Cloning, Expression, and Characterization of the Galα1,3Gal High Affinity Lectin from the Mushroom *Marasmius oreades*

Received for publication, January 7, 2002, and in revised form, February 6, 2002
Published, JBC Papers in Press, February 8, 2002, DOI 10.1074/jbc.M200165200

Robert P. Kruger‡‡, Harry C. Winter¶, Nancy Simonson-Leff, Jeanne A. Stuckey***, Irwin J. Goldstein‡‡, and Jack E. Dixon§§

From the Department of Biological Chemistry, ‡‡Neuroscience Program, ¶Biophysics Research Division, and §§Life Sciences Institute, University of Michigan, Ann Arbor, Michigan 48109-0606

The purification and unique carbohydrate binding properties, including blood group B-specific agglutination and preferential binding to Galα1,3Gal-containing sugar epitopes, of the *Marasmius oreades* agglutinin (MOA) are reported in an accompanying paper (Winter, H. C., Mostafapour, K., and Goldstein, I. J. (2002) *J. Biol. Chem.* 277, 14996–15001). Here we describe the cloning, characterization, and expression of MOA. MOA was digested with trypsin and endoprotease Asp-N, and the peptide fragments were purified by high performance liquid chromatography. Amino acid sequence data were obtained for eight peptides. Using oligonucleotides deduced from the peptide sequences for a reverse transcriptase-PCR, a 41-base pair cDNA was obtained. The 41-base pair fragment allowed the generation a full-length cDNA using 5′ rapid amplification of cDNA ends. MOA cDNA encodes a protein of 293 amino acids that contains a ricin domain. These carbohydrate binding domains were first described in subunits of bacterial toxins and are also commonly found in polysaccharide-degrading enzymes. Whereas these proteins are known to display a variety of sugar binding specificities, none to date are known to share MOA’s high affinity for Galα1,3Gal and Galα1,3Galβ1,4GlcNAc. Recombinantly expressed and purified MOA retains the specificity and affinity observed with the native protein. This study provides the basis for analyzing the underlying cause for the unusual binding specificity of MOA.

Specificity varies greatly among carbohydrate-binding proteins. Whereas some lectins broadly recognize all oligosaccharides containing particular terminal sugars, others show increasing affinity for specific di- and trisaccharides. Fewer still show almost no reactivity with a given sugar monomer yet bind strongly and specifically to particular oligosaccharides (1). As described in the accompanying paper (2), the *Marasmius oreades* agglutinin (MOA) binds to Galα1,3Gal-containing sugars and falls into this last category.

The Galα1,3Gal epitope has received considerable attention, stemming from its presence in the glycoproteins of most mammals and its conspicuous absence in humans, apes, and Old World monkeys (3). This absence is attributable to lack of the specific α1,3-galactosyltransferase because of frameshift mutations in its gene (4). The resulting immunogenicity of the Galα1,3Gal epitope is a significant barrier to xenotransplantation (5). Despite the importance of Galα1,3Gal epitope recognition, MOA is currently the only lectin known to have exclusive specificity for this disaccharide (2).

Few proteins have been shown to bind with any specificity to Galα1,3Gal. While *Clostridium difficile* toxin A and antibodies recognizing the α-galactosyl epitope both bind well to some Galα1,3Gal-containing oligosaccharides (6), the size and species of origin of MOA suggest that it is fundamentally dissimilar to these proteins. On these grounds, the blood group B-specific *Griffonia simplicifolia* I-B₂ isolecitin is perhaps more appropriate for comparison (7). A recent x-ray crystallographic structural analysis of *G. simplicifolia* I-B₂ isolecitin complexed with Galα1,3Gal revealed that its binding pocket is restricted to the terminal nonreducing sugar, consistent with data showing the lectin to have similar affinity for monosaccharide and the various positional isomers of the disaccharide (8). However, MOA is expected to have an extended binding site to explain its overwhelming preference for Galα1,3Gal-containing di- and trisaccharides.

To study the basis for its unique carbohydrate binding specificity, we have cloned, recombinantly expressed, and characterized MOA. These studies reveal that MOA is a member of the ricin superfamily.

**EXPERIMENTAL PROCEDURES**

Peptide Sequencing and Analysis—Peptide sequences were determined by the Macromolecular Structure Facility at Michigan State University. Briefly, purified protein was digested with trypsin or endoprotease Asp-N. Proteolytic fragments were bound to a C-18 column and eluted with a gradient of acetonitrile. Purified peptides were then sequenced by automated Edman degradation.

