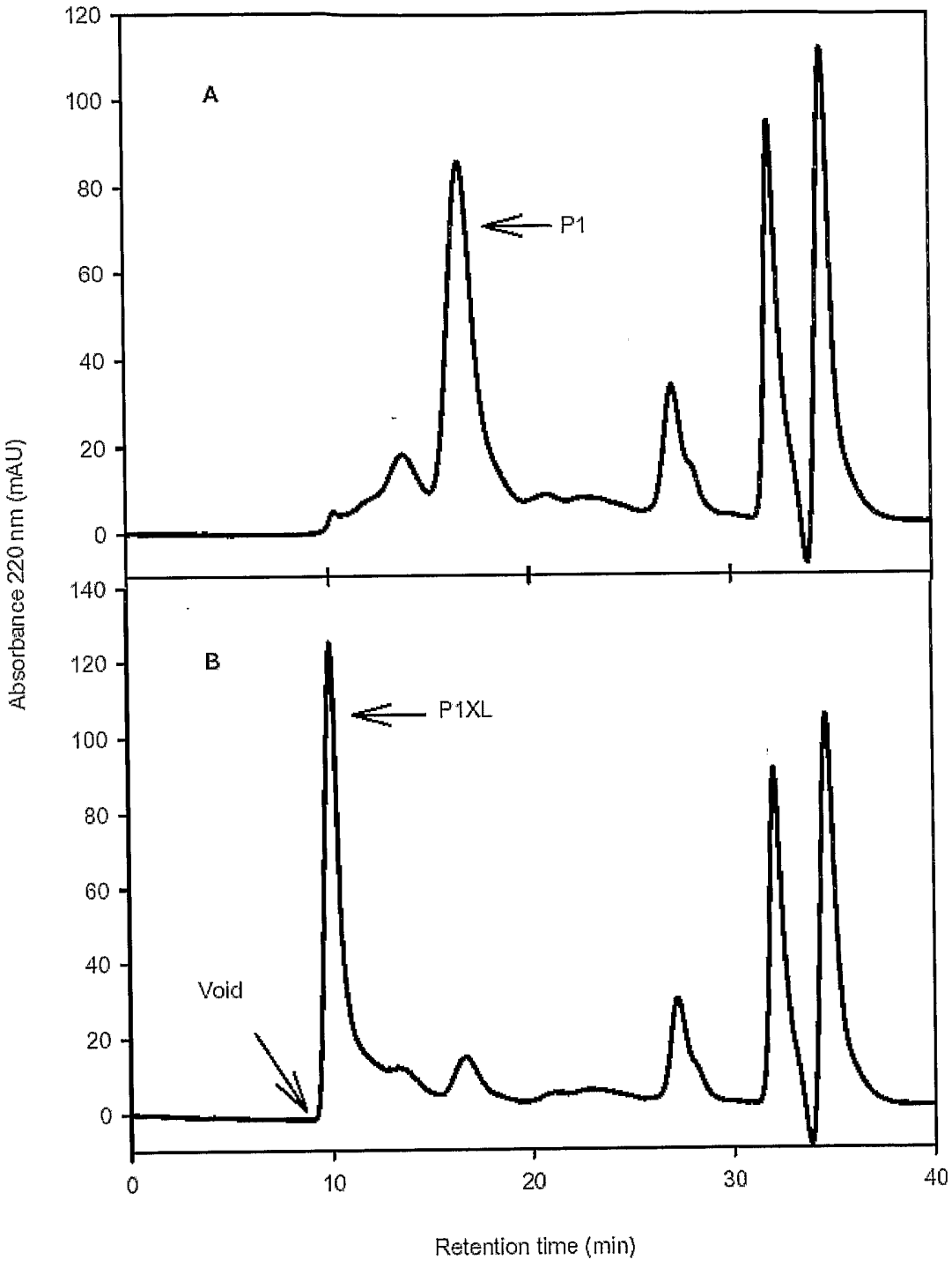

A
P1
B
P1XL
Void
Absorbance 220 nm (mAU)
Retention time (min)
0
10
20
30
40
0
20
40
60
80
100
120
140

Figure 24

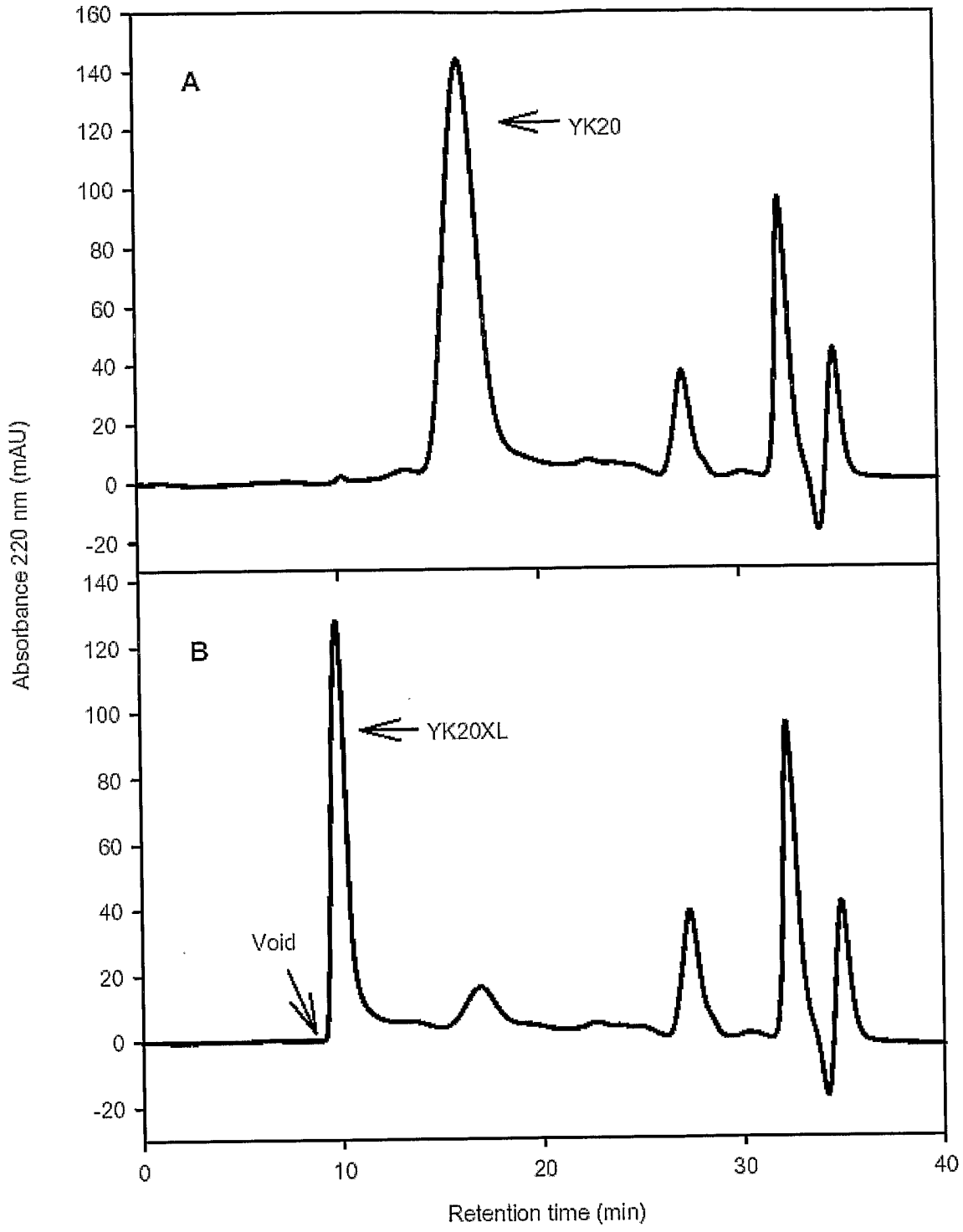

A
YK20
B
YK20XL
Void
Absorbance 220 nm (mAU)
Retention time (min)
160
140
120
100
80
60
40
20
0
-20
0
10
20
30
40

