Glycoconjugate Journal (1994) **11**:572–575

Cloning and sequence analysis of the *Maackia amurensis* haemagglutinin cDNA

KAZUO YAMAMOTO, CHIZU ISHIDA, MITSUE SAITO, YUKIKO KONAMI, TOSHIAKI OSAWA‡ and TATSURO IRIMURA*

Division of Chemical Toxicology and Immunochemistry, Faculty of Pharmaceutical Sciences, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

Received 19 May 1994, Revised 25 July 1994

Maackia amurensis haemagglutinin (MAH) is a leguminous lectin which preferentially binds to a cluster of sialylated *O*-linked carbohydrate chains (Konami Y, Yamamoto K, Osawa T, Irimura T (1994) *FEBS Lett* **342**:334–38). In the present study a 950 bp cDNA clone encoding MAH was isolated from a cDNA library constructed from germinated *Maackia amurensis* seeds. From the nucleotide sequence, MAH was predicted to consist of 285 amino acid residues containing a signal peptide of 29 amino acids. The results also confirmed our previous findings from the amino acid sequence analysis, which indicated that two highly conserved amino acid residues in all other well-known leguminous lectins were replaced in MAH. These residues were lysine-105 and aspartic acid-135. The corresponding amino acid residues in other leguminous lectins were glycine and asparagine, respectively. These differences were due to the presence of nucleotides AAA and GAT in place of AAT/C and GGA/T.

Keywords: *Maackia amurensis* haemagglutinin, cDNA cloning, amino acid sequence, carbohydrate recognition

Abbreviations: MAH, *Maackia amurensis* haemagglutinin.

Introduction

The presence of a lectin with haemagglutination activity in *Maackia amurensis* seeds was first described by Boyd *et al.* [1]. Two isolectins, haemagglutinin (MAH) and leukoagglutinin (MAL) were previously purified from seeds in our laboratory [2]. The carbohydrate binding specificity of these isolectins, particularly MAL, as compared with other leguminous lectins of related characteristics has been investigated in several laboratories [3, 4]. The properties of MAH was less understood than that of MAL until recently. It was previously reported that MAH primarily bound the sialic acid-containing carbohydrate chains of porcine submaxillary mucin. The results of competitive binding studies of MAH and other legume lectins to human erythrocytes [5] and the quantitative inhibition assay of MAH with various glycoproteins and glycopeptides [6] suggested that this lectin bound *O*-linked sugar chains. Recently, the precise carbohydrate-binding specificity of MAH was determined by lectin affinity chromatography. In these studies, the amino terminal octapeptide (CB-II) derived from human glycophorin A (containing three sialyl oligosaccharides linked to Ser-Thr-Thr) was found to have the strongest affinity for MAH among the different glycopeptides and oligosaccharides that were tested [7].

We have also taken another approach to elucidate the carbohydrate specificity of leguminous lectins, which is to determine the structure of carbohydrate-recognition domains. Based on this information, we have already identified the putative carbohydrate-recognition domain of MAH based on its amino acid sequence [8]. The location of the carbohydrate-recognition domain of leguminous lectins has been determined by a combination of physicochemical and molecular biological studies. For example, we have cloned cDNA coding for the *Bauhinia purpurea* lectin [9] and the carbohydrate-binding specificity of this lectin was altered by constructing a chimeric lectin with the carbohydrate-recognition domain of *Lens culinaris* lectin [10]. To elucidate eventually the relationship between the primary amino acid sequence of leguminous lectins and carbohydrate specificity, we aimed at obtaining cDNA clones corresponding to *Maackia amurensis* lectins. In the present report, we isolated and characterized a cDNA clone representing the entire coding sequence of MAH. The availability of this clone will enable us to examine the

* To whom correspondence should be addressed.
‡ Present address: Yakult Central Institute for Microbiological Research, Yaho, Kunitachi-shi, Tokyo 186, Japan.

correlation between its amino acid sequence and carbohydrate-binding specificity, especially in comparison with the interaction of other leguminous lectins with sialyl residues.

