Structural mechanism governing the quaternary organization of monocot mannose-binding lectin revealed by the novel monomeric structure of an orchid lectin

Wei Liu1,§, Na Yang1,2,§, Jingjin Ding1,2, Ren-huai Huang1, Zhong Hu3 and Da-Cheng Wang1*

1 Center for Structural and Molecular Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, P. R. China

2 Graduate School of Chinese Academy of Sciences, Beijing 100039, P. R. China

3 Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, P. R. China

§ Both authors contributed equally

* To whom correspondence should be addressed. Tel.: +86-10-64888547; Fax: +86-10-64888560; E-mail: dewan@sun5.ibp.ac.cn / wangdcibp@yahoo.com.cn

Running title: Structural mechanism governing quaternary state of bulb lectin

Abstract

Two isoforms of an antifungal protein, gastrodianin, were isolated from two subspecies of the orchid *Gastrodia elata*, belonging to the protein superfamily of monocot mannose-specific lectins. In the context that all available structures in this superfamily are oligomers so far, the crystal structures of the orchid lectins both at 2.0 Å revealed a novel monomeric structure. It resulted from the rearrangement of the C-terminal peptide inclusive of the 12th β-strand, which changes from the “C-terminal exchange” into a “C-terminal self-assembly” mode. Thus, the overall tertiary scaffold is stabilized with an intramolecular β-sheet instead of the hybrid one observed on subunit/subunit interface in all known homologous dimeric or tetrameric lectins. In contrast to the constrained extended conformation with a cis peptide bond between residue 98 and 99 commonly occurred in oligomers, a β-hairpin forms from position 97-101 with a normal trans peptide bond at the corresponding site in gastrodianin, which determines the topology of the C-terminal peptide and thereby its unique fold pattern. Sequence and structure comparison shows that residue replacement and insertion at the position where the β-hairpin occurs in association with cis-trans interconversion of the specific peptide bond (97-98) are possibly responsible for such a radical structure switch between monomers and oligomers. Moreover, this seems to be a common melody controlling the quaternary states among bulb lectins through studies on sequence alignment. The observations revealed a structural mechanism by which the quaternary organization of monocot mannose-binding lectins could be governed. The mutation experiment performed on MBP-gastrodianin fusion protein followed by a few biochemical detections provides a direct evidence to support this conclusion. Potential carbohydrate recognition sites and biological implications of the orchid lectin based on its monomeric state are also discussed in this
Introduction

Plant lectins are defined as proteins possessing at least one non-catalytic domain which binds reversibly to a specific mono- or oligosaccharide (1). Owing to the increasing availability of sequence and structure information, the majority of all known plant lectins have been recently subdivided into seven structural and evolutional related groups (2). Among these classes are monocot mannose-binding lectins that preferably bind to 1-3 or 1-6 linked D-mannoses with the highest affinity (3). Lectins from this protein superfamily having been isolated and characterized so far come from Amaryllidaceae, Alliaceas, Araceae, Liliaceae, Orchidaceae and recently reported Iridaceae families (2,4). All these proteins are found to present 3 potential carbohydrate-binding motifs per subunit, each of which contains a consensus sequence signature QXDXNXVXY essential for mannose binding (5).

All currently known monocot mannose-binding lectins with available three-dimensional structures in PDB are hololectins built from two or four identical one-domain promoters. It has been widely believed that oligomerization and multivalency thereby play an important role in variability of carbohydrate recognition among plant lectins. For instance, tetrameric proteins such as snowdrop and daffodil lectins bind with mannan epitopes on GP120, the major glycoprotein of HIV, with high degree of affinity (6,7), whereas garlic agglutinin, a dimer, does not (8,9). Galanthus nivalis agglutinin (GNA) from snowdrop bulbs was the first member of this protein superfamily to have its crystal structure determined (10), and since then, a series tetrameric or dimeric structures of monocot lectins from amaryllis (Hippeastrum hybrid) (11), bluebell (Scilla...
Campanulata (12), daffodil (Narcissus pseudonarcissus) (13) and garlic (Allium sativum) (14,15) have been reported, as well as a two-domain lectin from Scilla campanulata (16).

