Molecular analysis of *Physarum* haemagglutinin I: lack of a signal sequence, sulphur amino acids and post-translational modifications

Masashi Morita

Tel: +81 764 34 2281 ext. 2636. Fax: +81 764 34 4656.
e-mail: masa@ms.toyama-mpu.ac.jp

Faculty of Pharmaceutical Sciences, Toyama Medical and Pharmaceutical University, Sugitani, Toyama, 930-01, Japan

The cDNAs encoding haemagglutinin I from plasmodia of *Physarum polycephalum* have been cloned using PCR protocols. The composite haemagglutinin I cDNA sequence, derived from several overlapping clones from PCR fragments, spans 408 nt and the 315 bp ORF encodes a polypeptide of 104 aa without a typical signal sequence. The putative molecular mass deduced from the amino acid sequence (10760.76 Da) corresponds exactly to that determined by electrospray ionization MS (10759.86±0.15 Da), suggesting that haemagglutinin I is not subject to post-translational modification. Haemagglutinin I lacks sulphur amino acids and has a β -sheet as the major secondary structure. Expression of the coding sequence in *Escherichia coli* yielded a product that exhibits the same sugar-binding specificity as natural haemagglutinin I. The deduced amino acid sequence shows little similarity to that of any known lectins and thus apparently represents a novel type of lectin.

Keywords: haemagglutinin, lectin, *Physarum polycephalum*, cDNA cloning, post-translational modification

INTRODUCTION

Lectins are a class of proteins that agglutinate erythrocytes (so-called haemagglutinins) and the agglutination is inhibited by adding monosaccharide, oligosaccharide and/or glycoprotein. Lectins bind stereospecifically and reversibly to a carbohydrate and act as mediators of various physiological cellular events. To date, many amino acid sequences of lectins in higher plants, animals, fungi and bacteria have been revealed. Despite numerous studies and their wide distribution in nature, the physiological functions of these proteins have not yet been clearly established.

Several fungal lectins have been identified and characterized. Discoidins I and II, which are lectins produced by the cellular slime mould *Dictyostelium discoideum*,

have been reported to be involved in cell–substratum attachment (Barondes & Springer, 1987) and in spore coat formation (Cooper *et al.*, 1983), respectively. In contrast, a lectin in the nematophagous fungus *Arthrobotrys oligospora* has been reported to be involved in the capture of nematodes (Rosen *et al.*, 1996a, b). Lectins from plant-pathogenic fungi such as *Sclerotium rolfsii* and *Rhizoctonia solani* function in the recognition and attachment of these fungi by the mycoparasite *Trichoderma* (Inbar & Chet, 1994). A previous study (Morita & Nishi, 1992) has shown that the acellular slime mould *Physarum polycephalum* produces two types of lectins, termed haemagglutinins I and II, which differ in terms of their molecular masses and carbohydrate-binding properties. These haemagglutinating activities were not inhibited by various monosaccharides but were strongly inhibited by thyroglobulin. We have also reported that an acidic polysaccharide from the bacterial cell wall might be a relevant ligand for haemagglutinin (Morita *et al.*, 1995) and have hypothesized that *Physarum* haemagglutinins might play a physiological role in the interaction with bacteria in nature. However, the apparent role of haemagglutinins is not yet clear.

Abbreviations: ESI-MS, electrospray ionization MS; RACE, rapid amplification of cDNA ends; RP-HPLC, reverse-phase HPLC; TFA, trifluoroacetic acid.

The GenBank accession number for the sequence reported in this paper is U86761.

The slime mould is a suitable organism for the study of cell growth and development at the biochemical and molecular levels. It has been reported that the expression of numerous genes is regulated during the *Physarum* life cycle (Schreckenbach & Werenskiold, 1986). The expression of haemagglutinins is thought to be developmentally regulated because haemagglutinating activities have been found in the plasmodial stage (diploid) but not in the myxamoebae stage (haploid). Determination of the amino acid sequence of haemagglutinin I is needed to obtain more information on the role of the protein and to understand its sugar-binding specificity. Here we report the determination of the complete nucleotide sequence of haemagglutinin I cDNA and the expression of recombinant active haemagglutinin I in *Escherichia coli*.

METHODS

Cultivation of microplasmodia. Microplasmodia of *P. polycephalum* were grown in a semi-defined medium containing glucose, yeast extract, tryptone and haemin (Daniel & Baldwin, 1964). The axenic culture was maintained in 100 ml liquid medium and shaken at 26 °C in the dark.

Purification of haemagglutinin I. Purification of haemagglutinin I from microplasmodia was performed as described previously (Morita & Nishi, 1992). Haemagglutinating activity and haemagglutination inhibitory activity were measured with trypsin-treated, formalin-fixed rabbit erythrocytes as described previously (Morita *et al.*, 1995).