Figure 25

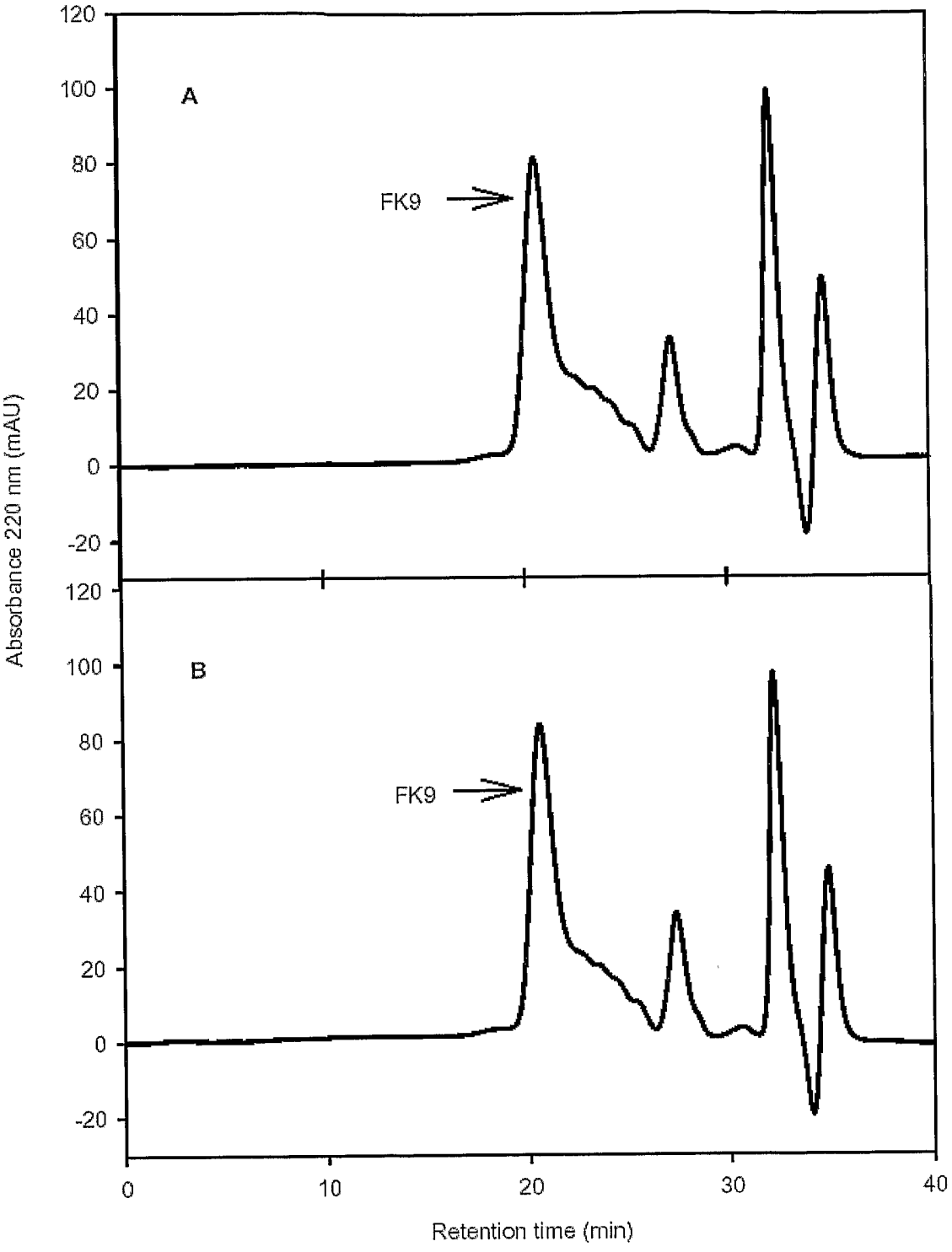

A
FK9
B
FK9
Absorbance 220 nm (mAU)
Retention time (min)
120
100
80
60
40
20
0
-20
0
10
20
30
40

Figure 26

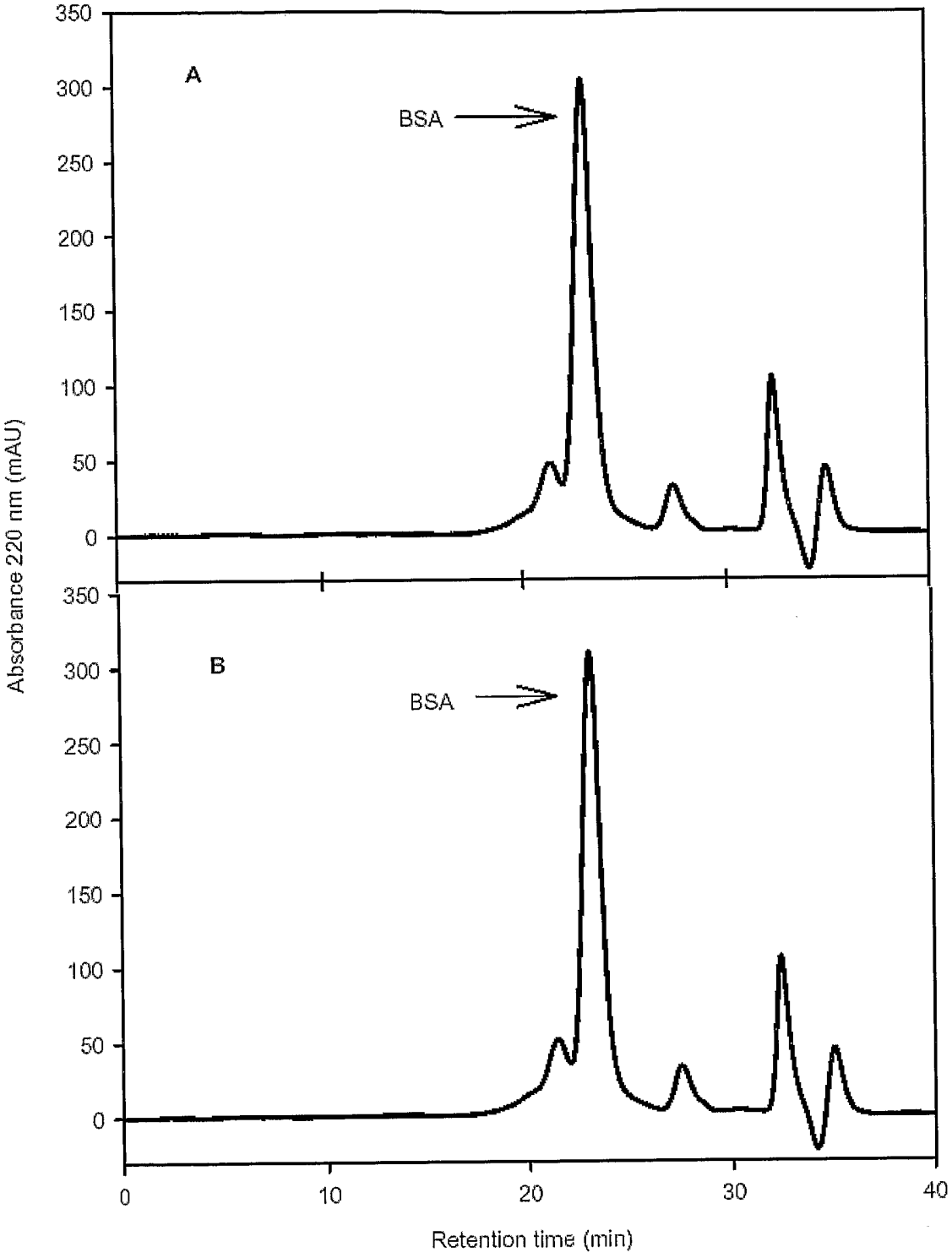

A
BSA
B
BSA
Absorbance 220 nm (mAU)
350
300
250
200
150
100
50
0
0
10
20
30
40
Retention time (min)