The basic structural unit for bulb lectins is a dimer as revealed in those literatures. In all studied structures, two subunits are related by a pseudo 2-fold symmetry axis and assemble into a tightly bound dimer by exchanging their C-terminal β-strands (residues 101 to 105 in GNA) to form a hybrid β-sheet (10). This mode often referred to as “C-terminal exchange” confers a large buried area on subunit/subunit interface through which a stable dimer is established. An unusual cis peptide bond between Gly98 and Thr99 is commonly involved in those oligomers and believed to play a vital role for strand exchange (10,12,13). From energetic and structural point of view, monomers cannot exist stably when a dimer dissociates, as expected and verified calorimetrically (17). However, three merolectins containing single protomer have already been isolated and sequenced from Orchidaceae, i.e. Lister ovata and Epipactis helleborine (18,19) and Gastrodia elata (20). In this context, it would be of great interest to uncover unique structural features of these monomers different from those presented in oligomers, for the purpose of gaining deeper insight into structural mechanisms governing quaternary associations for monocot mannose-binding lectins.

Gastrodianin, also known as Gastrodia antifungal protein (GAFP) previously, is purified from nutritive corms of the orchid Gastrodia elata, a traditional Chinese medicinal herb having been cultured for thousands of years. The plant lacks chlorophyll and leads a parasitic life on the fungus Armillaria mellea. The fungal hyphae invade the nutritive and the primary corm of G. elata during its development. The cortical cells in corms, however, capture and digest the infecting hyphae and transport the released nutrients into a terminal corm for sustaining its growth (21).
Histochemical localization studies in vivo showed that a mannose-binding lectin, later named gastrodianin, accumulates in nutritive corms where fungal infection takes place (20,22). Antifungal assays in vitro also confirmed a strong inhibitive activity of gastrodianin towards a wide-range of phytopathogenic fungi, such as Trichoderma viride, Valsaambiens, Rhizoctonia solani, Gibberella zeae, Ganoderma lucidum and Botrytis cinerea (23). As a common phenomenon that a number of isolectins simultaneously occur in one plant, several isoforms have been identified from G. elata, differing subtly from each other in their sequence. At least five of them have been distinguished on the level of cDNA sequences (24-26). The mature protein is estimated 112 amino acid long and behaves as a monomer in solution (25). Sequence alignment (Fig. 1) indicates that gratrodianins belong to the superfamily of monocot lectins, and shares the highest degree of identity with two other orchid merolectins, LOMBP and EHMBP (18).

We reported here the native crystal structures of two gastrodianin isoforms purified from two subspecies of Gastrodia elata, named gastrodianin-1 and gastrodianin-4 respectively according to their cDNA numbers. Under the background that a plethora of attentions have been paid on oligomeric plant lectins, the refined structures of gastrodianin provide us a scientific glimpse on a monomeric mannose-binding lectin for the first time. It also reveals a potential structural mechanisms governing protein assembly in this superfamily, with possible links to their biological roles in the plant. Furthermore, the conclusion from the structural analysis was generally confirmed by mutagenesis experiments on recombinant MBP-gastrodianin fusion protein.
Materials and Methods

Purification and crystallization—Gastrodianin-1 and Gastrodianin-4 were isolated from newborn terminal corms of two subspecies of the orchids, namely *G. elata* B1 f. *elata* and *G. elata* B1 f. *glauca*, respectively. The purification and crystallization of gastrodianin-1 has been described previously (27), and an identical scheme was applied on gastrodianin-4. The only difference in crystallization trials between the two proteins was different additives being used: 2.5% of dioxane for gastrodianin-1, while 3~5% of MPD for gastrodianin-4. Both crystals underwent a long period of growth before shooting.

Data collection—Data collection, processing and preliminary analysis for gastrodianin-1 were carried out as described previously (27). A 2.0 Å data set for gastrodianin-4 was collected at room temperature on the beamline of BL-6B at Photon Factory in KEK, Tsukaba, Japan, with synchrotron radiation source. X-ray diffraction patterns were record on a Weissenberg camera. Evaluation and scaling of intensity data were processed using DENZO and SCALEPACK in HKL program package (28). The completeness of the whole data sets reached 95.3%.