Reverse-phase (RP) HPLC. Purified haemagglutinin I (60 µg) was treated with 8 M urea in 0.1 M ammonium bicarbonate, pH 8.0, containing 2 mM β -mercaptoethanol for 45 min at 60 °C. When the concentration of urea reached 2 M, tryptic or chymotryptic cleavage was carried out with trypsin (Seikagaku Kogyo) or bovine pancreas chymotrypsin VII (Sigma), respectively, at an enzyme/substrate ratio of 1:50 (w/w) at 37 °C for 3 h. After lyophilization, the resultant peptides were dissolved in 30 µl 0.1% (v/v) trifluoroacetic acid (TFA) and applied to a Cosmosil 5C18-MS (C18) reverse-phase column (4.6 × 150 mm, Nacalai Tesque) equilibrated with 0.1% TFA for HPLC. Peptide fragments were separated using the following gradient of acetonitrile in 0.1% TFA: 10% at 0 min, followed by a linear gradient to 55% at 40 min and finally a linear gradient to 80% at 50 min. Peptides were detected at A_{220} . Peptide fragments were collected manually over several runs and the pure peptides were sequenced using a Shimadzu PPSQ-10 protein sequencer.

In another experiment, purified haemagglutinin I (300 µg) was denatured in 6 M guanidine hydrochloride and subjected to RP-HPLC, as above. Haemagglutinin I was eluted with a linear gradient (0–80%) of acetonitrile in 0.1% TFA at a flow rate of 0.8 ml min⁻¹ for 50 min. Elution was monitored at A_{280} .

Tricine-SDS-PAGE and blotting. Purified haemagglutinin I was subjected to Tricine-SDS-PAGE according to the procedure of Schagger & van Jagow (1987) as described previously (Morita & Nishi, 1992). For peptide sequencing, haemagglutinin I was transferred to a PVDF membrane (0.2 µm, Bio-Rad) using a semi-dry electroblotting system (Bio-Rad) and applied directly to an automated protein sequencer according to the procedure of Matsudaira (1987).

Electrospray ionization MS (ESI-MS). The molecular masses of protein or peptide were determined by ESI-MS using an AP-III triple quadrupole mass spectrometer (Perkin-Elmer Sciex Instruments). Calibration of the mass range was achieved by injecting a mixture of PEGs. RP-HPLC-purified proteins or peptides were dissolved in 50% acetonitrile containing 0.1% TFA at a concentration of 10–40 pmol µl⁻¹ and directly introduced into the ion source at a flow rate of 20 µl min⁻¹. The molecular mass spectrum was reconstructed from multiply charged ions in the m/z spectrum.

Purification of poly(A) RNA. Unless stated otherwise, all reagents used were treated with diethylpyrocarbonate. Total RNA was extracted from vegetatively growing microplasmodia of *P. polycephalum* according to the procedure of Nishimura *et al.* (1991). Isolation of poly(A) RNA from the total RNA was carried out using Oligotex-dT30 (Takara) according to the manufacturer's protocol.

cDNA cloning and sequencing. cDNA cloning was done using PCR-based protocols with primers given in Table 1. Restriction digests, DNA ligation and plasmid DNA preparation from *E. coli* were done using standard protocols described in Sambrook *et al.* (1989). First-strand cDNA was prepared by reverse transcription of poly(A) RNA using SuperScript TM II RT (Gibco-BRL) according to the manufacturer's protocol with oligo(dT)-containing Adapter Primer (Gibco-BRL). PCR was performed with the first-strand cDNA as the template using two sets of oligonucleotide mixture designed on the basis of the N-terminal 65 aa sequence. Amplification with primers 1 and 2 produced polynucleotides of approximately 200 bp. This product was then used as template for nested PCR performed with primers 3 and 4. Nested PCR yielded an amplification product of 180 bp. To obtain the 3' or 5' ends, 3'RACE or 5'RACE (rapid amplification of cDNA ends) was performed according to the procedure of Frohman *et al.* (1988). Using the 3'RACE System (Gibco-BRL), oligo(dT)-primed cDNA was amplified with Abridged Universal Amplification Primer (supplemented by the RACE System) and primer 5. The 5'RACE System (Gibco-BRL) was utilized to synthesize first-strand cDNA for the 5' end with primer 6, followed by ligation of the Abridged Anchor Primer. PCR was performed with primer 7 and Abridged Universal Anchor Primer to yield an amplification product. Nested PCR was then performed with the amplification product as the template using primer 8 and Abridged Universal Anchor Primer. All PCR amplifications were performed with *Taq* DNA polymerase (Gibco-BRL) in the buffer composition and other conditions suggested by the manufacturer. PCR was performed on a Perkin-Elmer Cetus DNA thermal cycler (model 480) with 30 cycles of 94 °C for 45 s (denaturation), 55 °C for 25 s (annealing) and 72 °C for 1 min (extension). After the final cycle, DNA was further extended at 72 °C for 7 min.