Figure 27

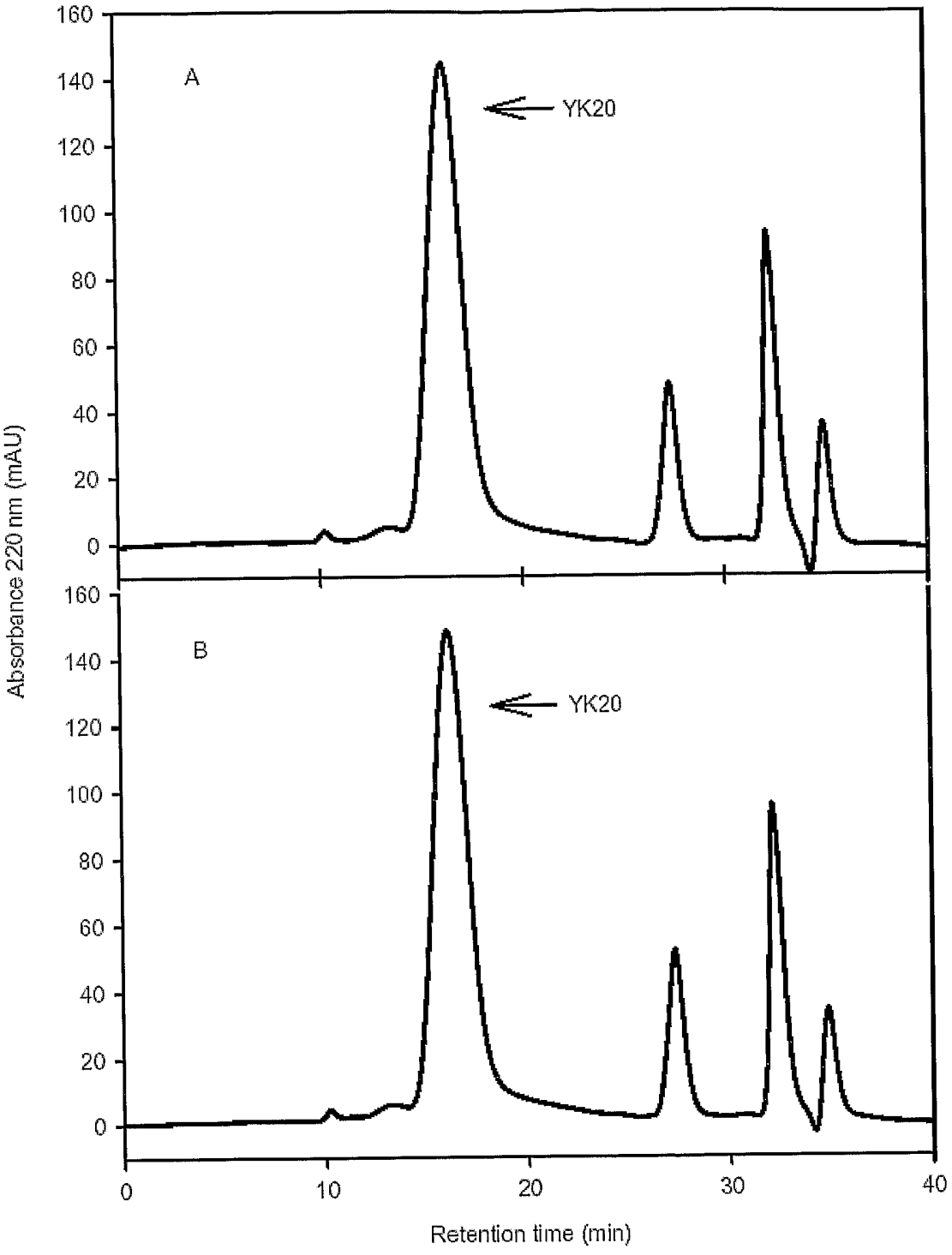

A
YK20
B
YK20
Absorbance 220 nm (mAU)
Retention time (min)
0
10
20
30
40
0
20
40
60
80
100
120
140
160

Figure 28

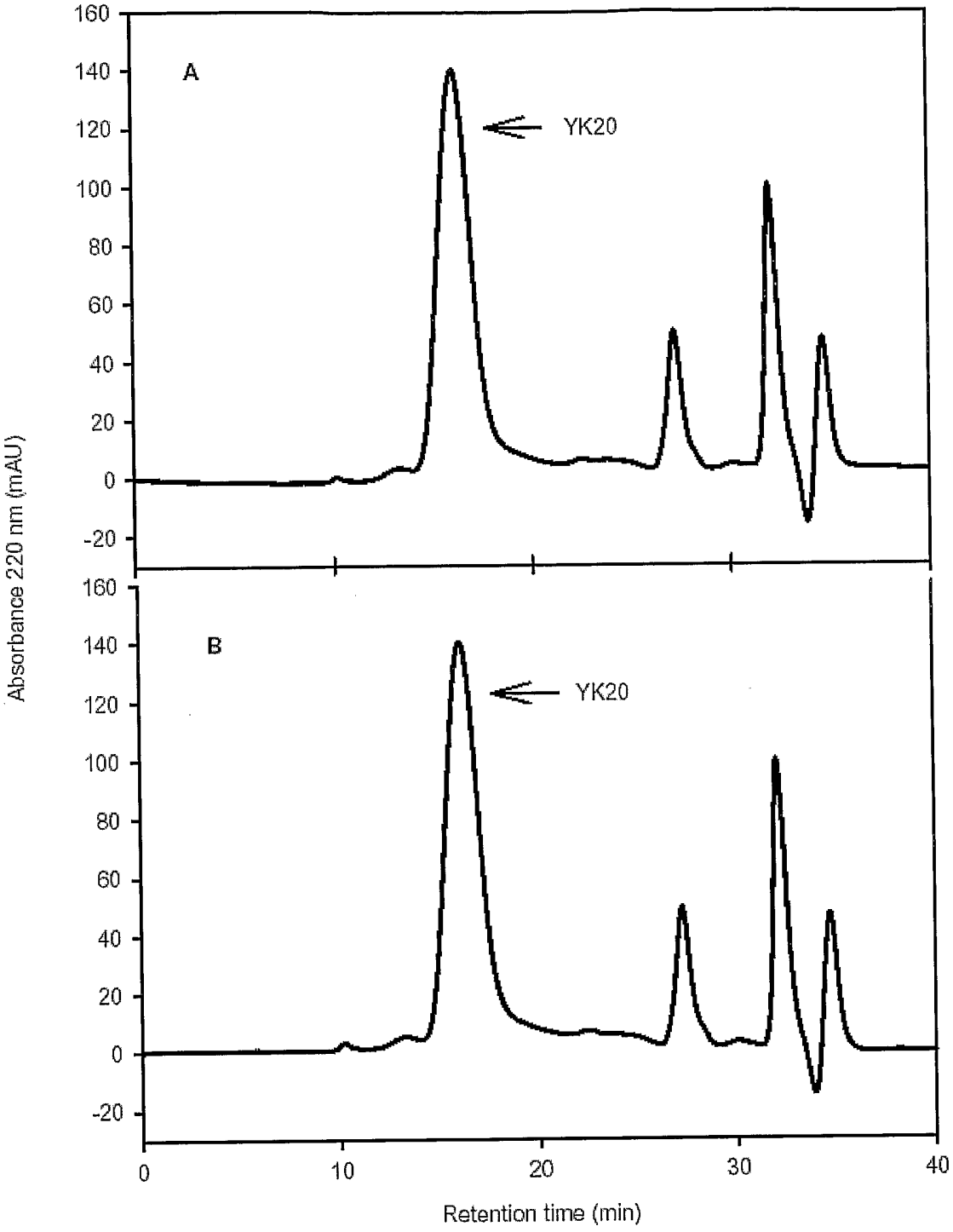

A
YK20
B
YK20
Absorbance 220 nm (mAU)
Retention time (min)
160
140
120
100
80
60
40
20
0
-20
0
10
20
30
40