Structure determination—Structure solution for gastrodianin-4 was first achieved by molecular replacement technique using the program AMoRe (29), with a search probe constructed on a truncated model of subunit C of GNA (PDB entry: 1MSA (10)) without the 11 residues at the C terminus. A reasonable R-factor of 40.8% and a correlation coefficient of 60.7% were given from the top solution after rigid body refinement. In the case of gastrodianin-1, four molecules were estimated in an asymmetric unit of the crystal (27) and the program MOLREP (30) in CCP4 program suite (31) was used instead of AMoRe with the refined model of gastrodianin-4 as the search probe. Correlation factor rose from 25.7% to 58.5% in four cycles, while R-factor went
down from 55.1% to 41.3%. Very few close contacts indicated the final combined coordinates from MOLREP were acceptable.

*Model building and refinement*—Although a truncated search model from a subunit of snowdrop lectin was used in molecular replacement, extra electron densities clearly present at the C-terminus, accommodating additional residues beyond residue 98. Proper residues in accordance with the sequence were manually fitted one by one into the density map with the program O (32), and several rounds of positional refinement were needed at intervals to improve map quality.

After the completion of model building at the C-terminus and the replacement of all various residues which differ between gastrodianin and GNA, the final models from molecular replacement for both orchid lectins were refined by simulated annealing at initial stages followed by energy minimization with CNS program suite (33), and with REFMAC (34) at the final round. Anisotropic scaling for initial B factor and solvent contribution was included in the process of refinement with both CNS and REFMAC. The thermal parameters were refined on individual atoms.

NCS restraints were applied between four monomers in gastrodianin-1 structure during first simulated annealing runs to prevent over-refinement, but with decreasing $W_i$ from cycle to cycle until last rounds of refinement when neither positional nor B-factor restraints were applied. This allowed for conformational variability among monomers. In both cases, solvent molecules as well as sulfate ions were incorporated when the polypeptides became stable, until no significant feature was remained in the difference map. The occupancy of sulfate ions was refined within a reasonable range of thermal factor.

The agreement between the atomic models and X-ray data was checked by SFCHECK (35),
and proper geometries were verified using PROCHECK (36).

Expression of MBP-gastrodianin fusion protein — Due to unavailability of soluble recombinant gastrodianin from various expression system including *E.coli* and yeast, oligomeric studies had to be performed with resort to fusion protein expression. The cDNA encoding gastrodianin-1 was obtained from topical corm of *Gastrodia elata* B1 elata using 3’-RACE as reported previously (24). Nucleotide sequence corresponding to the mature lectin from residue 1 to 112 was inserted into pMAL-p2 (New England Biolabs), downstream the gene encoding the maltose-binding protein (MBP), following the protocol described by Wang et al. (25). MBP-gastrodianin fusion protein was produced in *E. coli* host strain TB1 and purified through an affinity chromatography on a column packed with amylose resin (New England Biolabs) in the first step. The fraction of interest was subsequently concentrated with ular-filtration (Millipore) and then loaded onto a Superdex 75 10/300 GL column (Amersham Pharmacia) pre-equilibrated with 50 mM phosphate buffer, pH 7.0 and 0.15 M NaCl at room temperature.

Mutation experiments — In order to confirm the conclusion from the structure analysis, the mutation from D97-N98-S99-N100-N101 into G97-T98 on the loop sequence of gastrodianin gene was fulfilled by multiple-round PCR. Four mutagenic primers were synthesized for the purpose of stepwise nested amplification. A pair of primers, 5’-primer (5’-TCAGATCGGTTGAATTCGGAATCCGGG-3’) and one of 3’-primers (primer 1: 5’-ACGGTCAGTACCGTATATGACGAGTGTCGCCGGG-3’; primer 2: 5’-AGGTCAGTACCGTATATGACGAGTGTCGCCGGG-3’ and primer 3: 5’-GTGGATCCATTAATTTCCAAACGTGTGGGTGGGGCTTTCCAGGT-3’) was applied for each run in sequence, and the intermediate amplified product was used as the PCR template for the
following round after purification from agarose gel. DNA sequencing was performed to guarantee the correctness of the final mutant gene. The expression and purification of the mutant fusion protein were carried out with the same protocol as that for the wild type.