Each amplified PCR fragment was isolated by electrophoresis using low-melting-point agarose and was directly cloned into the pT7Blue(R)T-vector (TA-cloning system, Novagen). The recombinant plasmid was transformed into *E. coli* XL-1 Blue MRF' (Stratagene) competent cells. Several clones were identified and some of them were sequenced by a conventional dideoxy chain-termination method (Sanger *et al.*, 1977) using Sequencing High -Cycle- (Toyobo) and Sequenase Images (Amersham) according to the manufacturer's protocol. Several clones were sequenced in both directions.

Computer analysis. The GENETYX-SV/DB (Software Development) program was used for analysis of the primary and secondary structure of the protein. The SWISS-PROT and

Table 1. Oligonucleotides used for the cloning of haemagglutinin I cDNA

Primer	Sequence (5'–3')*	Direction†	Target nucleotides‡
1	GTNTGGWSNGTNCARATHGTNGAY	Sense	34–57
2	RTACCANSWYTGRTGNACNGTYTT	Antisense	195–172
3	GTNCARATHGTNGAYAAAYGCNNGN	Sense	43–66
4	YTGRTGNACNGTYTTNNGNCCDATYTC	Antisense	186–160
5	AATTCTTCCACTGTCCTCG	Sense	100–119
6	GAGGGTCAAATGCTGGG	Antisense	330–314
7	ATACTCACATTCCCACCTGCTG	Antisense	293–272
8	TGCTCTGTCATCTCCATTGTGC	Antisense	225–204
9	GAAGCATATGGTCTGGTTCGGTGCAAATA	Sense	24–51
10	TTGTGGATCCTTATGCAAGTTTCTTGAG	Antisense	355–328

* D denotes A or G or T; M, A or C; R, A or G; S, C or G; W, A or T; Y, C or T. Sequences of primers 1, 2, 3 and 4 were designed on the basis of 65 aa N-terminal sequences. Restriction enzyme recognition sites incorporated into primer sequences are indicated by underlining (*Nde*I in primer 9; *Bam*HI in primer 10).

† With respect to the sequence shown in Fig. 1.

‡ Corresponding cDNA sequence numbered as shown in Fig. 1.

NBRF protein sequence databases and GenBank were used for sequence homology searching. The prediction of the hydropathy profile was performed according to the method of Kyte & Doolittle (1982) with a window size of eight residues. The prediction of secondary structure was made using the program of Chou & Fasman (1978).

Northern blot analysis. Northern blot analysis was performed according to the standard method (Sambrook *et al.*, 1989) using total RNA prepared from microplasmodia. About 15 µg total RNA was separated by electrophoresis on a 1% (w/v) agarose/2.2 M formaldehyde gel and transferred onto a nylon membrane (Hybond-N⁺, Amersham) by capillary action. The filter was hybridized with fluorescein-labelled probe, prepared using Gene Images (Amersham) by specific primer extension labelling. A 0.2 kbp cDNA fragment amplified with primers 5 and 7 was labelled with fluorescein-dUTP using primer 8 and used as probe. After hybridization the blot was treated with an alkaline-phosphatase-labelled anti-fluorescein antibody followed by treatment with detection reagents and the filter was exposed to XAR-5 film (Kodak) for 2 h. An RNA ladder (0.16–1.77 kb) was used as a marker.

Expression of haemagglutinin I in *E. coli*. Expression of recombinant protein was done using the pET-3a vector in the BL21 host (pET Expression System 3, Novagen). DNA fragments containing an *Nde*I site covering the start codon and a *Bam*HI site downstream of the desired stop codon were generated by PCR using primers 9 and 10, as shown in Table 1. After digestion with *Nde*I and *Bam*HI, the PCR-generated DNA fragment was inserted between the same sites of pET-3a. *E. coli* NovaBlue (Novagen) was used as a host for the first cloning. The cloned plasmid was then introduced into *E. coli* BL21(DE3)plysS cells grown on LB plates containing 50 µg ampicillin ml⁻¹ and 15 µg chloramphenicol ml⁻¹. The BL21(DE3)plysS cells containing the recombinant plasmid were grown to mid-exponential phase at 37 °C in LB medium, induced by adding IPTG to a concentration of 0.4 mM and incubated at 26 °C. After an additional 3 h culture period, *E. coli* cells were collected by centrifugation and resuspended in 50 mM Tris/HCl, pH 8.0, containing 0.5 mM EDTA, 0.4 M

NaCl, 5 mM MgCl₂, 5% (w/v) glycerol, 0.5 mM PMSF, 1 mM DTT and 1 mg lysozyme ml⁻¹. Then the cell lysate was prepared by repeated freezing and thawing, and sonicated for 20 s 10 times with an Ultrasonic Disrupter (model UR-200P, Tomy Seiko) after the addition of EDTA and Nonidet P40 to final concentrations of 1 mM and 0.5% (w/v), respectively. The cell lysate was centrifuged at 15 000 g for 30 min to remove insoluble materials. After desalting by ultrafiltration (PM 10, Amicon) the supernatant was applied to a column of DEAE-Sephacel (3 × 15 cm, Pharmacia) equilibrated with 10 mM Tris/HCl, pH 7.5. The column was washed with the same buffer to remove unadsorbed materials and subsequently the adsorbed proteins were eluted with the same buffer containing 1 M NaCl. Those exhibiting haemagglutinating activity were pooled and the active fraction was further purified using a thyroglobulin-immobilized Toyopearl HW55 column (3 × 15 cm, Tosoh) as described previously (Morita & Nishi, 1992). In some experiments, the pellet of intact bacteria was directly solubilized in sample buffer and analysed by Tricine-SDS-PAGE.