Figure 29

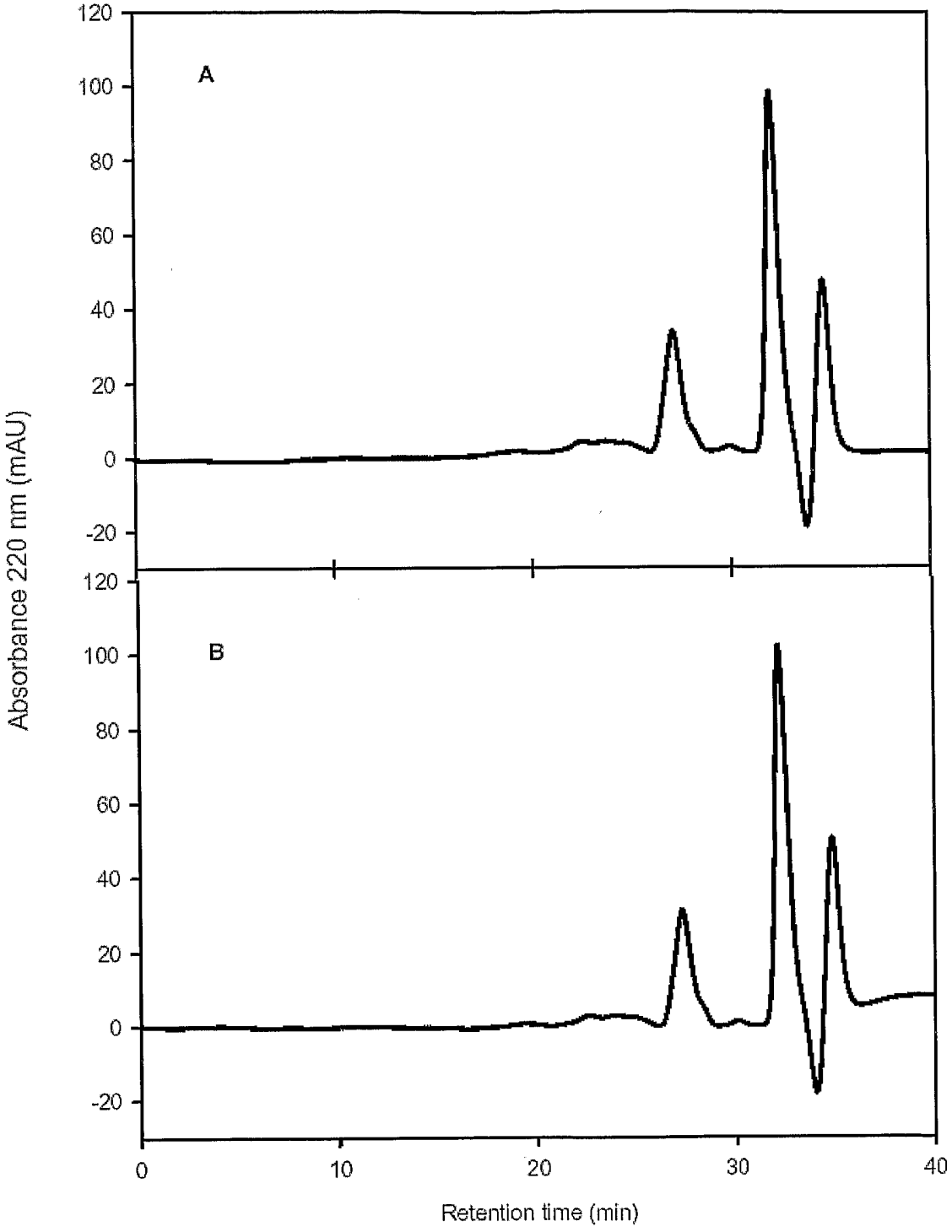

A
B
120
100
80
60
40
20
0
-20
Absorbance 220 nm (mAU)
0
10
20
30
40
Retention time (min)