RNA Isolation, cDNA Cloning, and Northern Analysis—*M. oreades* mushrooms were collected in grassy plots in Ann Arbor, Michigan, frozen immediately in dry ice, and stored at −80 °C until extracted. The frozen tissue was ground under liquid nitrogen to a medium-fine powder with a mortar and pestle resting in dry ice. Subsequent steps in the RNA purification followed recommendations given with the Plant RNA Isolation Aid as an accessory to the RNAqueous-Midi kit (Ambion). Using this protocol, 7.2 μg of total RNA/μg of mushroom was isolated. Oligonucleotides for RT-PCR were designed from the available peptide sequences. The two regions with lowest degeneracy were within a region of four peptides whose sequences overlap one another. The forward primer (5′-GGNTGCGARTYACNC-C-3′) was reverse transcribed by rapid amplification of cDNA ends; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight.
lated from the amino acid sequence GWQFTP. The reverse primer (5'ARYTGRTGCCARTTDAT-3') is the reverse complement of the reverse translated amino acid sequence INWHQL. The degeneracies of the forward and reverse primers are 64- and 48-fold, respectively.

RT-PCR was conducted with Moloney murine leukemia virus reverse transcriptase (Invitrogen) and AmpliTaq Gold polymerase (Applied Biosystems). Template mRNA was purified from 0.86 µg of total RNA using the mRNA capture kit (Roche Molecular Biochemicals). A PCR product of appropriate size (~40 bp) was cloned using the TA TOPO PCR cloning kit (Invitrogen). Sequencing of this product yielded a total of 11 unambiguous bases.

5' and 3' RACE was performed essentially as described in the First-Choice RLM-RACE kit (Ambion). Two overlapping primers were designed for each 5' and 3' RACE. These primers include all or part of the 11 unambiguous bases. The two primers used for subsequent amplification steps in 5' RACE (primer 1, 5'-ARYTGRTGCCARTTDAT-3'; primer 2, 5'-TGGCARTRATGGTCTGG-3') are 32- and 4-fold degenerate, respectively, with the degeneracy weighted toward the 5'-ends. Similarly, the two primers used for 3' RACE (primer 1, 5'-GNGTGGCARTTYACRCCAGA-3'; primer 2, 5'-CARTTYACRCCACAGCAG-3') are 32- and 8-fold degenerate, respectively.

For Northern analysis, total RNA (10 µg) was run on a prerun formamide gel and transferred to a nylon membrane (Nytran) with the Bios blotting system. The cDNA probe was generated by random primer labeling with Klenow (Roche Molecular Biochemicals) incorporating 

| Peptide | Sequence |
|---------|----------|
| 1       | DGSSSDTPIVG |
| 2       | DGTPIVGQFTP |
| 3       | QTPTFDINNH |
| 4       | JTNWHQLL |
| 5       | EAQAAIAARNPHHTGYRGY |
| 6       | M/G/A/ADSXK |
| 7       | DLYNGSEAEATVNG |
| 8       | IAAGXGTG |

**TABLE I**

Sequences of *M. oreades* lectin peptides

| Peptide | Sequence |
|---------|----------|
| 1       | DGSSSDTPIVG |
| 2       | DGTPIVGQFTP |
| 3       | QTPTFDINNH |
| 4       | JTNWHQLL |
| 5       | EAQAAIAARNPHHTGYRGY |
| 6       | M/G/A/ADSXK |
| 7       | DLYNGSEAEATVNG |
| 8       | IAAGXGTG |

**FIG. 1.** MOA cloning strategy. Relative position of degenerate primers with respect to derived sequence from the overlapping peptides 1–4 is shown by arrows labeled *F* and *R*. RT-PCR from *M. oreades* total RNA yields a product of the expected size. Subsequent cloning and sequencing confirm the size of the product (41 bp) and show it to encode the intervening amino acids. Sequencing of this product yielded 11 nondegenerate bases. Primers utilizing this nondegenerate sequence were used for 5' and 3' RACE to generate a full-length sequence.