RESULTS

Partial amino acid sequence

Our previous study (Morita & Nishi, 1992) reported that plasmodia of *P. polycephalum* produce two types of lectins, haemagglutinins I and II, which migrate to about 6 and 11 kDa upon Tricine-SDS-PAGE, respectively. In this study, direct Edman analysis of the band migrating to 6 kDa revealed the sequence of the first 23 N-terminal residues (Table 2). To obtain additional sequence, haemagglutinin I was subjected to tryptic or chymotryptic digestion and the resultant peptides were analysed by Edman degradation and ESI-MS after purification with RP-HPLC (Table 2). ESI-MS can measure molecular masses of polypeptide chains to an accuracy of about 0.01%. Therefore, we applied this method to confirm the amino acid sequence determined by Edman

Table 2. Amino acid sequences and masses of intact haemagglutinin I, and fragments generated from purified haemagglutinin I by trypsin and chymotrypsin treatment

Haemagglutinin I and its tryptic fragments purified by RP-HPLC on a C₁₈ column were subjected to MS to determine their molecular mass. ND, not determined.

Sequence	Molecular mass from ESI-MS (Da)
Intact haemagglutinin I*	
VWSVQIVDNAGLGANLALYPSGN-----	10759.86 ± 0.15
Trypsin fragments†	
Major peak	
T3 TVHQSWY	919.44 ± 0.08
T4 TVHQSWYITVHNGDDR	1927.68 ± 0.14
T6 YVTVTGYAPITFSEIGPK	1943.14 ± 0.35
T8 VWSVQIVDNAGLGANLALYPSGNSSTVPR	2986.73 ± 0.36
Minor peak	
T1 KLA	ND
T2 ITVHNGDDR	ND
T5 TAGGNVSISTGFGDAQHLTLK	2074.23 ± 0.35
T7 YVTVT-----	1331.49 ± 0.70
Chymotrypsin fragments†	
Major peak	
C1 Negative	
C2 Negative	
C5 VW	
C6 SEIGPKTVHQS-----	
C7 SEIGPKTVHQS-----	
C9 APITF-----	
C10 AL-----	
C11 ALYPSGNSSTVPRYV-----	
C13 ALYPS-----	
C14 SVQIVDNA-----	
Minor peak	
C3 ALYPSGN-----	
C4 ITVHNGDDRAF-----	
C8 IVDNAGLGA-----	
C12 Negative	

* Haemagglutinin I blotted on PVDF membrane was directly analysed by an automated protein sequencer.

† Trypsin (T) or chymotrypsin (C) fragments were numbered in accordance with their elution time on RP-HPLC. Dashes indicate that the peptide was not analysed further.

degradation. Digestion of haemagglutinin I with trypsin yielded eight fragments from which most of the N-terminal sequence could be established. The molecular masses of the T3, T4, T5, T6 and T8 fragments, as determined by ESI-MS, were in good agreement with those calculated from the amino acid sequences (919.97, 1928.02, 2074.25, 1943.17 and 2986.29 Da, respectively). Peptide T8 showed a sequence corresponding to a region in the N-terminal part of haemagglutinin I. Most of the peptide fragments were preceded by arginine or lysine, whereas T3 was preceded by tyrosine. This unexpected cleavage may be due to an abnormal action of the enzyme. Tryptic fragments T8, T6 and T4 overlapped with chymotryptic fragments C11, C7 and C4, which led to an unambiguous identification of residues 1–65 (Fig.

1). ESI-MS analysis of intact haemagglutinin I (Table 2) suggested that there should be about another 30 aa residues in the C-terminal region.

Cloning of cDNA encoding haemagglutinin I

Based on the N-terminal 65 aa sequence of haemagglutinin I, two sets of degenerate oligonucleotide primers (1 and 2, 3 and 4) were designed and used to amplify internal cDNA fragments. Several clones were identified and some of them were sequenced on both strands. Using primers designed from the cloned internal cDNA fragments, 5' and 3' end clones were isolated by RACE methods. The sequence of these clones together with internal cDNA fragments showed significant overlap-

```

1 aaacaaaactaatcacacaagaagaagaatggtctggtcggtgcaaatagtagataat 60
  (M) V W S V Q I V D N
10
61 gcagggttaggagcctaacttagctctctatccatccggaattotccactgtccctcgc 120
  A G L G A N L A L Y P S G N S S T V P R
20
121 tatgttactgtgacggctatgctcctatcacttttagcgaattggaccaagacagtt 180
  Y V T V T G Y A P I T F S E I G P K T V
30
181 caccaaagctggtacataactgtgcacaatggagatgacagagcattccaactggatat 240
  H Q S W Y I T V H N G D D R A F Q L G Y
40
241 gaaggaggaggtgtagcaacagcaacttttacagcaggtgggaatgtgagtattagcact 300
  E G G G V A T A T F T A G G N V S I S T
50
301 ggggttgggagtcacagcatttgacctcaagaactgcaaaatttttccacaaattgt 360
  G F G D A Q H L T L K K L A
60
361 ggcacataaataattttgttttaacccttaaaaaaaaaaaaaaaaaaaaaa 408
  
