Identification of *N*-acetylglucosamine binding residues in *Griffonia simplicifolia* lectin II

Keyan Zhu[a], Ray A. Bressan[b], Paul M. Hasegawa[b], Larry L. Murdock[a,*]

[a]*Department of Entomology, Purdue University, West Lafayette, IN 47907, USA*
[b]*Department of Horticulture, Purdue University, West Lafayette, IN 47907, USA*

Received 24 April 1996; revised version received 13 June 1996

Abstract Primary structure and crystallographic data of several legume lectins were used to predict the involvement in carbohydrate binding of six amino acid residues (Asp^{88}, Glu^{108}, Tyr^{134}, Asn^{136}, Leu^{226} and Gln^{227}) in *Griffonia simplicifolia* lectin II (GS-II). The functional involvement of these residues was evaluated by assessing GlcNAc binding of modified forms of GS-II in which these residues were eliminated in truncated peptides or systematically substituted with other amino acids by site-specific mutations. Mutations at Asp^{88}, Tyr^{134} or Asn^{136} eliminated GlcNAc binding activity by GS-II, while those at Glu^{108}, Leu^{226} or Gln^{227} did not alter the activity. The former three amino acids were functionally essential for carbohydrate binding by GS-II presumably through hydrogen bonding to and hydrophobic interactions with GlcNAc. Although an Asp or Gly substitution for Tyr^{134} eliminated GlcNAc affinity, substitution with Phe did not appreciably affect binding. Despite the fact that mutations to Leu^{226} and Gln^{227} did not alter carbohydrate binding, a truncated form of GS-II lacking these residues no longer exhibited carbohydrate binding affinity.

Key words: Lectin; *N*-acetylglucosamine; Carbohydrate binding residue; Site-directed mutagenesis; *Griffonia simplicifolia*

1. Introduction

Lectins are carbohydrate binding proteins that recognize specific sugar moieties in polysaccharides, glycoproteins and glycolipids [1]. These proteins are ubiquitously distributed among plants, animals and microbes. Presently, plant lectins can be categorized into two groups: those that have carbohydrate binding domains and those that have the functional sugar binding residues distributed spatially throughout the primary sequence of the protein [2]. Legume lectins are the largest known family of the latter group of proteins [3]. Most commonly, these lectins exist in native form either as homo- or heterodimers or tetramers. Each subunit exhibits carbohydrate binding properties and contains Mn^{2+} and Ca^{2+}, which are required for interaction with sugars [3]. Although there are differences in sugar substrate specifities, the legume lectins share extensive primary sequence homology and 3-dimensional structure similarity [3,4]. The tertiary structure is composed of two antiparallel pleated sheets of six and seven strands each that are connected by loops and β-turns [5]. The majority of plant lectins are secretory proteins that either accumulate in the vacuole or extracellular matrix [2].

Analyses of available legume lectin crystallographic data indicate that, regardless of carbohydrate specificity, the monosaccharide binding sites are very similar and are composed of six residues that interact with the sugars [5–11]. Included among the legume lectins that have been analyzed are concanavalin A (conA), *Lathyrus ochrus* isolectin 1 (LOL 1), *Griffonia simplicifolia* lectin IV (GS-IV), pea lectin, *Vicia faba* lectin (Favin), phytohemagglutinin (PHA), and *Erythrina corallodendron* lectin (ECorL). Arcelin and α-amylase inhibitor from common bean have sequence similarity to legume lectins but internal deletions have eliminated one and two of the domains, respectively, containing carbohydrate-interacting residues [11]. Consequently, arcelin has weak lectin activity while α-amylase inhibitor has none.

Legume lectins may function in plant defense against insects [12–15]. Proteins with specificities for *N*-acetylgalactosamine/galactose residues (GalNAc/Gal) and *N*-acetylglucosamine residues (GlcNAc) have been shown to exhibit insecticidal activities. Interaction of lectins with carbohydrate substrates in the gut may disrupt nutrient acquisition and other cellular functions and be the basis for anti-insect activity [12,13].

Recently, we determined [15] that leaf and seed isoforms of GlcNAc binding GS-II from *G. simplicifolia* have substantial insecticidal activities when included in cowpea weevil (*Callosobruchus maculatus* (F.)) bioassay diets. Leaf GS-II is composed of two unique subunits of M_r 22 and 30 kDa and the seed GS-II is a tetramer of a 30 kDa subunit. The leaf and seed 30 kDa subunits are immunologically related. Bacterial recombinant protein of the leaf large GS-II subunit exhibited both GlcNAc specificity and insecticidal activity [15].