**FIG. 2.** Analysis of the primary amino acid sequence for the MOA. A, the deduced full-length coding sequence is shown. The underlined regions denote the positions of the MOA peptide sequences (Table I). The boxed region shares homology with the ricin B chain lectin domain. B, alignment of MOA with an assortment of bacterial toxins and carbohydrate-degrading enzymes indicates the presence of three conserved Q(X/W/F) motifs (each identified with a line and number). Amino acids identical to the consensus sequence are boxed and shaded. Amino acids similar to the consensus are shaded lightly. Numbers indicate the amino acid positions of the selected full-length sequences.

**FIG. 3.** SDS-PAGE of recombinant MOA. Gel was 12.5% (0.8% cross-linked). Lane 1, standard protein ladder; lane 2, native MOA; lane 3, affinity-purified recombinant MOA; lane 4, cell lysate of *E. coli* cells expressing recombinant MOA.
Comparison of binding of oligosaccharides by native and recombinant M. oreades agglutinin

| Ligand       | Cytosolic MOA | Native MOA | Recombinant MOA |
|--------------|---------------|------------|----------------|
| LacNac       | 0.29          | 0.38       | 0.67           |
| Galα1,3Gal   | 5.48          | 27.3       | 26.7           |
| Galα1,3Galβ1,4GlcNac | 9.75          | 39.4       | 30.3           |
| Galα1,3Galβ1,4Glc | ND a          | 16.7       | 13.2           |
| Galα1,3 (Fucα1,2) Gal | 27.8          | ND         | 46.1           |

a Data from Ref. 2.

**FIG. 4.** Alignment of MOA subdomains with the ricin and ebulin subdomains that bind galactose. Key residues that hydrogen-bond with the galactosyl third and fourth oxygens are denoted with Δ. The key sugar stacking residue is denoted with a plus sign. The loop region referenced under “Results and Discussion” is also labeled. Residues that are identical to the consensus sequence are boxed and shaded. Residues that are similar to the consensus are shaded. Numbers indicate the amino acid positions of the selected full-length sequences.

[32P]dATP (PerkinElmer Life Sciences). The template used was a full-length coding sequence PCR product. The blot was probed and washed according to the membrane manufacturer’s protocol and exposed to film for 2 h.

Expression and Characterization of Recombinant MOA—A full-length coding sequence PCR product incorporating NdeI and EcoRI sites into its forward and reverse primers, respectively, was cloned into PCR Blunt II using topoisomerase (Invitrogen) and subsequently subcloned into an isopropyl-1-thio-β-d-galactopyranoside-inducible pT7 expression vector (MOAP/T7LO). Recombinant MOA was expressed in a Nova Blue DE3 strain of *Escherichia coli*. Induced bacteria were collected by centrifugation and resuspended in a lysis buffer consisting of 50 mM NaH2PO₄, 300 mM NaCl, 10 mM imidazole, 10 mM 2-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, 1% Nonidet P-40, and a protease inhibitor mixture. The extract was run twice through a French press. The insoluble fraction was removed by centrifugation (10,000 × g, 15 min).

**RESULTS AND DISCUSSION**

Here we report the deduced amino acid sequence and recombinant expression of the only known Galα1,3Gal-specific lectin. This is the first recorded protein sequence from the fairy ring mushroom *M. oreades*.

Enzymatic digestion, purification of peptide fragments, and Edman degradation of the native protein yielded eight peptide sequences (Table I). Inspection of the peptides reveals that four have overlapping amino acid sequences, designated peptides 1–4. The two low degeneracy oligonucleotides used for RT-PCR were designed from the overlapping region (Fig. 1). These oligonucleotides were used to obtain a 41-base pair product whose sequence generated 11 unambiguous bases. The 11 unambiguous nucleotides proved a sufficient starting point for the generation of a full-length sequence via 5’ and 3’ RACE. Cloning and sequencing of 5’ and 3’ RACE products generated 169 and 881 bp of 5’ and 3’ sequence, predicting a total message size, not including polyadenylation, of 1062 bp (GenBank™ accession number AY06613). This corresponds well with Northern analysis showing a major band at ~1.5 kb and a minor band at ~1.1 kb (data not shown). Sequencing of multiple clones revealed that the mRNA apparently contains four nucleotide polymorphisms, only one of which confers an amino acid ambiguity. Specifically, position 200 can be either aspartic acid or asparagine (Fig. 2A). This polymorphism seems unlikely to alter binding specificity, since it lies outside of the predicted ricin domain (discussed below).