```

Fig. 1. The cDNA nucleotide sequence of *Physarum* haemagglutinin I and its deduced amino acid sequence. Nucleotide numbers are shown on both sides of each line. The amino acid sequence is shown below the nucleotide sequence. The number of amino acid residues starting from Val is given below the amino acid sequence. The N-terminal 65 aa sequence determined by Edman degradation is overlined. The box and double underlining denote the termination codon and the polyadenylation signal, respectively.

Fig. 2. Northern blot analysis of plasmodial RNA. Plasmodial RNA (15 µg) was electrophoresed on a formaldehyde-agarose gel, transferred to a nylon membrane by capillary action and hybridized with the fluorescein-labelled cDNA insert.

ping regions and spanned a cDNA sequence of 408 bp (Fig. 1). Hybridization of total RNA with a fluorescein-labelled nucleotide oligomer from a cloned cDNA is shown in Fig. 2. The oligomer hybridized with a single RNA band of approximately 450 nt (Fig. 2). The size of the RNA was consistent with the length of the cloned cDNA.

cDNA sequence and deduced amino acid sequence

The complete nucleotide sequence of the full-length cDNA contains 408 nt with a protein-coding region of

Fig. 3. Hydropathy profile (bottom) and prediction of α -helices, β -sheets and β -turns (top) in the amino acid sequence of haemagglutinin I using the computer programs of Kyte & Doolittle (1982) and Chou & Fasman (1978), respectively. Data plotted above and below the vertical line indicate hydrophobic and hydrophilic portions, respectively. Predicted β -sheets and β -turns are shown by diagonal-lined and solid black boxes, respectively. Open boxes at the C terminus indicate the predicted α -helices.

315 nt, beginning with an initiation codon, ATG, at positions 31–33 and ending with a stop codon, TAA, at positions 343–345 (Fig. 1). The fragment has a 5' non-coding region, as well as a 3' non-coding region before the poly(A) tail containing a typical polyadenylation signal, AATAAA, at positions 366–371. The ORF specified a protein whose amino acid sequence matched that determined by Edman degradation (Table 2). We therefore identified this gene as that of haemagglutinin I. The 315 bp ORF encodes a polypeptide of 104 aa without a typical signal sequence. Lack of a secretory signal peptide in haemagglutinin I might be evidence that it is a cytosolic protein (Morita & Nishi, 1992). The N-terminal amino acid sequence from Edman degradation coincided with the deduced sequence starting from valine at the second residue, indicating the removal of the N-terminal methionine residue. ESI-MS analysis revealed that haemagglutinin I has an apparent molecular mass of 10759.86 ± 0.15 Da, corresponding exactly to that predicted from the cDNA sequence (10760.76 Da). This suggests that haemagglutinin I is not post-translationally modified, except for the removal of the initiation methionine. Therefore, two consensus sequences (Asn-X-Ser/Thr) for the attachment of asparagine-linked oligosaccharide at amino acid residues 23–25 and 84–86 (Fig. 1) will be not glycosylated. The appearance of a phenylthiohydantoin-amino acid peak in the normal cycle by Edman degradation provides additional evidence for the lack of post-translational modification. The theoretical isoelectric point (pI) of the predicted amino acid sequence of haemagglutinin I is 6.32, consistent with the value of 6.9 determined by two-dimensional IEF/SDS-PAGE (data not shown). Haema-

Fig. 4. Expression of recombinant haemagglutinin I in *E. coli*. Protein samples were analysed by Tricine-SDS-PAGE, followed by staining with Coomassie brilliant blue R-250. Lanes: 1, haemagglutinin I (lower) and II (upper) purified from plasmodia; 2, *E. coli* cell lysate before induction; 3, *E. coli* cell lysate after induction with 0.4 mM IPTG; 4, purified recombinant haemagglutinin I. Positions of molecular mass markers are indicated. Molecular mass markers (R) used were phosphorylase b (97400 kDa), bovine serum albumin (66200), ovalbumin (45000), carbonic anhydrase (31000), soybean trypsin inhibitor (21500) and lysozyme (14400). Myoglobin fragments (r) were also run on the same gel.

gglutinin I is rich in hydrophobic amino acid residues such as glycine (13.59%), alanine (9.71%) and valine (9.71%) but no cysteine or methionine residues are found in the amino acid composition. A predominantly hydrophilic core of haemagglutinin I was observed at residues 42–70 (Fig. 3). The predicted secondary structure, analysed by the methods of Chou & Fasman (1978), suggests that haemagglutinin I is mainly built up of β -sheets and two α -helices (Fig. 3). Upon Tricine-SDS-PAGE, haemagglutinin I gave a single band with a relative molecular mass of about 6 kDa, though ESI-MS analysis revealed a mass of 10759.86 Da. It is unlikely that the protein was digested by proteolytic enzymes during gel electrophoresis because the migration position was not changed in the presence of various proteinase inhibitors. The reason for the anomalously high electrophoretic mobility of haemagglutinin I is not clear at present.