Materials and methods

Materials

The *Maackia amurensis* seeds were obtained from F. W. Schumacher (Sandwich, MA, USA). Sephadex G-25 was purchased from Pharmacia (Uppsala, Sweden). Restriction enzymes and DNA modifying enzymes were obtained from Pharmacia, New England Biolabs (Beverley, MA, USA), and Boehringer GmbH (Mannheim, Germany). DNA sequencing reagents were obtained from ABI (Foster City, CA, USA). All other reagents were of analytical grade.

RNA isolation and cDNA cloning

Maackia amurensis seeds were germinated and grown on wet cotton at 31°C for 5–8 days. Total RNA was isolated from germinated *Maackia amurensis* seeds by the phenol-sodium dodecyl sulfate method [11], followed by guanidium thiocyanate-CsCl centrifugation [12]. PolyA$^+$ RNA was prepared by oligo dT-cellulose column chromatography [13]. PolyA$^+$ RNA (5 μg) was used to construct a double-stranded cDNA by the method of Gubler and Hoffman [14]. Double-stranded cDNA thus obtained was methylated with *Eco* RI methylase, followed by the addition of *Eco* RI linkers. The linked cDNA was ligated into the *Eco* RI site of λgt10 cloning vector and packaged with *in vitro* packaging extract (Stratagene, La Jolla, CA, USA). The cDNA library, consisting of 1.2×10^6 independent clones was used to transform *E. coli* C600hfl$^-$.

Isolation of cDNA clones coding MAH

Recombinant phages with host cells were plated on 2 × YT agar plates and incubated at 37°C for 14 h. The amplified library was transferred to nylon membrane filters (Pall, Glen Cove, NY, USA). The replicated filters were hybridized with a synthetic DNA mixed probe (5′GTNGAA/GACNCAC/TGAC/TGT3′) representing all possible DNA sequences predicted for the amino acid sequence, VETHDV, determined by peptide sequencing of a Lys-C fragment [8]. Following hybridization and washing as described by Maniatis *et al.* [15], the filters were exposed to Kodak XAR-5 film at −80°C. Positively hybridizing clones with the above probe were isolated and phages collected according to standard procedures [16]. The cDNA inserts were amplified by means of polymerase chain reaction (PCR) using λgt10 primers (5′AGCAAGTTCAGCCTGGTTAAG3′ and 5′CTTATGAGTATTTCTTCCAGGGTA3′, New England Biolabs), and examined by Southern blot hybridization using the labelled probe described above [17]. To obtain full length cDNA clones, further screening was carried out using a positively hybridizing cDNA insert, after the DNA was labelled with ^{32}P-dCTP by a random primer labelling kit (Amersham, Buckinghamshire, England).

Subcloning of cDNA inserts

cDNA inserts positively hybridizing with the probe were ligated with the vector, pT7T3-18U (Pharmacia LKB Biotechnology, Uppsala, Sweden), and then transformed into competent *E. coli* JM109 cells. Recombinant plasmids were isolated by the alkaline lysis method [18].

DNA sequence analysis

The DNAs were sequenced by the dideoxy chain termination method [19]. MAH cDNA fragment was cleaved with *Bam* HI and the resulting smaller fragments were subcloned into pBluescript SK II$^+$ (Clonetech, Palo Alto, CA, USA) followed by sequence analysis from double-stranded templates. The sequencing strategy for determining the nucleotide sequence is given in Fig. 1.

Results and discussion

Poly A$^+$ RNA from germinated *Maackia amurensis* seeds was used to construct a cDNA library in the λgt10 vector. The bacteriophage was propagated in *E. coli* C600hfl$^-$, and the resultant 5.0×10^4 plaques were screened with a mixed oligonucleotide probe as described in the Materials and methods section. From the library, two clones (MAH1 and

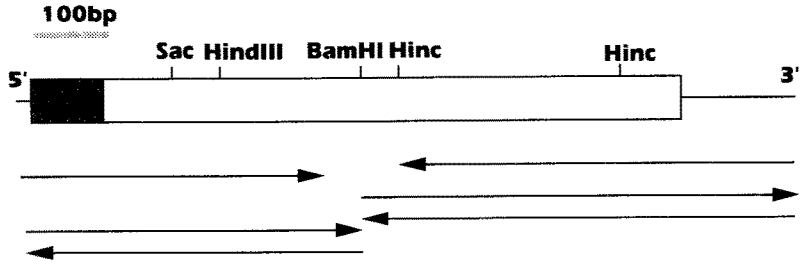

Figure 1. Restriction map and sequencing strategy of the MAH cDNA clone. The coding region is boxed. The black box indicates the signal peptide. Restriction sites are indicated. Arrows indicate the length of the fragments and the direction of sequencing.