*Detection of oligomerization*—Oligomerization of the mutant MBP-gastrodianin fusion protein was detected by the following experiments with the wild type fusion protein used as the control. Firstly the wild and mutant proteins were comparatively analysed by gel filtration chromatography on HIlode 16/60 Superdex 75 column. The molecular size of wild and mutant fusion proteins were then measured by using Superdex 75 column (separation range from 3000 to 70000) and Superdex 200 10/300 HR column (separation range from 10000 to 600000) (Amersham Pharmacia). The two columns were calibrated with low and high MW protein standard kits respectively, the former comprising Ribonulease A (MW 13.7 kDa), Chymotrypsinogen A (MW 25.0 kDa), Ovalbumin (MW 43.0 kDa) and Bovine Serum Albumin (MW 67.0 kDa) while the later including Aldolase (MW 158 kDa), Catalase (MW 232 kDa), Ferritin (MW 440 kDa) and Thyroglobulin (MW 690 kDa). The elution solution used in the experiments contained 0.15 M NaCl in 50 mM phosphate buffer, pH 7.0, which was eluted at the flow rate of 0.5 ml/min.

**Results and discussion**

*Quality of model*—Both structures determined by molecular replacement and refined at 2.0 Å resolution showed acceptable values of either crystallographic or free R-factors. Reasonable r.m.s statistics concerning bond length and angles indicated the stereochemical correctness of the final coordinates. All the relative data are listed in Table 1.
Four molecules of gastrodianin-1, named A, B, C and D, resided in the asymmetric unit and behaved somewhat dissimilarly during refinement. High quality electron density accommodated all 112 amino acids for monomer A and B, whereas one or two residues of C and D were missing due to the lack of density at their C-termini. Molecule A and B both have far lower averaged B-factors than that of C or D as a result of more stabilizing contacts from neighboring molecules. Two disordered loop regions 45 to 52 and 77 to 82 of molecule D, which protrude from molecular surface into solvent, are poorly defined in the $2F_o - F_c$ map.

Given the different packing mode, the crystal structure of gastrodianin-4 contained only one molecule in the asymmetric unit belonging to a different space group (Table 1). The model of gastrodianin-4 was refined to a little higher R-factor and has a lower degree of precision than gastrodianin-1, ascribed from higher solvent content and fewer stabilizing contacts among protein chains. Still, the loop region 45 to 52 is stabilized through a lattice contact, while the loop 77 to 82 keeps disorder and the density for two residues at the C-terminus is missing.

_C-terminus_—Bulb lectins are believed to undergo a posttranslational modification during maturation. As for gastrodianin, even if the cDNA sequences corresponds to a 171-AA pre-polypeptide, the precise cleavage site at the C-terminus is still in debate (24,25). The same pending problems were left on other orchid lectins inclusive of LOMBP and EHMBP as well, both of which were thought to exist as monomers (19). In the final model of gastrodianin-1, the C-termini of two monomers (A and B) can be clearly traced in density map, because they are strongly fixed by neighboring monomers (D and C respectively) via salt bridges from their N-terminal cationic groups. In this case, definite electron densities reasonably accounts for the presence of the carboxyl group at Asn112, as the last residue of the chain (Fig. 2). We thereby
identified that the mature gastrodianin is comprised of 112 residues as the total length, and could even presume that the mature peptides of LOMBP and EHMBP are also composed of 112 amino acids since the sequence identity among them reaches as high as 83% (Fig. 1).