The amino acid sequence of haemagglutinin I was subjected to comparative analysis with the entire SWISS-PROT and NBRF sequence databases and the translated GenBank database for possible homologies between haemagglutinin I and other known proteins, including lectins. No significant sequence similarity was found between haemagglutinin I and other plant, animal and fungal lectins, and the bulk of other proteins examined.

Expression of haemagglutinin I in *E. coli*

Expression of haemagglutinin I in *E. coli* was performed using a bacteriophage T7 RNA polymerase/promoter system by induction with IPTG. Fig. 4 shows the results of Tricine-SDS-PAGE analysis of lysates of *E. coli*

BL21(DE3) containing pET with haemagglutinin I cDNA before (lane 2) and after (lane 3) induction with IPTG. *E. coli* appeared to produce a recombinant protein (lane 3), the position of which corresponds to that of native haemagglutinin I (lower band in lane 1). The recombinant protein, purified using a thyroglobulin-immobilized Toyopearl column, was detected as a single band (lane 4). The recombinant protein with marked haemagglutinating activity was isolated as a soluble protein at a yield of 3 mg (litre culture)⁻¹. The haemagglutinating activity of the recombinant protein was not inhibited by simple sugars (at 0.2 M) and glycoproteins such as ovalbumin, ovomucoid and mucin (at 5 mg ml⁻¹). However, the activity was inhibited by thyroglobulin, orsomucoid and fetuin, the minimum amounts of which for inhibiting four haemagglutinating doses were 0.24, 0.73 and 1.61 mg ml⁻¹, respectively. The inhibitory activity of these glycoproteins on recombinant haemagglutinin I is very similar to that on natural haemagglutinin I but not haemagglutinin II (Morita & Nishi, 1992).

DISCUSSION

In this study we have cloned a cDNA encoding haemagglutinin I and have obtained over 0.4 kbp of sequence information, spanning the protein-coding region and the 5' and 3' untranslated regions. Haemagglutinin I consists of 103 aa with a molecular mass of 10760.76 Da and the primary structure indicates the lack of a typical signal sequence. Expression of cDNA in *E. coli* revealed that the recombinant protein has a molecular mass and a sugar-binding specificity similar to that of native haemagglutinin I. This result indicates that the cDNA clone obtained is truly that for haemagglutinin I and post-translational modifications such as N-glycosylation (Gowda *et al.*, 1994), C-terminal proteolytic cleavage (Young *et al.*, 1995) or N-terminal acetylation (Hirabayashi & Kasai, 1993) are not essential for the haemagglutinating activity. Lack of post-translational modification was demonstrated by comparing the molecular masses deduced from the cDNA with those determined by ESI-MS.

Computer analysis revealed that haemagglutinin I has not been classified into any lectin family such as animal C-type lectin, galectin and plant legume lectin. However, a comparison of the amino acid composition and the secondary structure of haemagglutinin I with those of other lectins revealed some similar features. These include the abundance of glycine, alanine, aspartic acid/asparagine, threonine and valine residues, the absence or low number of methionine, cysteine and arginine residues and the presence of an extensive network of β -sheets. The small number of sulphur-containing amino acids has been reported in plant legume lectins (Sharon, 1993), fungal lectins (Rosen *et al.*, 1992; Vranken *et al.*, 1987; Yatohgo *et al.*, 1988) and bacterial lectins (Avichezer *et al.*, 1992). The lack of cysteine residues in haemagglutinin I supports the result that the haemagglutinating activity is not affected by

thiol reagents (Morita & Nishi, 1992). The major secondary structure in soluble plant lectins (Van Damme *et al.*, 1996; Young *et al.*, 1989) and animal galectins (Lobsanov *et al.*, 1993; Muramoto & Kamiya, 1992) is the β -sheet. Similarly, the presence of an extensive network of β -sheets bound by α -helices has been shown in both *Pseudomonas aeruginosa* lectin (Avichezer *et al.*, 1994) and *Arthrotrix oligospora* lectin (Rosen *et al.*, 1996b). These features were also found in the predicted secondary structure of haemagglutinin I (Fig. 3).