MAH2) which gave strong hybridization signals with the probe were detected. From the southern blot hybridization, it was apparent that the two clones contained inserts of about 335 and 220 bp, respectively. These cDNAs were amplified by means of PCR and ligated into the sequencing plasmid pT7T3-18U. Plasmids carrying cDNA inserts were recovered and designated as pMAH1 and pMAH2, respectively. Nucleotide sequences of these cDNA clones indicate that both clones encoded MAH. However, the size of the cDNAs in pMAH1 and pMAH2 was smaller than expected for a full-length cDNA encoding the lectin. Thus, further screening of full-length cDNA coding MAH was performed by use of pMAH1 cDNA as a hybridization probe. Twelve of the 5×10^4 recombinant phages hybridized with the cDNA probe. The largest clone of those isolated from the 12 positive clones, pMAH-F, contained an insert of approximately 1000 bp. The sequencing strategy and restriction map of the pMAH-F cDNA clone are given in Fig. 1. The nucleotide sequence of pMAH-F and its deduced amino acid sequence are shown in Fig. 2. The cDNA included a coding region of 855-nucleotides with a single initiation codon ATG at position 3–5. The primary sequence analysis of MAH [6] previously showed that serine was residue 1 and aspartic acid residue 2. The open reading frame of the cDNA encoded a polypeptide of 285 amino acids including a signal peptide composed of 29 amino acids (upstream of Ser-Asp). The estimated molecular weight of the deduced sequence minus the signal sequence was 27 591. Figure 3 shows that the deduced sequence was identical to the sequence of MAH, as determined previously by peptide sequencing, except at positions 50 and 203 which contained Ala and Gly instead of Val and Glu, respectively. The sequences of the other positive clones in these regions were examined and found to be identical to that of pMAH-F. The differences between the deduced sequence and the previously determined sequence might be explained by a possibility that haemagglutinins of *Maackia amurensis* seeds might be coded by a family of closely related genes. Alternatively, the differences might be due to an error during the peptide sequencing. As underscored by solid lines in Fig. 2, the potential glycosylation sites of MAH were Asn-111, Asn-177 and Asn-189. These sites were located in the sequences of -Phe-Asn-Asn-Ser-, -Pro-Asn-Lys-Thr- and Ser-Asn-Gln-Thr-. In the previous protein sequence analyses the Asn residues at these sites were not detectable, as shown in Fig. 3 by X. Therefore, it was plausible that all three sites are glycosylated in the mature lectin.