Analysis of the cDNA indicates an open reading frame encoding a protein of 293 amino acids (Fig. 2A). Inspection of the predicted amino acid sequence shows the presence of all eight of the sequenced peptides. MOA also apparently lacks a signal peptide and is therefore probably cytosolic. The predicted molecular weight of this protein is 32,299. This is consistent with analysis of the native protein by MALDI-TOF mass spectrometric analysis giving a molecular mass for the full-length native protein of approximately 32,290. Mass spectrometric analysis of the native protein and tryptic digests thereof showed remarkable correlation between the observed molecular weights and those predicted from the deduced amino acid sequence (data not shown). This strongly suggests that the isolated lectin and the cloned cDNA product are the same protein.

The MOA open reading frame was cloned into a T7 expression vector. The protein was produced in *E. coli* and purified as described above. Recombinant MOA had an electrophoretic mobility in SDS-PAGE identical to that of native protein at 32 kDa (Fig. 3) and eluted as a single, symmetrical peak at the same elution volume as native MOA from a G2000 SWXL molecular sieve column (not shown; see Ref. 2). Moreover, polyclonal rabbit antiserum prepared against either the native MOA or the recombinant MOA formed precipitin bands of identity with the native and recombinant MOA preparations. The recombinant protein and native protein were also subjected to MALDI-TOF mass spectrometry. Recombinant MOA showed a molecular mass of 32,090 ± 20 Da, whereas the native protein had a slightly higher mass, 32,132 ± 17 Da. Both the native and recombinant proteins were found to contain less than 0.25 mol of neutral sugar/mol of 32-kDa protein by the phenol-sulfuric acid assay (9), provided that the proteins were purified by absorption to Synsorb B and elution at high pH with diaminopropane, a procedure not involving elution with
MOA, a Ricin Superfamily Galα,3Gal-binding Lectin

sugar. Similarly, neither protein was stained on SDS-PAGE gels by the periodate-Schiff stain. Since the native protein appears to be blocked at the N terminus, the difference in the molecular mass of the native versus recombinant proteins might be caused by the presence of a blocking group, such as an N-acetyl moiety (M = 42 Da) on the native protein.

Binding constants of several relevant oligosaccharides to recombinant and intact native MOA were determined calorimetrically. As shown in Table II, little or no difference was observed between the two preparations. The original isolation of MOA had produced protein containing a mixture of full-length (32 kDa) and “clipped” protein (23- and 10-kDa fragments). Not surprisingly, both intact MOA and recombinant MOA show slightly stronger binding to Galα1,3-linked oligosaccharides than the “clipped” form of the native lectin.

Blast searches of the MOA sequence showed highest similarity to the ricin domains of a xylanase-arabinofuranosidase from Streptomyces chattanoogensis (NCBI accession number AAD32559), a β-mannanase from Polysaccharium cellulose (NCBI accession number AAK19890), the α chain of coagulation factor G from horseshoe crab (NCBI accession number BAA04044), and mosquitocidal toxin 21 from Bacillus sphaericus (NCBI accession number S27514). The presence of a ricin domain is best shown by the alignment of MOA with this subset of ricin domain-containing proteins (Fig. 2B). Other ricin domain-containing proteins showed less identity with MOA. Outside of the prospective ricin domain, however, no convincing homology was observed with these or any other known proteins. Many of the ricin domain-containing proteins, like the ricin B chain itself, promote the internalization of disulfide-linked toxic protemers through their binding to glycosylated cell surface receptors (10); however, there is no evidence that MOA functions in this manner.

Structural analysis of ricin domains suggests that they are composed of three repeating subdomains that may have originated from an ancestral galactose-binding motif (11). Close analysis of the three subdomains of MOA indicates strong conservation with the key residues in the 1α and 2γ subdomains of ricin and ebulin (Fig. 4). Structural determination of these proteins in the presence of sugar shows binding to these two subdomains (11, 12). All of the MOA subdomains have the conserved QXX motif (13). In ricin, the conserved tryptophan is necessary for hydrophobic packing of the core structure, whereas the glutamine coordinates the conserved aspartic acid that hydrogen-bonds with the third and fourth oxygens of the galactosyl moiety. The asparagine prior to the QXX motif also hydrogen-bonds with the O-3 and O-4 of the sugar. The corresponding histidine found in the MOA subdomains could function similarly. Additionally, there is a conserved hydrophobic position occupied by tryptophan, tyrosine, or phenylalanine between the conserved aspartic acid and asparagine. This residue forms a stacking interaction with the sugar ring. In the MOA subdomains, this position is also occupied by a tryptophan.