Many saline-soluble lectins with a relatively low molecular mass and low affinity to monosaccharides have been found in several fungal species. Mushroom lectins in *Flammulina velutipes* (Yatohgo *et al.*, 1988) and *Coprinus cinereus* (Cooper *et al.*, 1997) are developmentally regulated with high levels of expression in fruiting bodies and low levels of expression in vegetative mycelia, suggesting that these lectins may function in fruiting body formation. In *P. polycephalum*, however, it is unlikely that haemagglutinin I plays a role in the differentiation of plasmodia because acceleration of haemagglutinin synthesis was not always observed during macrocyst formation and fruiting body formation (Morita *et al.*, 1995). Discoidin I, a lectin in the cellular slime mould *D. discoideum*, has an RGD sequence which participates in cell-substratum attachment (Barondes & Springer, 1987). The RGD sequence, a cell-binding signal common to eukaryotic cell adhesion proteins, has also been found in several lectins (Ozeki *et al.*, 1991; Relman *et al.*, 1989). However, in spite of the developmental relationship to *D. discoideum*, no RGD sequence was found in *Physarum* haemagglutinin I. We therefore think that haemagglutinin I fulfils other important roles in this particular stage of the life cycle of this slime mould. As shown for galectin (Barondes *et al.*, 1994), several fungal lectins (Cooper *et al.*, 1997; Rosen *et al.*, 1996b) have been reported to accumulate extracellularly despite the lack of a classical secretion signal sequence. The lack of a typical signal sequence in haemagglutinin I is notable because previous studies have indicated that haemagglutinin I is present in the slime layer that exists on the surface of the plasmodia (Morita & Nishi, 1991). It is likely that haemagglutinin I is also externalized by non-classical mechanisms and functions in some physiological events by interaction with glycoconjugates on the plasmodial surface. Because of its preferential binding to acidic polysaccharide from the bacterial cell wall, we assume that the function of haemagglutinin may be to interact with other organisms (Morita *et al.*, 1995). In spite of several similarities, the amino acid sequence of haemagglutinin I is unrelated to those of fungal lectins and therefore represents a novel type of lectin.

In this study, we have expressed haemagglutinin I cDNA in *E. coli* BL21(DE3)pLysS using the T7 RNA polymerase expression system. Site-directed mutagenesis using this system will provide us with valuable information concerning lectin-carbohydrate recognition. In addition, the isolation of haemagglutinin I cDNA will facilitate the investigation of several aspects of lectin

function, including the regulation of its expression in growth and differentiation.

ACKNOWLEDGEMENTS

I thank Dr Shoji Okamura and Dr Teruhito Yasui for their helpful discussions and advice during the course of this investigation. I thank Naoko Shimada and Yukiko Iwado for their technical assistance.

REFERENCES

- Avichezer, D., Katcoff, D. J., Garber, N. C. & Gilboa-Garber, N. (1992). Analysis of the amino acid sequence of the *Pseudomonas aeruginosa* galactophilic PA-I lectin. *J Biol Chem* **267**, 23023–23027.
- Avichezer, D., Gilboa-Garber, N., Garber, N. C. & Katcoff, D. J. (1994). *Pseudomonas aeruginosa* PA-I lectin gene molecular analysis and expression in *Escherichia coli*. *Biochim Biophys Acta* **1218**, 11–20.
- Barondes, S. H. & Springer, W. R. (1987). An endogenous lectin and an oligosaccharide participate in adhesion mechanisms in *Dictyostelium*. In *Genetic Regulation of Development*, pp. 129–140. Edited by W. F. Loomis. New York: Alan R. Liss.
- Barondes, S. H., Cooper, D. N. W., Gitt, M. A. & Leffler, H. (1994). Galectins. Structure and function of a large family of animal lectins. *J Biol Chem* **269**, 20807–20810.
- Chou, P. Y. & Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol* **47**, 45–148.
- Cooper, D. N. W., Lee, S.-C. & Barondes, S. H. (1983). Discoidin-binding polysaccharide from *Dictyostelium discoideum*. *J Biol Chem* **258**, 8745–8750.
- Cooper, D. N. W., Boulianne, R. P., Charlton, S., Farrell, E. M., Sucher, A. & Lu, B. C. (1997). Fungal galectins, sequence and specificity of two isolectins from *Coprinus cinereus*. *J Biol Chem* **272**, 1514–1521.
- Daniel, J. W. & Baldwin, H. H. (1964). Axenic culture of *Physarum polycephalum* on soluble media. *Methods Cell Physiol* **1**, 15–29.
- Frohman, M. A., Dush, M. K. & Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* **85**, 8998–9002.
- Gowda, L. R., Savithri, H. S. & Rao, D. R. (1994). The complete primary structure of a unique mannose/glucose-specific lectin from Field Bean (*Dolichos Lab lab*). *J Biol Chem* **269**, 18789–18793.
- Hirabayashi, J. & Kasai, K. (1993). The family of metazoan metal-independent beta-galactoside-binding lectins: structure, function and molecular evolution. *Glycobiology* **3**, 297–304.
- Inbar, J. & Chet, I. (1994). A newly isolated lectin from the plant pathogenic fungus *Sclerotium rolfsii*: purification, characterization and role in mycoparasitism. *Microbiology* **140**, 651–657.
- Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132.
- Lobsanov, Y. D., Gitt, M. A., Leffler, H., Barondes, S. H. & Rini, J. M. (1993). X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9-Å resolution. *J Biol Chem* **268**, 27034–27038.
- Matsudaira, P. (1987). Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J Biol Chem* **262**, 10035–10038.