We have initiated research to characterize the carbohydrate binding and insecticidal structure/function of GS-II, the first GlcNAc binding legume lectin proven to have insecticidal activity[15]. This report describes structure/function results pertaining to GlcNAc binding affinity obtained from experiments using truncated peptides or proteins containing site-specific mutations. Although the tertiary structure of GS-II is unknown, crystallographic, mutagenesis and primary sequence data of other legume lectins were used to identify six residues (Asp^{88}, Glu^{108}, Tyr^{134}, Asn^{136}, Leu^{226} and Gln^{227}) in GS-II that likely mediate monosaccharide binding [15]. Site-specific modifications to Asp^{88} and Asn^{136} in GS-II eliminated GlcNAc binding. Replacement of Tyr^{134} with Phe did not affect carbohydrate binding but elimination of the aromatic side chain by substitution with Asp or Gly abolished binding. Mutations to Glu^{108}, Leu^{226} and Gln^{227} did not perturb

*Corresponding author. Fax: (317) 496-1219.
E-mail: larry-murdock@entm.purdue.edu

Abbreviations: GlcNAc, *N*-acetylglucosamine; GS-II, *Griffonia simplicifolia* lectin II

This is publication No. 15 020, Purdue University Agricultural Experimental Station, West Lafayette, IN. This research was supported by a grant from National Research Initiative USDA Competitive Grant, Award No. 95-37302-1804.

0014-5793/96/$12.00 © 1996 Federation of European Biochemical Societies. All rights reserved.
PII S0014-5793(96)00671-0

GlcNAc binding but a truncated peptide that did not contain the latter two amino acids failed to bind to GlcNAc.

2. Materials and methods

2.1. Construction of truncated GS-II peptides

Procedures for PCR and bacterial expression of recombinant protein were as described previously [15]. DNAs encoding three truncated peptides of the large subunit of leaf GS-II were synthesized by PCR using the cDNA clone encoding the mature protein [15]. These DNAs encoded the following peptides: F1, Ala^{1} to Ser^{217}; F2, Thr^{117} to Met^{254}; and F3, Thr^{117} to Ser^{217}. The PCR primers were:

F1	sense 5′-ATGGCCAAGCTTCA-TATGG(1)CTGATACAGTTTGCTTC(18) antisense 5′-ATCCGGATCCTCAA(651)GA-GACCCATTCCGG(637)
F2	sense 5′-CCGGCATATGA(349)CTG-GAAACGACCCTTCC(366) antisense 5′-CCGGCTCGAGGATCCT(765)-CACATGTCCGTTATGGC(748)
F3	sense 5′-primer of F2 antisense 5′-primer of F1

Numbers in parentheses are the bp numbers in the GS-II cDNA that encodes the mature protein [15]. The bacterial expressed truncated peptides were contained in inclusion bodies. These peptides were solubilized in a solution of 8 M urea in 100 mM Tris-HCl buffer (pH 8.0) and renatured through a series of dialyses against 4 M, 2 M and 1 M urea in 100 mM Tris-HCl buffer (pH 8.0) and, finally, against 100 mM NaCl, 20 mM $MgCl_2$, 1 mM $CaCl_2$ and 1 mM $MnCl_2$. Soluble bacterial recombinant GS-II was also added to a solution of 8 M urea and subsequently dialyzed in the same manner for use as an appropriate control. GS-II and peptide fragments were fractionated by GlcNAc affinity chromatography to assess carbohydrate binding [15].

2.2. Site-specific mutations in GS-II

The template for PCR-based site-directed mutagenesis (Chameleon mutagenesis kit, Stratagene) was pET9c containing the insert that encoded the mature leaf large subunit of GS-II [15]. Point mutations were made to codons that encoded Asp^{88}, Glu^{108}, Tyr^{134}, Asn^{136}, Leu^{226} and Gln^{227} (Table 1) that were predicted to be monosaccharide binding residues in GS-II [15]. PCR was conducted using a selection primer and mutagenic primers that incorporated the desired point mutations into the cDNA. All mutations were confirmed by DNA sequence analyses. Recombinant proteins were produced and isolated as described previously [15]. Recombinant protein-producing bacterial cells were collected by a brief centrifugation at $5000 \times g$ for 5 min and resuspended in extraction buffer (50 mM Tris-HCl, pH 8.0, 2 mM EDTA, 100 mM NaCl, 20 mM $MgCl_2$, 1 mM $CaCl_2$ and 1 mM $MnCl_2$). Lysozyme and Triton X-100 were added for 15 min at 30°C to lyse the bacteria. This treatment was followed by a 1 min sonication. After centrifugation at $10{,}000 \times g$ for 20 min, the supernatant was loaded onto a GlcNAc affinity column (E.Y. Laboratory, Inc., San Mateo, CA). The column was washed with solution containing 100 mM NaCl, 20 mM $MgCl_2$, 1 mM $CaCl_2$ and 1 mM $MnCl_2$ and then eluted with GlcNAc dissolved in the above buffer. The proteins in the wash and the eluate were separated by SDS-PAGE in order to assess whether GS-II was in the fraction.