Overall structure of gastrodianin—The four crystallographic independent molecules of gastrodianin-1 form two pseudo-dimers (AD and BC) in the asymmetric unit, which related by a non-crystallographic 2-fold axis almost parallel to the y axis (179.36°) (Fig. 3a). Seven hydrogen bonds between A and D or B and C, and two salt bridges between A and B are involved in stabilizing inter-molecular contacts (Fig. 3b). In the crystallographic lattice of gastrodianin-4, a number of hydrogen bonds are involved in contacts between symmetric related molecules, among which Asn36 is supposed to be most notable. Both main chain and side chain of the residue contribute to interactions with neighboring molecules, and very probably for this reason, its dihedral angles fall in disallowed region of Ramachandran plot, as often observed in oligomeric structures of bulb lectins (12, 13). Least-square superpositions of main chain atoms among all independent molecules in gastrodianin-1 structure give r.m.s differences ranging from 0.27 Å (A-B) to 0.43 Å (A-D), and the r.m.s. of 0.51 Å between the two isoelectins, gastrodianin-1 and gastrodianin-4, indicating the absolute structural consensus among these molecules despite different crystal packing states (Fig. 4). Since monomer A in the model of gastrodianin-1 has been refined to the highest degree of precision, its atomic model will accordingly be used in following descriptions and discussions as a representative of the tertiary structure of the orchid lectin.

Despite being a monomer, gastrodianin resembles the general fold of known bulb lectins, the β-prism II fold established since the first structure of GNA (10) (Fig. 4b). It consists of three subdomains (I, II and III), each of which forms a flat four-stranded β-sheet. A total of twelve-
β-strands are arranged to form a β-barrel where a pseudo internal 3-fold axis is located in the center and relates the three faces of the triangular prism. All the strands are perpendicular to the quasi symmetric axis, and in consistent with others, quite a number of non-polar side chains point to the center of β-barrel (Fig. 5a), forming a hydrophobic core stabilized by a network of strong van der Waals interactions. Among these conserved residues are three tryptophan residues, Trp41, Trp72 and Trp104, located in subdomain III, II and I, respectively (Fig.5c). They are kept invariant through all sequences studied so far (Fig. 1) and believed to play a crucial role in stabilizing of the overall fold.

*Unique structural feature as a monomer*—As expected, the first monomeric model in monocot mannose-binding lectin superfamily displays some novel structural properties distinct from those observed in known oligomeric structures, mainly in structural organization of subdomain I (Fig. 5). In all studied models of bulb lectins, a stable dimer forms a basic structural and functional unit, regardless of whether two such dimers further assemble into a tetramer (10,13) or not (14). A hybrid β-sheet observed unexceptionally in subdomain I of these structures consists of 3 strands (the 1st, 10th and 11th) from one chain while the other (the 12th) from the opposite subunit (Fig. 5b). By virtue of this arrangement, a considerable buried surface is given at the dimeric interface, thereby conferring great stability on subunit assembly. This is often referred to as “C-terminal exchange” in which a *cis* peptide bond between residue 98 and 99, e.g. Gly98 and Thr99 in GNA, is generally involved (10,12,13). Although reasonable links between this unusual peptide bond and the dimerization appear to exist considering its extensive occurrence in this protein superfamily, the definite role has not been understood rigorously.

In principle, however, such a mechanism to exchange the C-terminal peptide does not make
any sense in a structure of merolectin, say, gastrodianin. As shown in Fig. 5, within our expectation, the most distinctive feature of the monomeric structure in question lies in subdomain I where a homogeneous sheet is formed instead of the hybrid one described above. Its 12th strand (residue 102 to 106) forms an intramolecular sheet together with the 1st (3 to 5), 10th (84 to 87) and 11th (92 to 96) strand from the same chain (Fig. 5a, b), providing a structural basis for a stable monomer (Fig. 5c). We call it “C-terminal self-assembly” in contrast to above “C-terminal exchange”. Interestingly, the conformation of the C-terminal strand, the 12th, in gastrodianin strictly coincides with its counterpart (residue 101 to 105) in subunit B of GNA (r.m.s.d. 0.25 Å), if the overall main chain of gastrodianin is superimposed on its subunit C. This guarantees the intactness of the topological invariance of sheet I in gastrodianin model. The hydrophobic core of the β-barrel hence also keeps integral and stable in such a fold, especially considering the side chain of Trp104, which is located in strand 12 and adopts almost the same position and orientation compared with that in other models (Fig. 5c). Clearly, such a rearrangement of the C-terminal peptide would bring about a radical decrease or even complete vanishment of the buried surface at the interface between two molecules, which in turn disable the orchid lectin to form a dimer like other bulb lectins.