Figure 30

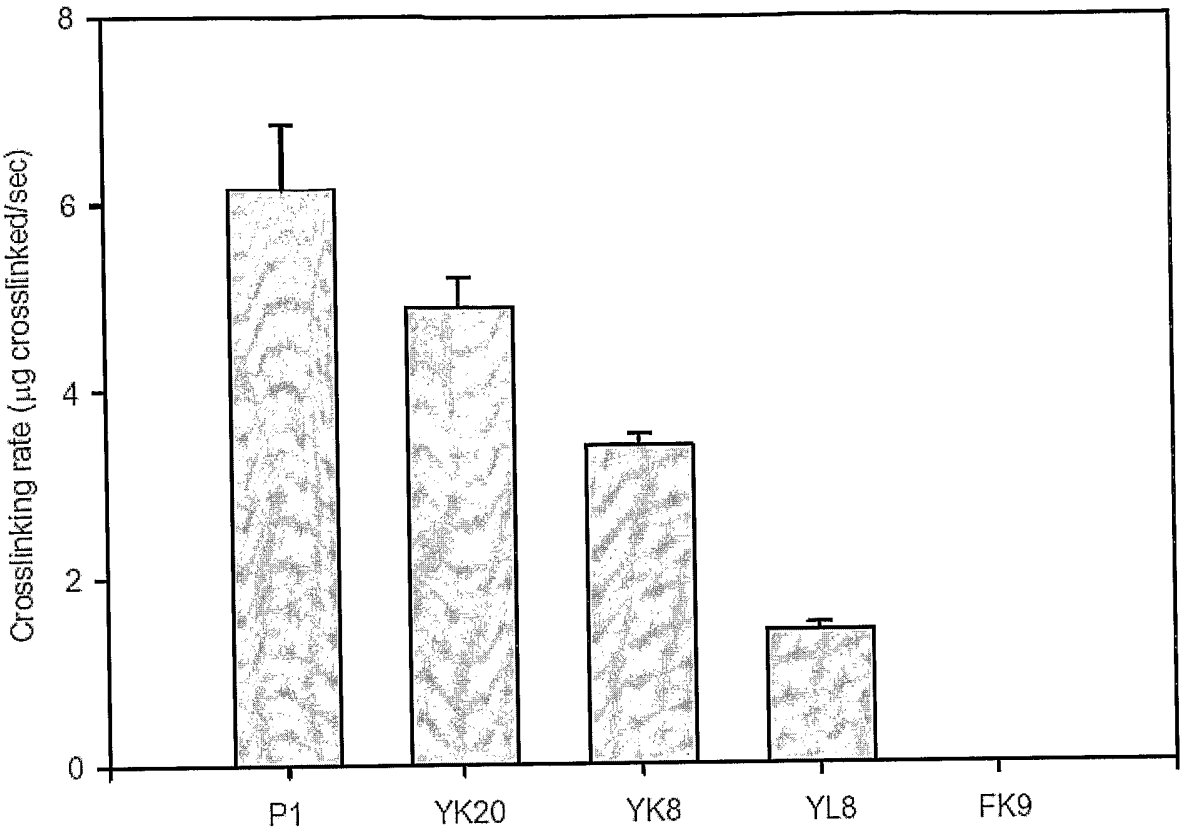

Crosslinking rate (µg crosslinked/sec)
8
6
4
2
0
P1
YK20
YK8
YL8
FK9

Figure 31

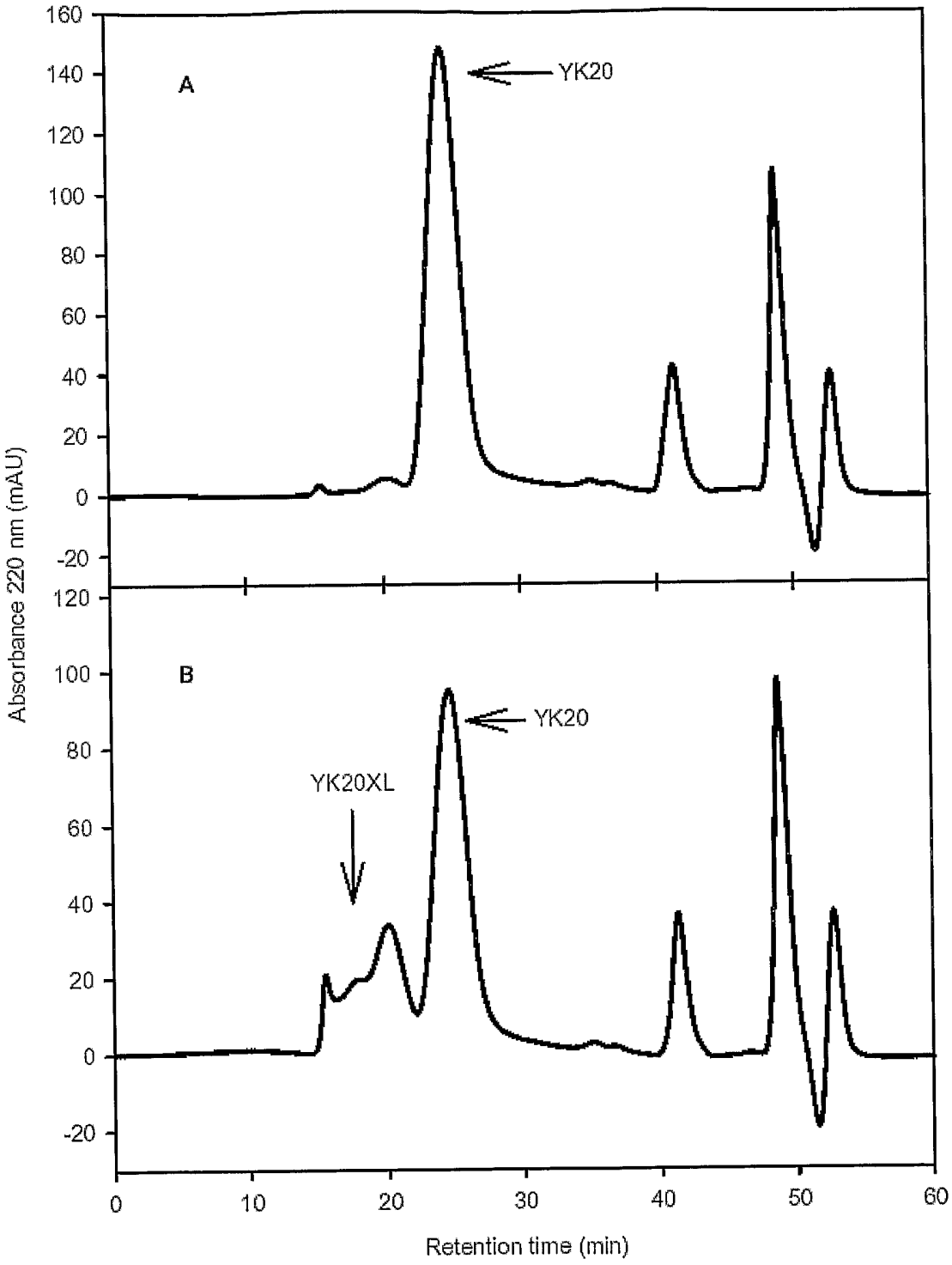

A
YK20
B
YK20
YK20XL
Absorbance 220 nm (mAU)
Retention time (min)
160
140
120
100
80
60
40
20
0
-20
120
100
80
60
40
20
0
-20
0
10
20
30
40
50
60

Figure 32

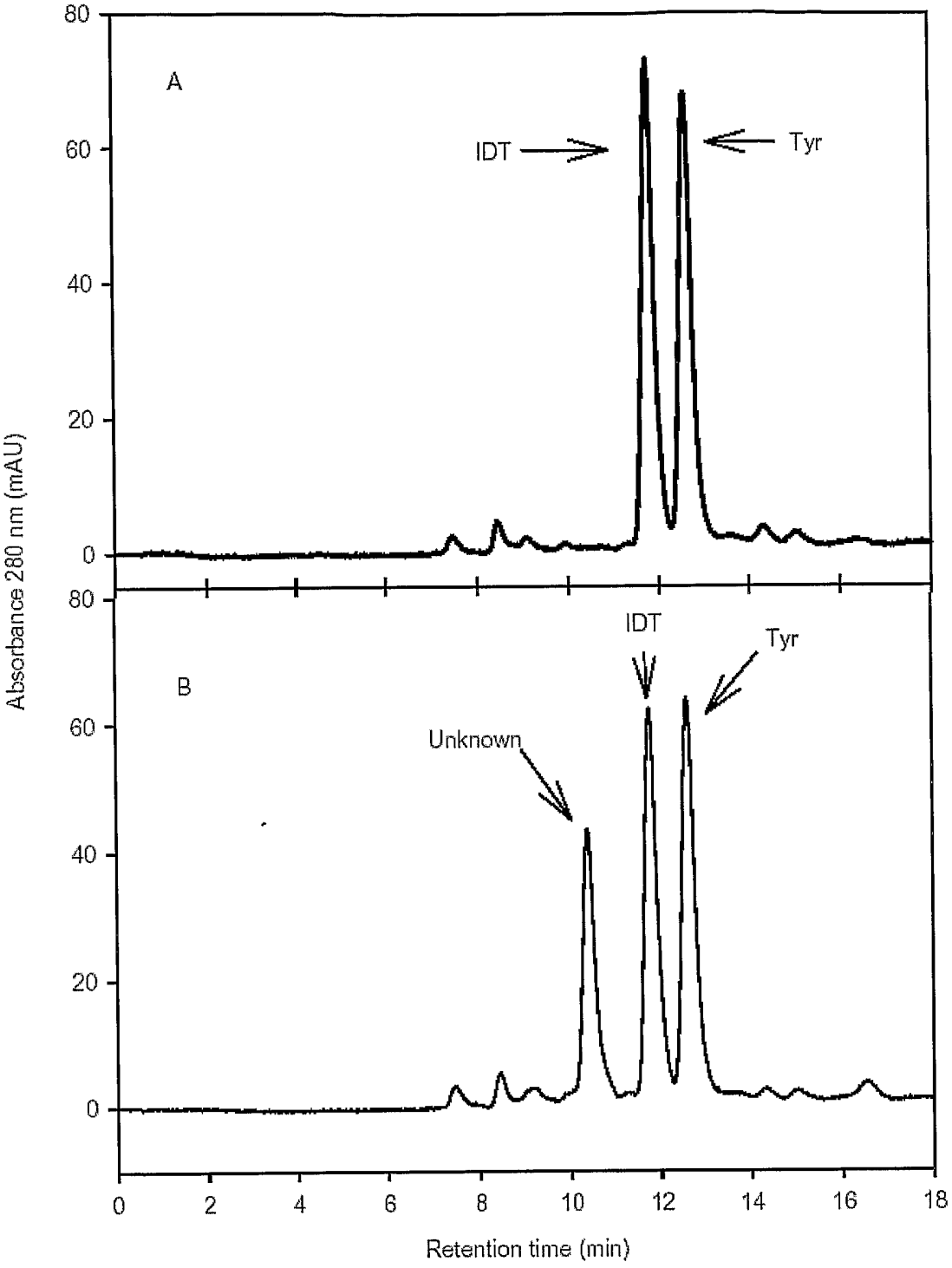

A
IDT
Tyr
B
IDT
Tyr
Unknown
Absorbance 280 nm (mAU)
Retention time (min)
0
20
40
60
80
0
2
4
6
8
10
12
14
16
18

Figure 33

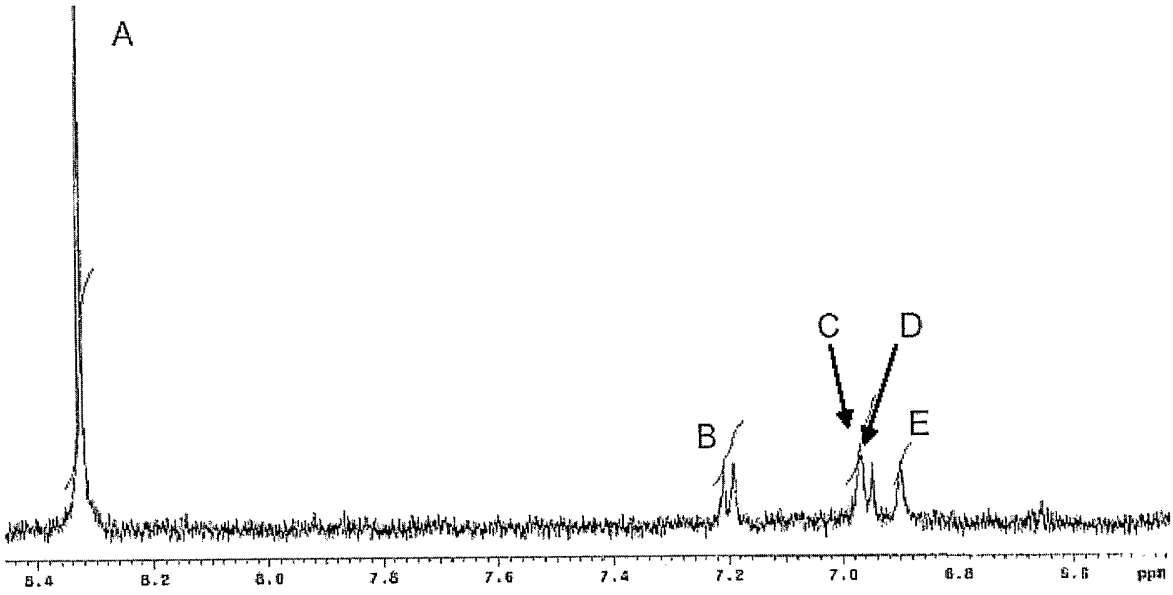


Figure 34

ChemNMR Prediction Result:

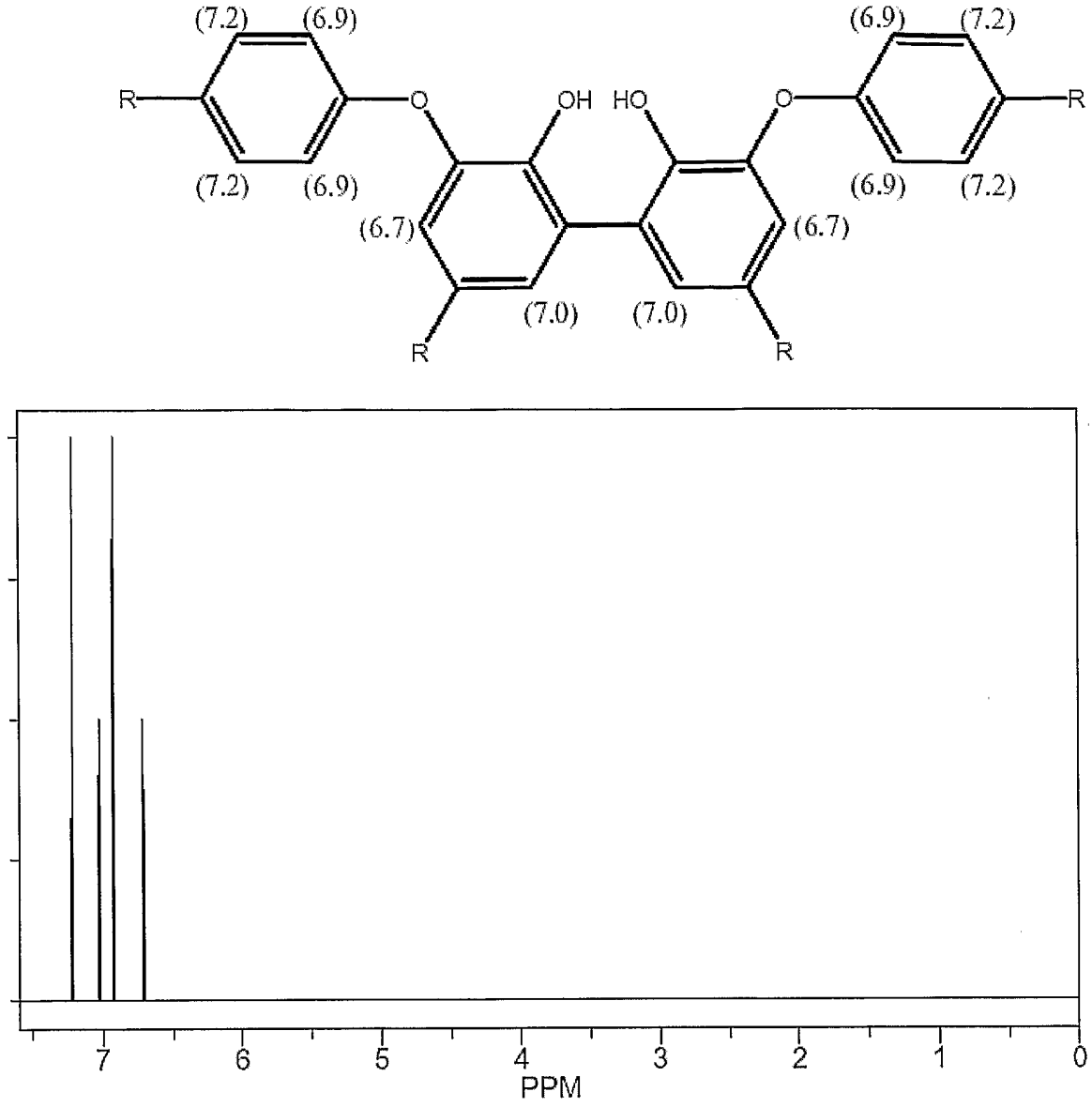


Figure 35

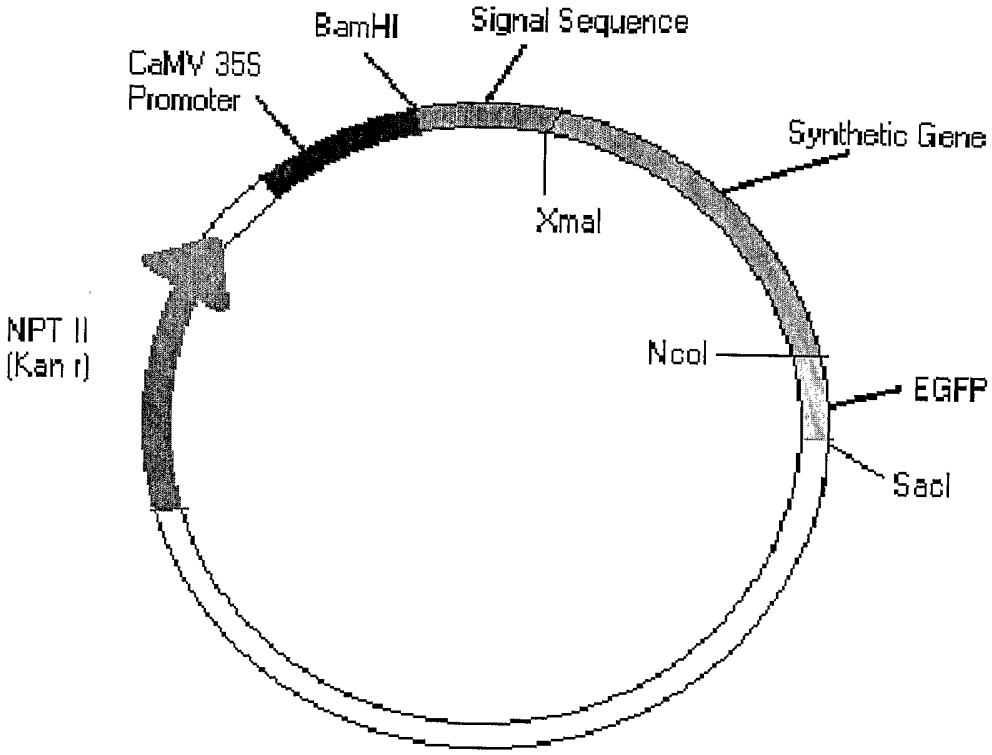

BamHI
Signal Sequence
CaMV 35S
Promoter
Synthetic Gene
XmaI
NPT II
(Kan r)
NcoI
EGFP
SacI