3. Results and discussion

3.1. Truncated GS-II peptides

Three GS-II fragments were produced as insoluble proteins in bacteria (Fig. 1), solubilized and evaluated for carbohydrate binding activity. None of the peptides were able to bind GlcNAc (Fig. 2). Recombinant mature GS-II that was subjected to 8 M urea treatment retained carbohydrate binding activity. Consequently, it is unlikely that the solubilization treatment was the basis for lack of GlcNAc binding affinity by the truncated peptides. Analyses of these data indicate that residues within the deleted regions are essential for carbohydrate binding or indicate that substantial *N*- or *C*-terminal deletion to mature GS-II destabilizes the tertiary structure of the protein resulting in attenuation of carbohydrate binding.

A small motif (7–11 amino acids) that is present in virtually all legume lectins is purported to be the carbohydrate and metal binding domain [16–19]. The individual domain peptides retained the same substrate specificity as the proteins from which they were derived. The corresponding carbohydrate binding motif (Asp^{132} to Ile^{138}) was included in all three of the GS-II deletion fragments, yet these did not retain GlcNAc binding activity indicating that other amino acids are involved in carbohydrate binding. These contradictory results may be due to different procedures that were used by the individual research groups for evaluation of carbohydrate binding. Alternatively, the lack of tertiary structure in the 7–11 amino acid peptides may allow some interaction with sugars that cannot occur with a larger peptide that has conformational restrictions (e.g. F1, F2 and F3).

3.2. Lectin activities of site-specific mutated GS-II proteins

Listed in Table 1 are the different mutated GS-II proteins that were evaluated and the effect of each specific mutation on GlcNAc binding affinity. Mutations to Asp^{88}, Tyr^{134}, and Asn^{136} all resulted in a loss of GlcNAc binding implicating these amino acids as functional residues in carbohydrate binding. Proteins containing mutations to Glu^{108}, Leu^{226} or Gln^{227} maintained affinity for GlcNAc despite the fact that these residues were substituted with amino acids with substantially different chemical properties, e.g. from a charged to an uncharged R group, or vice versa. Consequently, it is unlikely that these residues are functionally involved in carbohydrate binding. Asn^{196} was not predicted to be involved in carbohydrate binding so it was not surprising that a mutation to this residue had no effect on GlcNAc binding.

Asp^{88} is conserved in most legume lectins and our results confirm the importance of this residue in carbohydrate binding [3]. By analogy with LOL1, Asp associates with the sugar by hydrogen bonding of the side chain COOH with the C6 of

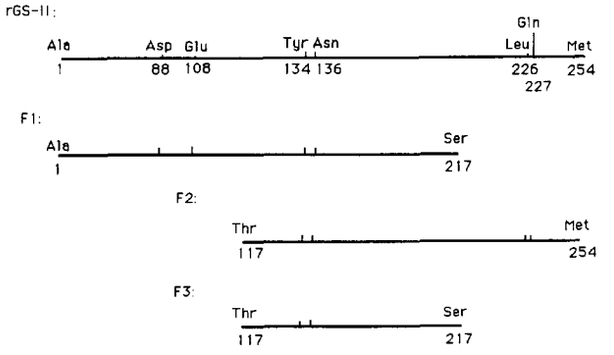

Fig. 1. Schematic diagram of truncated peptides of GS-II (F1, F2 and F3) which were evaluated for GlcNAc binding. Illustrated are the amino acid residues at the beginning and end of each deleted peptide and the amino acids in GS-II which were subjected to site-directed mutagenesis. The rGS-II peptide is the mature protein expressed as a bacterial recombinant protein [15].