The distinct conformation of five-residue motif (97-101) of gastrodianin is crucial for the C-terminal self-assembly. In oligomeric lectins only two residues, Gly98 and Thr99 (Ser99 in ASAI) present in the corresponding region (Fig. 1) and take an unusual cis configuration in association with an extended conformation of the following C-terminal peptide 100-109 (Fig. 6c). In monomers, i.e. gastrodianin, LOMBP and EHMBP, these two residues are replaced by an aspartic acid and an asparagine residue (labeled with uparrows in Fig.1) followed by a
three-residue insertion (boxed region in Fig. 1). Notably, in place of the *cis* peptide bond commonly occurring in oligomeric monocot lectins is a *trans* bond presented at the corresponding site in the refined model of gastrodianin. The short loop of Asp97-Asn98-Ser99-Asn100-Asn101 forms a β-turn (Fig. 6b) that makes the following C-terminal polypeptide chain flank the 11th strand (Fig. 5b) rather than protrude from molecular surface. The β-turn connecting the two strands (the 11th and 12th) close to the C-terminus belongs to a 3:5 β-hairpin with a β-bulge inside (Fig. 6b) according to the category system of β-hairpins established by Sibanda *et al.* (37).

The backbone of the β-hairpin is well stabilized by a local H-bond network mediated by the hydrophilic side chains and 3 ordered water molecules in this region (Fig. 6b). In line with the statistics showing that aspartic acid, asparagine and serine are the most commonly-occurring residues in β-hairpins located at molecular surface of peptides (38), all these residues are presented in this case. From structural point of view, it is due to the presence of such a β-hairpin arising from residue replacement and insertion in gastrodianin sequence that the peptide beyond this position has to shift radically from its original location and orientation as in oligomeric states to where it presents in the current model. Consequently, this shift in turn contributes greatly to stabilization of its monomeric state. The unique fold pattern of gastrodianin as a merolectin is observed for the first time (Fig. 5a), obviously differing from those of the oligomers reported previously (39) in structural organization of subdomain I, and thus the structure of gastrodianin should represent a novel type of monomeric monocot mannose-binding lectin.

**Structure determinants and switch governing quaternary state of monocot mannose-binding lectin**—Much effort has gone into the studies of the relationship among sequence motif, structure signature and protein oligomerization in several groups of plant lectins for years (2). For instance,
legume lectins, exhibiting considerable variations in their oligomeric structures with small changes in their sequences, have been well-documented as “natural mutants of quaternary state” (40). Likewise, the superfamily of monocot mannose-binding lectins is also quite heterogeneous in their quaternary organizations, in spite of the high degree of sequence conservation as well as the overwhelming “β-prism II” tertiary fold among all known members. However, information concerning potential sequence elements and structural features that might play crucial roles in subunit assembly is still meager. The first monomeric structure availably presented here offers us an opportunity for deeper insight into the main structure determinants governing the protein oligomerization.

Evidently, the architecture of subdomain I of the orchid lectins is the structure determinant for the distinct quaternary state of the monocot mannose-binding lectins. In the case of “C-terminal self-assembly”, like gastrodianin in the current study, an intramolecular sheet forms in that the 12th strand comes from the intrinsic ten C-terminal residues of the same polypeptide chain, and the lectin adopts the monomeric state accordingly. While if an intersubunit hybrid sheet is established via “C-terminal exchange” instead, like GNA and other oligomers, the 12th strand stretches into the partner subunit and the proteins will fold into dimers or oligomers.