Figure 36

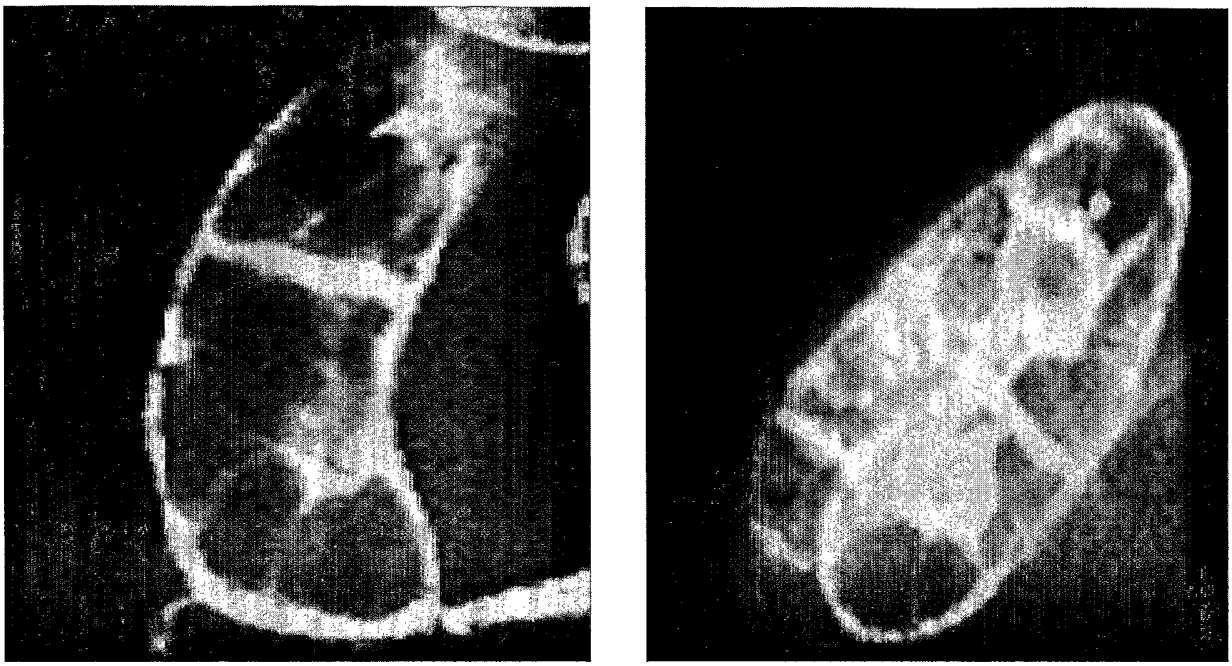


Figure 37

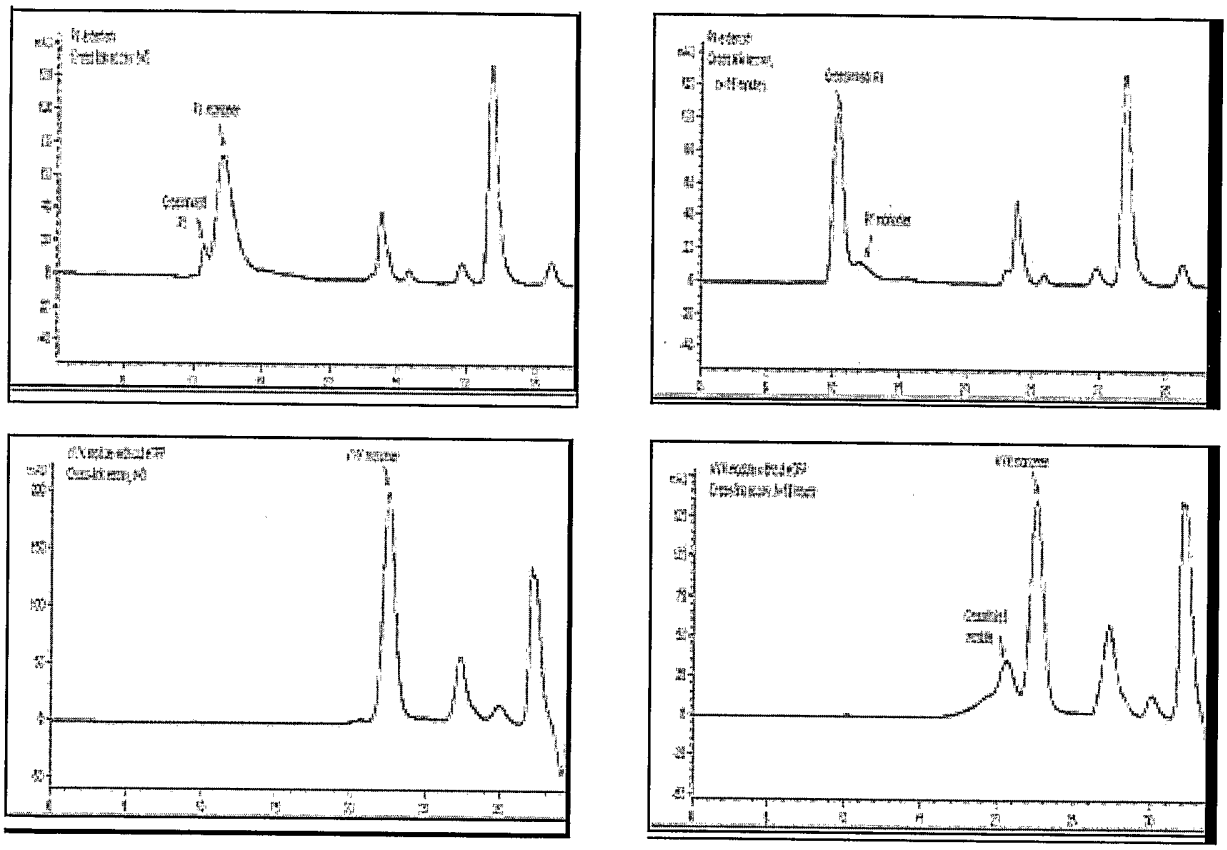


Figure 38

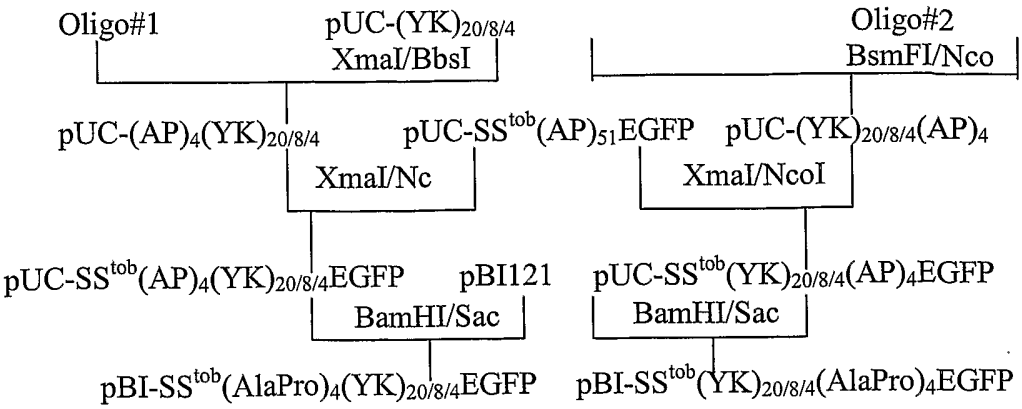

Oligo#1
pUC-(YK)20/8/4
XmaI/BbsI
Oligo#2
BsmFI/Nco
pUC-(AP)4(YK)20/8/4
pUC-SStob(AP)51EGFP
pUC-(YK)20/8/4(AP)4
XmaI/Nc
XmaI/NcoI
pUC-SStob(AP)4(YK)20/8/4EGFP
pBI121
pUC-SStob(YK)20/8/4(AP)4EGFP
BamHI/Sac
BamHI/Sac
pBI-SStob(AlaPro)4(YK)20/8/4EGFP
pBI-SStob(YK)20/8/4(AlaPro)4EGFP