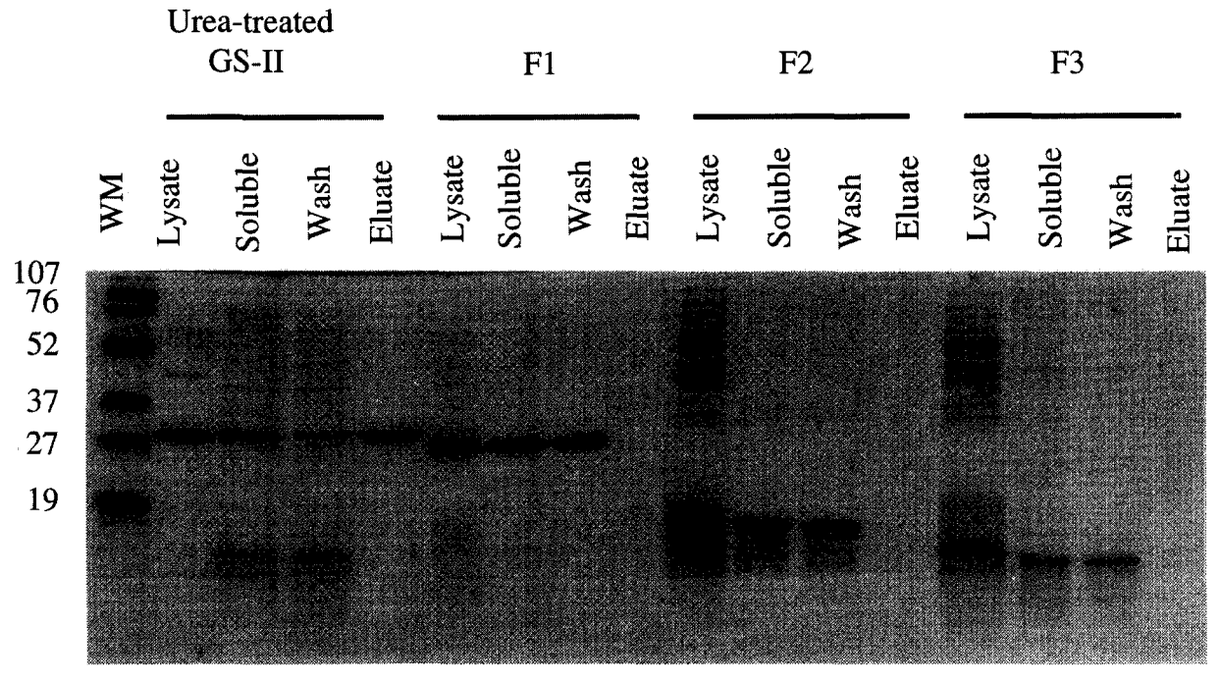

Fig. 2. Truncated GS-II peptides do not have GlcNAc binding affinity. IPTG-induced bacterial cells were lysed and the proteins (GS-II, F1, F2 and F3) dissolved in SDS-PAGE protein loading buffer (lysate). All truncated peptides were contained in inclusion bodies. Consequently, prior to fractionation, the proteins were solubilized in urea and extensively dialyzed to obtain the following fractions: soluble proteins (soluble), GlcNAc affinity column wash after loading of soluble proteins (wash) and affinity column GlcNAc eluate (eluate); MW identify the lane containing molecular weight standards. Soluble GS-II was subjected to the same urea treatment as the truncated peptides. Illustrated are Coomassie blue-stained gels.

hydroxyl group of the monosaccharide [9]. Accordingly, the terminal amide in Asn blocked this interaction. Replacement of Asn^{136} either with an opposite (Asp) or similarly (Gln) charged residue disrupted GlcNAc binding. These results confirm the essential function for this amino acid in carbohydrate binding and indicate that precise positioning of this residue is critical as even a small increase in side chain length (one carbon) eliminated activity. This residue is the most conserved amino acid present in legume lectins [3]. Also, mutation of Asn to Asp in both pea lectin and PHA-L eliminated carbohydrate binding and leucoagglutinating and mitogenic activities [4,20].