- Morita, M. & Nishi, A. (1991). Glycoproteins and enzymes associated with plasmodial membrane and slime layer of *Physarum polycephalum*. *J Gen Appl Microbiol* 37, 93–109.
- Morita, M. & Nishi, A. (1992). Isolation and partial characterization of haemagglutinins from plasmodia of *Physarum polycephalum*. *J Gen Microbiol* 138, 619–625.
- Morita, M., Minami, E., Zheng, L. & Nishi, A. (1995). *Physarum polycephalum* haemagglutinins: effect of nutrition on synthesis, and their possible role in nature. *Microbiology* 141, 2315–2322.
- Muramoto, K. & Kamiya, H. (1992). The amino-acid sequence of a lectin from conger eel, *Conger myriaster*, skin mucus. *Biochim Biophys Acta* 1116, 129–136.
- Nishimura, M., Tanigaki, C. & Okamura, S. (1991). Changes in β -tubulin isoforms and their RNA level in synchronized tobacco cells. *Cell Struct Funct* 16, 489–494.
- Ozeki, Y., Matsui, T., Suzuki, M. & Titani, K. (1991). Amino acid sequence and molecular characterization of a D-galactoside-specific lectin purified from sea urchin (*Anthocidaris crassispina*) eggs. *Biochemistry* 30, 2391–2394.
- Relman, D. A., Domenighini, M., Tomanen, E., Rappuoli, R. & Falkow, S. (1989). Filamentous hemagglutinin of *Bordetella pertussis*: nucleotide sequence and crucial role in adherence. *Proc Natl Acad Sci USA* 86, 2637–2641.
- Rosen, S., Ek, L., Rask, L. & Tunlid, A. (1992). Purification and characterization of a surface lectin from the nematode-trapping fungus *Arthrobotrys oligospora*. *J Gen Microbiol* 138, 2663–2672.
- Rosen, S., Bergstrom, J., Karlsson, K.-A. & Tunlid, A. (1996a). A multispecific saline-soluble lectin from the parasitic fungus *Arthrobotrys oligospora*. Similarities in the binding specificities compared with a lectin from the mushroom *Agaricus bisporus*. *Eur J Biochem* 238, 830–837.
- Rosen, S., Kata, M., Persson, Y., Lipniunas, P. H., Wikstrom, M., van den Hondel, C. A. M. J. J., van den Brink, J. M., Rask, L., Heden, L.-O. & Tunlid, A. (1996b). Molecular characterization of a saline-soluble lectin from a parasitic fungus. Extensive sequence similarities between fungal lectins. *Eur J Biochem* 238, 822–829.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74, 5463–5467.
- Schagger, H. & van Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* 166, 368–379.
- Schreckenbach, T. & Werenskiold, A.-K. (1986). Gene expression during plasmodial differentiation. In *The Molecular Biology of Physarum polycephalum*, pp. 131–150. Edited by W. F. Dove, J. Dee, S. Hatano, F. B. Haugli & K.-E. Wohlfarth-Bottermann. New York: Plenum Press.
- Sharon, N. (1993). Lectin-carbohydrate complexes of plants and animals: an atomic view. *Trends Biochem Sci* 18, 221–226.
- Van Damme, E. J. M., Barre, A., Verhaert, P., Rouge, P. & Peumans, W. J. (1996). Molecular cloning of the mitogenic mannose/maltose-specific rhizome lectin from *Calystegia sepium*. *FEBS Lett* 397, 352–356.
- Vranken, A. M., Van Damme, E. J. M., Allen, A. K. & Peumans, W. J. (1987). Purification and properties of an N-acetylgalactosamine specific lectin from the plant pathogenic fungus *Rhizoctonia solani*. *FEBS Lett* 216, 67–72.
- Yatohgo, T., Nakata, M., Tumuraya, Y., Hashimoto, Y. & Yamamoto, S. (1988). Purification and properties of a lectin from the fruitbodies of *Flammulina velutipes*. *Agric Biol Chem* 52, 1485–1493.
- Young, N. M., Johnston, R. A. Z., Szabo, A. G. & Watson, D. C. (1989). Homology of the D-galactose-specific lectins from *Artocarpus integrifolia* and *Maclura pomifera* and the role of an unusual small polypeptide subunit. *Arch Biochem Biophys* 270, 596–603.
- Young, N. M., Watson, D. C., Yaguchi, M., Ader, R., Arango, R., Rodriguez-Arango, E., Sharon, N., Blay, P. K. S. & Thibault, P. (1995). C-Terminal post-translational proteolysis of plant lectins and their recombinant forms expressed in *Escherichia coli*. *J Biol Chem* 270, 2563–2570.

Received 5 September 1997; revised 6 November 1997; accepted 24 December 1997.