Figure 39

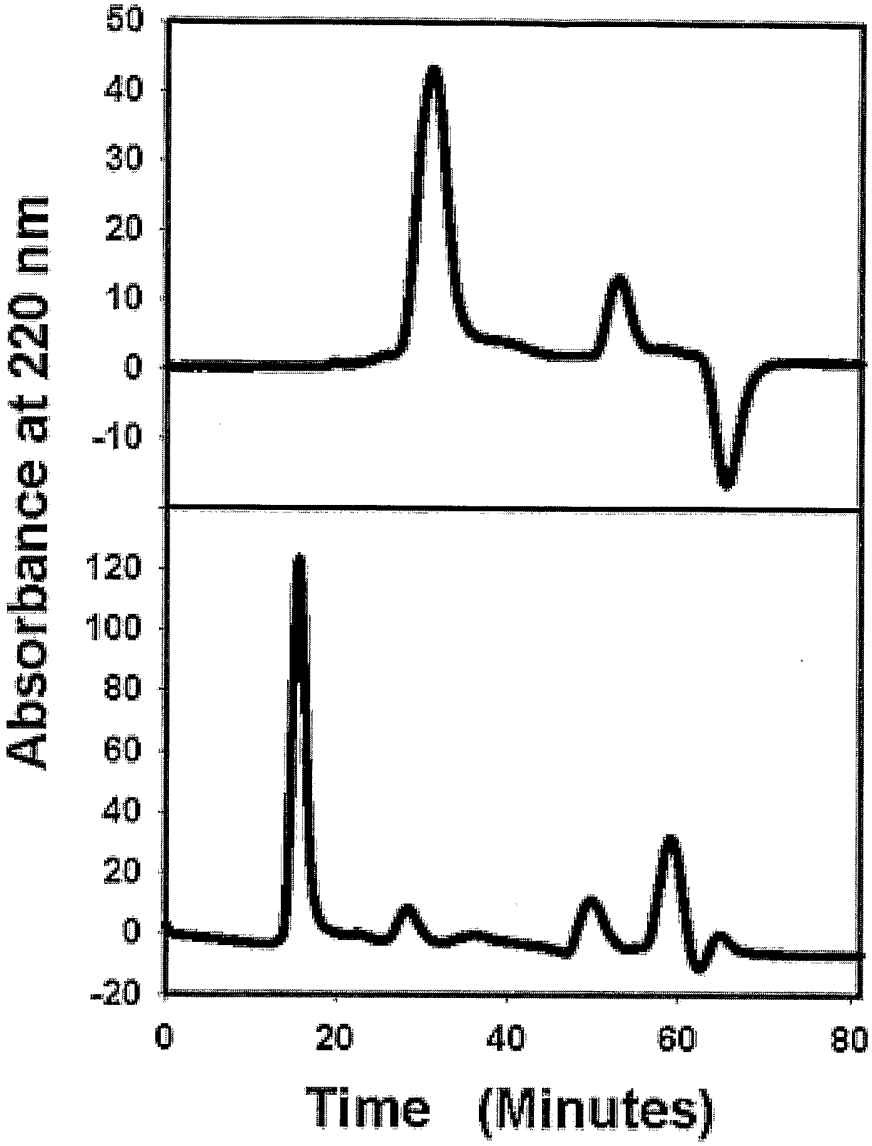

Absorbance at 220 nm
50
40
30
20
10
0
-10
120
100
80
60
40
20
0
-20
0
20
40
60
80
Time (Minutes)

Figure 40

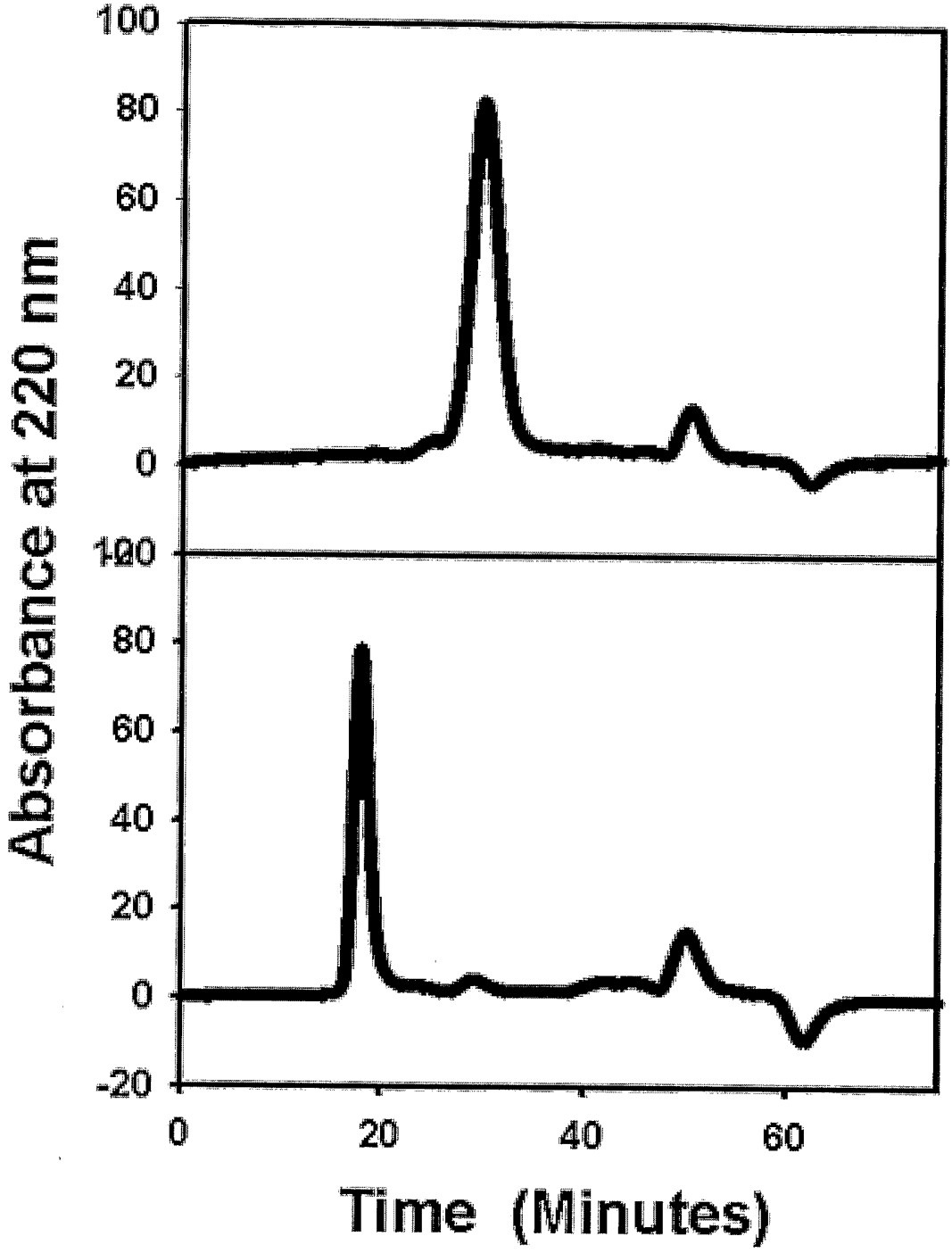


Figure 41

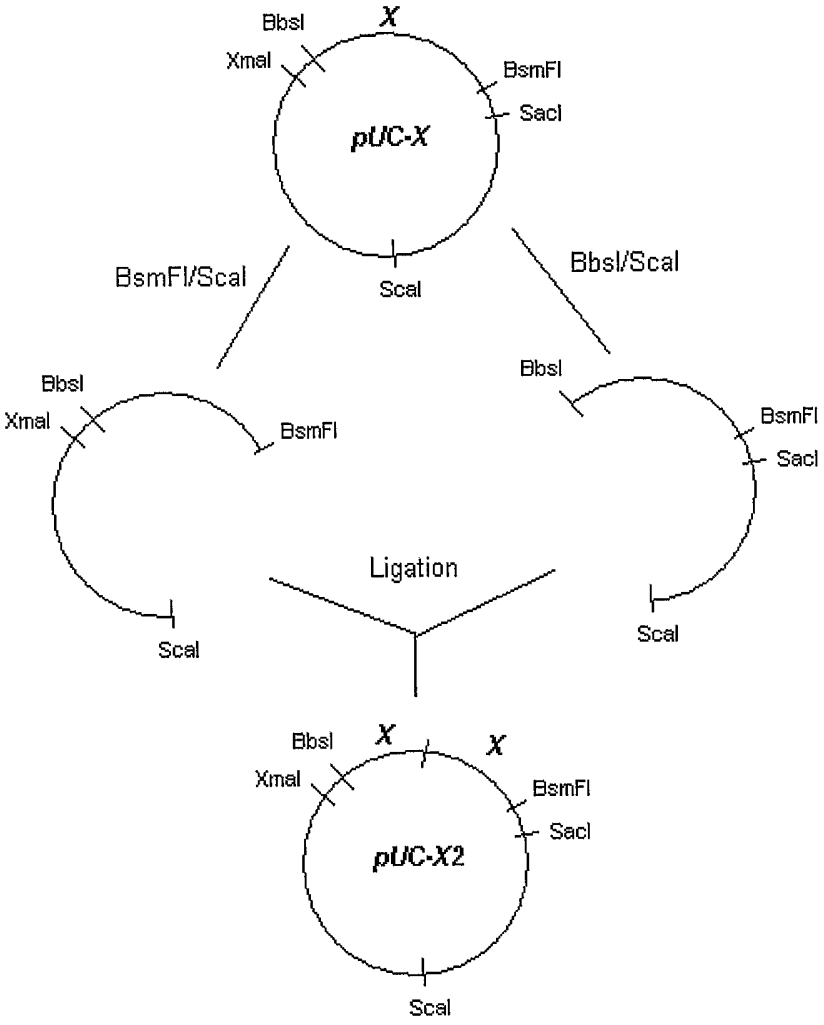

BbsI
X
XmaI
BsmFI
SacI
pUC-X
ScaI
BsmFI/ScaI
BbsI/ScaI
BbsI
XmaI
BsmFI
ScaI
BbsI
BsmFI
SacI
ScaI
Ligation
BbsI
X
X
XmaI
BsmFI
SacI
pUC-X2
ScaI

Figure 42