Tyr^{134} is not highly conserved in the primary sequence of legume lectins. Leu, Phe, Trp and even Cys occupy this topological position in carbohydrate binding legume proteins [3]. It seems that any amino acid with an uncharged R group, that is of sufficient length, can occupy this position, regardless of whether the side chain is aliphatic or contains an aromatic ring, and the protein will retain sugar binding activity. Replacing Tyr with Phe in GS-II did not affect GlcNAc binding (Fig. 3) indicating that the hydroxyl group is dispensable. The small increase in hydrophobicity which results from the substitution of Phe for Tyr also did not affect binding affinity appreciably. However, replacing Tyr with Asp or Gly abolished binding activity (Table 1). Thus, introducing a charged side chain or markedly reducing the size of the side chain at this position in GS-II substantially impacts GlcNAc binding activity, presumably by affecting the non-polar interaction between the aromatic ring and the C5 and C6 atoms of the sugar [8,9].

GS-II proteins with mutations at Glu^{108}, Leu^{226} and Gln^{227} unpredictably retained capacity to bind GlcNAc. Perhaps the amino acids involved in sugar binding were incorrectly identified. It is possible also that none of these residues individually is essential but are functional collectively as a motif for carbohydrate binding [10]. The crystallographic data of other analogous legume lectins do indicate that residues in this spatial topology are linked to monosaccharide binding [9]. Perhaps the inconsistency of our results with predictions that these residues are involved in carbohydrate binding can be resolved by introducing alternative or multiple mutations into residues surrounding Glu^{108}, Leu^{226} and Gln^{227}. It is interesting to note that mutations at Leu^{226} and Gln^{227} retained GlcNAc binding activity, whereas a deletion that removed these two residues from a peptide fragment that retained the other residues proven to be essential for carbohydrate binding (Asp^{88}, Tyr^{134}, Asn^{136}), resulted in a peptide (F1) which was incapable of binding GlcNAc. It is possible that other residues

Table 1
Site-specific mutations in GS-II that eliminate GlcNAc binding affinity

Amino acid		GlcNAc binding activity
Original	Mutation	
Asp^{88}	Asn	No
Glu^{108}	Gln	Yes
Tyr^{134}	Asp	No
Tyr^{134}	Gly	No
Tyr^{134}	Phe	Yes
Asn^{136}	Asp	No
Asn^{136}	Gln	No
Leu^{226}	Lys	Yes
Gln^{227}	Glu	Yes
Asn^{196}	Asp	Yes

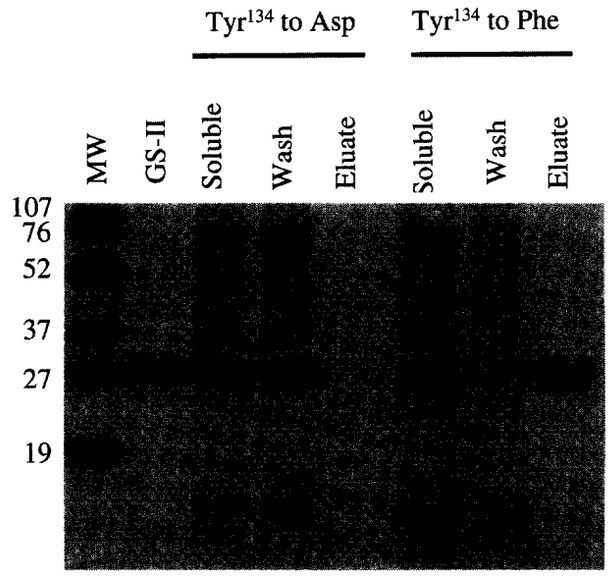

Fig. 3. Tyr^{134} is an essential residue for GlcNAc binding of GS-II. Shown are Coomassie blue-stained SDS-PAGE separated GS-II or GS-II proteins with Tyr^{134} to Asp or Phe mutations. MW = molecular weight standards; GS-II = GlcNAc affinity-purified recombinant protein; soluble = bacterial lysate supernatant, wash = GlcNAc affinity column wash after loading soluble proteins; eluate = affinity column eluate.

in the fragment deleted from F1 directly interact with GlcNAc or that the tertiary structure of GS-II is disturbed by the deletion, which consequently prevents sugar binding.

Three-dimensional structure analyses indicate that legume lectins have similar sugar binding motifs [9], although they may have substantially different carbohydrate specificities [3]. The amino acid residues that are presumed to interact with sugars in the carbohydrate binding site are spatially located in the primary sequences of the lectins. However, with the exception of residues that correspond to Asp^{88} and Asn^{136} in GS-II, the carbohydrate binding amino acids are not absolutely conserved, even amongst proteins with the same carbohydrate binding specificity [21,22]. Our results indicate that Asp^{88}, Tyr^{134} and Asn^{136} are essential for carbohydrate binding affinity, probably because of their direct interaction with GlcNAc. However, these residues may not be responsible for GlcNAc specificity, since these amino acids are rather invariant in the primary structure of legume lectins.