Because the essential features required for galactosyl binding are conserved in MOA, it is interesting that the specificity of ricin is very different from that of MOA. While MOA is specific for Galα1,3Gal-containing sugars, ricin binds well with β-1,3- or β-1,4-linked galactose-terminated sugars (14). Like MOA, ricin shows higher affinity for larger, more complex saccharides than for simple sugars (2, 15). The affinity constant for lactose binding to ricin is 10-fold greater than for galactose alone (15). Similarly, MOA binds Galα1,3Gal with an affinity constant 44-fold greater than that for Meα1,3-Gal (2). While the structure of ricin shows hydrogen bonding exclusively to the terminal sugar, it is clear that elements outside of the main binding pocket are important for determining the strength and specificity of binding.

Of particular interest in explaining the difference between MOA and other ricin domain proteins could be the loop region between the stacking hydrophobic residue and the sugar binding asparagine/histidine. Unlike other subdomain segments, this loop does not model well onto ricin. It is longer in MOA than in ebulin and ricin by 1–3 residues and appears structurally different, since it lacks a conserved proline following the hydrophobic stacking residue. This region could provide an additional hydrophobic stacking interface or hydrogen bonding specific for Galα1,3Gal-containing sugars either through direct side chain contact or water-mediated interactions and would be appropriately positioned to sterically block sugars not in the 1,3 orientation.

The cloning and expression of the recombinant MOA provides a route for understanding the structure and unique carbohydrate binding specificity of this novel lectin. Crystallographic structure determination of MOA in the presence of bound sugar should provide the rationale for the specific binding of Galα1,3Gal. This study also emphasizes the flexibility of the ricin domain in sugar binding specificity and suggests that the ricin superfamily will be a continuing source for the discovery of novel lectins that, like MOA, are specific in recognition for both sugar moiety and linkage.

REFERENCES

1. Weis, W. I., and Drickamer, K. (1996) Annu. Rev. Biochem. 65, 441–473
2. Winter, H. C., Mostafapour, K., and Goldstein, I. J. (2002) J. Biol. Chem. 277, 14996–15001
3. Galili, U., Clark, M. R., Shohet, S. B., Buchler, J., and Macher, B. A. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 1369–1373
4. Larsen, R. D., Rivera-Marrero, C. A., Ernst, L. K., Cummings, R. D., and Lowe, J. B. (1990) J. Biol. Chem. 265, 7055–7061
5. Sandrin, M. S., Vaughan, H. A., Babekowski, P. L., and McKenzie, I. F. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 11391–11395
6. Teneberg, S., Lonnroth, I., Torres Lopez, J. F., Galili, U., Halvarsson, M. O., Angstrom, J., and Karlsson, K.-A. (1996) Glycoconjugate J. 13, 599–609
7. Murphy, L. A., and Goldstein, I. J. (1979) Biochemistry 18, 4999–5005
8. Tempel, W., Tchampel, S., and Woods, R. (2002) J. Biol. Chem. 277, 6615–6621
9. Dubois, M. A., Gilles, K. A., Hamilton, J. K., Rebers, P. A., and Smith, F. (1956) Anal. Chem. 28, 350–356
10. Olesen, S., and Pihl, A. (1973) Biochemistry 12, 3121–3125
11. Montfort, W., Villafranca, J. E., Monzino, A. F., Ernst, S. R., Kaltzin, B., Rutner, E., Xiong, N. H., Hamlin, R., and Robertus, J. D. (1987) J. Biol. Chem. 262, 5396–5403
12. Pascal, J. M., Day, P. J., Monzino, A. F., Ernst, S. R., Robertus, J. D., Iglesias, R., Perez, Y., Ferreras, J. M., Citores, L., and Girbes, T. (2001) Protein Sci. 43, 319–326
13. Hazes, B. (1996) Protein Sci. 5, 1490–1501
14. Baenziger, J. U., and Fiete, D. (1979) J. Biol. Chem. 254, 9795–9799
15. Zentz, C., Frenoy, J. P., and Bourrillon, R. (1979) Biochimie (Paris) 61, 1–6