As mentioned above, the deduced sequence showed 29 amino acid residues before Ser-1 which most likely represents a signal leader peptide needed for the translocation of the immature lectin into the endoplasmic reticulum. All legume lectins studied so far have such a signal peptide which is usually between 20 and 30 amino acids long and characteristically very hydrophobic. At the C-terminal end of the deduced sequence, there were nine additional amino

```
       10        20        30        40        50        60
gccatggctacttccaactcaaaaccaactcaagtccttcttgccaccttcttaactttc
   M  A  T  S  N  S  K  P  T  Q  V  L  L  A  T  F  L  T  F

       70        80        90        100       110       120
ttccttttgctactcaacaacgtaaactcatcagatgagctttcttttaccatcaacaat
F  L  L  L  L  N  N  V  N  S  S  D  E  L  S  F  T  I  N  N

       130       140       150       160       170       180
ttcatgccaaatcaaggcgatctactcttccaaggtgtagccactgtttcaccaacaggg
F  M  P  N  Q  G  D  L  L  F  Q  G  V  A  T  V  S  P  T  G

       190       200       210       220       230       240
gtattacaacttaccagcgaagaaaacggtcaacccctggagtattctgttggcagagct
V  L  Q  L  T  S  E  E  N  G  Q  P  L  E  Y  S  V  G  R  A

       250       260       270       280       290       300
ctatatactgcccctgtgcgcatttgggacagtaccactggcgccgtagcaagcttctcc
L  Y  T  A  P  V  R  I  W  D  S  T  T  G  A  V  A  S  F  S

       310       320       330       340       350       360
acttccttcacctttgttgtgaaagcagctaggggagcttctgacggtttagccttcttt
T  S  F  T  F  V  V  K  A  A  R  G  A  S  D  G  L  A  F  F

       370       380       390       400       410       420
cttgcaccacctgattctcagatcccttcgggcagcgtatcgaaatacctaggacttttt
L  A  P  P  D  S  Q  I  P  S  G  S  V  S  K  Y  L  G  L  F

       430       440       450       460       470       480
aacaactcaaattccgatagttccaaccaaattgttgctgtagagtttgacacttacttc
N  N  S  N  S  D  S  S  N  Q  I  V  A  V  E  F  D  T  Y  F
*
       490       500       510       520       530       540
ggccatagttatgatccctgggatccaaattatcgacatatcggaattgatgtcaacggt
G  H  S  Y  D  P  W  D  P  N  Y  R  H  I  G  I  D  V  N  G

       550       560       570       580       590       600
attgagtcgataaaaactgtgcaatgggattggattaacggcggagttgcctttgctacc
I  E  S  I  K  T  V  Q  W  D  W  I  N  G  G  V  A  F  A  T

       610       620       630       640       650       660
ataacctatctagctcccaacaaaacgttaatagcatctctagtttaccccttccaatcaa
I  T  Y  L  A  P  N  K  T  L  I  A  S  L  V  Y  P  S  N  Q
                 *                                       *
       670       680       690       700       710       720
acaagtttcattgtcgctgcttctgttgatttgaagggaatcctccctgagtgggttaga
T  S  F  I  V  A  A  S  V  D  L  K  G  I  L  P  E  W  V  R

       730       740       750       760       770       780
gttggtttctctgctgccacgggtgctcctaaagcagttgaaacccacgatgttcgttcc
V  G  F  S  A  A  T  G  A  P  K  A  V  E  T  H  D  V  R  S
                                    ----------------
       790       800       810       820       830       840
tggtctttcacgtcaactttggaagccaacagccctgctgatgtggataataatgtgcat
W  S  F  T  S  T  L  E  A  N  S  P  A  D  V  D  N  N  V  H

       850       860       870       880       890       900
atcgcacgttacactgcatgatctcgtgagctttcgtatgtattaggtgtttatgtaaat
I  A  R  Y  T  A  *

       910       920       930       940       950
taaataaaaatgacctgaaataatggttatcggcgcagctatacaaaaat
```

Figure 2. The nucleotide sequence and the deduced amino acid sequence of the MAH cDNA clone. The nucleotides are listed in the 5′ to 3′ direction. The termination site is marked with a star. The signal sequence is indicated by the thick solid line. The amino acid sequences used for probes are underlined with dotted lines. Asterisks represent sites of potential asparagine-linked glycosylation.

acid residues not found in the sequence determined at the protein level. The difference might be due to incompleteness of the chemical sequence or might suggest the possibility of a post-translational modification at the C-terminal of the protein, as has been proposed for other lectins [20].

The cDNA sequence of MAH was 59.0% homologous to

```
A)  -29  MATSNSKPTQVLLATFLTFFFLLLNNVNSSDELSFTINNFMPNQGDLLFQGVATVSPTGV
                                          ********************************
B)  1'                                    SDELSFTINNFMPNQGDLLFQGVATVSPTGV

A)  32   LQLTSEENGQPLEYSVGRALYTAPVRIWDSTTGAVASFSTSFTFVVKAARGASDGLAFFL
         ****************** ****************************************
B)  32'  LQLTSEENGQPLEYSVGRVLYTAPVRIWDSTTGAVASFSTSFTFVVKAARGASDGLAFFL

A)  92   APPDSQIPSGSVSKYLGLFNNSNSDSSNQIVAVEFDTYFGHSYDPWDPNYRHIGIDVNGI
         ****************** ****************************************
B)  92'  APPDSQIPSGSVSKYLGLFXNSNSDSSNQIVAVEFDTYFGHSYDPWDPNYRHIGIDVNGI

A)  152  ESIKTVQWDWINGGVAFATITYLAPNKTLIASLVYPSNQTSFIVAASVDLKGILPEWVRV
         ************************ *********** ************* ********
B)  152' ESIKTVQWDWINGGVAFATITYLAPXKTLIASLVYPSXQTSFIVAASVDLKEILPEWVRV

A)  212  GFSAATGAPKAVETHDVRSWSFTSTLEANSPADVDNNVHIARYTA
         ****************************************---------
B)  212' GFSAATGAPKAVETHDVRSWSFTSTLEANSPADVDN
```