Several legume lectins have been shown to have insecticidal activities [13]. Presumably, the anti-insect properties are linked to the carbohydrate binding properties of these proteins [12,13]. Interaction of lectins with glycoproteins in the insect alimentary tract may inhibit nutrient uptake, or otherwise disrupt gut function or integrity resulting in mortality or delayed development. The present study provides the foundation for determining the relationship between GlcNAc binding of GS-II and its insecticidal activity. If carbohydrate binding is the basis for insecticidal activity, presumably modifications to the sugar binding properties of lectins would increase or decrease affinity for complimentary molecules in the insect guts resulting in increased or decreased insecticidal efficacy.

References

[1] Barondes, S.H. (1988) TIBS 13, 480–482.
[2] Chrispeels, M.J. and Raikhel N.V. (1991) Plant Cell 3, 1–9.
[3] Sharon, N. and Lis, H. (1990) FASEB J. 4, 3198–3208.
[4] Mirkov, T.E. and Chrispeels, M.J. (1993) Glycobiology 3, 581–587.
[5] Delbaere, L.T.J., Vandonselaar, M., Prasad L., Quail, J.W., Wilson, K.S. and Dauter, Z. (1993) J. Mol. Biol. 230, 950–965.
[6] Einspahr, H., Parks, E.H., Suguna, K., Subramanian, E. and Suddath F.L. (1986) J. Biol. Chem. 261, 16518–16527.
[7] Reeke, G.N. Jr. and Becker, J.W. (1986) Science 234, 1108–1111.
[8] Derewenda, Z., Yariv, J., Helliwell, J.R., Kalb, A.J., Dodson, E.J., Papiz, M.Z., Wan, T. and Campbell, J. (1989) EMBO J. 8, 2189–2193.
[9] Bourne, Y., Roussel, A., Frey, M., Rouge, P., Fontecilla-Camps, J. and Cambillau, C. (1990) Proteins: Struct. Funct. Genet. 8, 365–376.
[10] Arango, R., Rodriguez-Arango, E., Adar, R., Belenky, D., Loontiens, F.G., Rozenblatt, S. and Sharon, N. (1993) FEBS Lett. 330, 133–136.
[11] Rouge, P., Barre, A., Causse, H., Chatelain, C. and Porthe, G. (1993) Biochem. Syst. Ecol. 21, 695–703.
[12] Czapla, T.H. and Lang, B.A. (1990) J. Econ. Entomol. 83, 2480–2485.
[13] Murdock, L.L., Huesing, J.E., Nielsen, S.S., Pratt, R.C. and Shade, R.E. (1990) Phytochemistry 29, 85–89.
[14] Habibi, J., Backus, E.A. and Czapla, T.H. (1993) J. Econ. Entomol. 86, 945–951.
[15] Zhu, K., Huesing, J.E., Shade, R.E., Bressan, R.A., Hasegawa, P.M. and Murdock, L.L. (1996) Plant Physiol. 110, 195–202.
[16] Yamamoto, K., Konami, Y., Kusui, K. and Osawa, T. (1991) FEBS Lett. 281, 258–262.
[17] Yamamoto, K., Konami, Y., Osawa, T. and Irimura, T. (1992) J. Biochem. 111, 436–439.
[18] Konami, Y., Yamamoto, K., Osawa, T. and Irimura, T. (1992) FEBS Lett. 304, 129–135.
[19] Konami, Y., Yamamoto, K., Osawa, T. and Irimura, T. (1992) J. Biochem. 112, 366–375.
[20] Van Eijsden, R.R., Hoedemaeker, F.J., Diaz, C.L., Lugtenberg, B.J.J., de Pater, B.S. and Kijne, J.W. (1992) Plant Mol. Biol. 20, 1049–1058.
[21] Strosberg, A.D., Buffard, D., Lauwereys, M. and Foriers, A. (1986) in: Legume Lectins: a Large Family of Homologous Proteins (Liener, I.E., Sharon, N. and Goldstein, I.J. eds) The Lectins Properties, Functions, and Applications in Biology and Medicine. pp. 249–264, Academic Press, Orlando, FL.
[22] Pusztai, A. (1991) in: Structure of Lectins (Phillipson, J.D., Ayres, D.C. and Baxter, H. eds) Plant Lectins. pp. 32–38, Cambridge University Press, Cambridge, UK.