Figure 3. Comparison of the deduced amino acid sequence of MAH cDNA (A) with the sequence of the lectin as determined at the protein level (B). Amino acid numbers are shown in the second column. Asterisks indicate identical amino acid residues. The leader sequence of the lectin is underlined. There are nine additional amino acids at the C-terminal of the protein in the deduced sequences underscored by a dashed line.

the nucleotide sequence of the coding region of soybean agglutinin [21], 55.7% homologous to the cDNA of the *Dolichos biflorus* lectin [22], 60.7% homologous to the cDNA of the *Erythrina corallodendron* lectin [23], 57.0% homologous to the cDNA of the *Pisum sativum* lectin [20], and 56.8% homologous to the cDNA of *Bauhinia purpurea* lectin [8].

The isolation and characterization of the cDNA for MAH reported in this paper now enables us to utilize recombinant DNA techniques in the future elucidation of the correlation between the structure and the function of this lectin. These studies are currently in progress.

Acknowledgement

This study was supported by grants-in-aid from the Ministry of Education, Science, and Culture of Japan (04670138 and 05274101), the Research Association for Biotechnology, the Hayashi Memorial Foundation for Female Natural Scientists and an Esso Research Grant for Women.

References

1. Boyd WC, Waszczenko-Zacharczenko E, Goldwasser SM (1961) *Transfusion* **1**:374–82.
2. Kawaguchi T, Matsumoto I, Osawa T (1974) *J Biol Chem* **249**:2786–92.
3. Wang WC, Cummings RD (1988) *J Biol Chem* **263**:4576–85.
4. Knibbs RN, Goldstein IJ, Ratcliffe RM, Shibuya N (1991) *J Biol Chem* **266**:83–88.
5. Kawaguchi T, Matsumoto I, Osawa T (1974) *Biochemistry* **13**:3169–73.
6. Kawaguchi T, Osawa T (1976) *Biochemistry* **15**:4581–86.
7. Konami Y, Yamamoto K, Osawa T, Irimura T (1994) *FEBS Lett* **342**:334–38.
8. Konami Y, Yamamoto K, Osawa T, Irimura T (1994) *J Biochem* **115**:767–77.
9. Kusui K, Yamamoto K, Konami Y, Osawa T (1991) *J Biochem* **109**:889–903.
10. Yamamoto K, Konami Y, Osawa T, Irimura T (1992) *J Biochem* **111**:87–90.
11. Watanabe A, Price CA (1982) *Proc Natl Acad Sci USA* **79**:6304–8.
12. Chirgwin JW, Przybla AE, MacDonald RJ, Putter WJ (1979) *Biochemistry* **18**:5294–99.
13. Aviv H, Leder P (1972) *Proc Natl Acad Sci USA* **69**:1408–12.
14. Gubler U, Hoffman BJ (1983) *Gene* **25**:263–69.
15. Maniatis T, Fritsh EF, Sambrook J (1982) In *Molecular Cloning*, pp. 326–28. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
16. Maniatis T, Fritsh EF, Sambrook J (1982) In *Molecular Cloning*, pp. 371–72. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
17. Southern EM (1975) *J Mol Biol* **98**:503–17.
18. Birnboim HC, Doly J (1979) *Nucleic Acids Res* **7**:1513–23.
19. Sanger F, Nicklen S, Coulson AR (1977) *Proc Natl Acad Sci USA* **74**:5463–67.
20. Higgins TJV, Chandler PM, Zurawski G, Button SC, Spencer D (1983) *J Biol Chem* **258**:9544–49.
21. Vodkin LO, Rhodes PR, Goldberg RB (1983) *Cell* **34**:1023–31.
22. Schnell DJ, Etzler ME (1988) *J Biol Chem* **263**:1463–53.
23. Arango R, Rozenblatt S, Sharon N (1990) *FEBS Lett* **264**:109–11.