The structural comparison between gastrodianin and the counterpart of GNA shows that the inflexion point of the dramatic structure change occurs at position 97 (corresponding to 98 in GNA) (Fig. 6a), from which the C-terminal polypeptide chain of monomeric gastrodianin goes into the direction completely different from that of dimeric GNA with a deviation of 131°. This C-terminal peptide can go into either the intramolecular subdomain I, or the subdomain I of the partner subunit according to whether it is in a monomeric or oligomeric lectin. The comparison
revealed that the distinct configuration of the peptide bond between residues 97 and 98 (corresponding to 98 and 99 of GNA) in association with three residue insertion or deletion followed this dipeptide could be considered as a structural switch for the above structural ramification, and in turn, a determinant of the quaternary state of the monocot mannose-binding lectins. When the crucial peptide bond adopts a normal trans conformation associated with the three residue insertion (Fig. 1) as in gastrodianin sequence, a five-residue reverse turn belonging to a 3:5 β-hairpin with a β-bulge inside forms (Fig. 6b), which renders the following C-terminal segment reversible to contact the flanking peptide chain 11th β-strand of the same subdomain to establish an intramolecular homogenous four-stranded β-sheet (Fig.5b). In contrast, however, if this peptide bond takes an unusual cis form instead or a strained conformation with the deletion that three residues in sequences, as the case of GNA (Fig. 6c), the followed C-terminal peptide is directed to take an extended conformation and straightly inserts into the subdomain I of the partner subunit to build an intermolecular hybrid β-sheet (Fig.5b) by which dimerization occurs.

It appears notable that the structure switch on monomer/dimer conversion may be sequence-dependent, as revealed by sequence alignment with Phi-blast search (41). The five-residue motif, DNSNN, composing the crucial β-hairpin in gastrodianin model seems to represent a consensus sequence signature occurring in all three known monomers (Fig. 1). In oligomeric structures, to the contrary, any side chains at the crucial position (98 in GNA) could be harmful for the cis bond formation with the followed residue because of introducing close contacts, and therefore glysine is supposed to be the optimum amino acid to be placed at this position. This agrees well with the common occurrence of the two-residue motif GX (X is any amino acids) in oligomeric sequences (Fig. 1). Taken together of the above statements, these two
characteristic motifs separately existing in monomers or oligomers can be considered as sequence determinants that govern the quaternary states of monocot mannose-binding lectins.

The fine structure of gastrodianin in comparison with those of other bulb lectins in oligomeric state uncovers a structural mechanism by which the distinct quaternary states of monocot mannose-binding lectins could be governed, as shown in Fig. 6d. The observations also manifested a special way to achieve the distinct quaternary state of proteins through the strained backbone geometry.

The mutation experiment confirming the structure determinant on protein oligomerization—

In order to confirm the main conclusion from the structural analysis regarding quaternary organization of bulb lectins, a mutation experiment was performed. Since satisfying results have never been achieved in our efforts to obtain soluble recombinant gastrodianin from several E. coli and yeast expression systems, expression of MBP-gastrodianin-1 fusion protein in E.coli was established (see materials and methods). Subsequently, a mutagenesis design was figured out on this fusion gene, the loop from 97 to 101 being concerned in the first place. In our performance, Asp97 and Asn98 was mutated into glycine and threonine residues respectively, while the coden of Ser99, Asn100 and Asn101 were deleted from the gene, to match the sequence of GNA in this segment.

After expression and purification, recombinant fusion proteins with high purity of the wild type and the mutant were both applied to gel filtration chromatography to compare their molecular sizes in solution. On the elution profile from Superdex 75 column, the peak of native MBP-gastrodianin appeared at the elution volume 10.1 ml (the expected time of 20.2 min) corresponding to an apparent molecular weight of about 53 KD (Fig. 7A), which is generally
consistent with the monomeric molecular size of the recombinant fusion protein (43 KD and 13 KD for MBP and gastrodianin, respectively). The mutant, however, was eluted in the flow-through fraction with the elusion volume of 8.2 ml (the expected time of 16.4 min) (Fig. 7C), indicating its molecular size must be greater than 70 kDa. On the elution profile from Superdex 200 column, the peak of mutant MBP-gastrodianin was eluted at the elution volume of 13.4 ml (the expected time of 26.8 min) corresponding to an apparent molecular weight of about 130 KD (Fig. 7E), which is roughly coincident to the molecular size of a dimeric form of the muted fusion protein. The SDS-PAGE analyses of the elution peaks of the wild type and the mutant fusion protein were showed in Figure 7B and 7D, respectively, which clearly identified the results from the chromatography. These results unquestionably confirm the change of the protein quaternary states caused by mutation, and most likely, the association form of the mutant fusion protein can be assumed as dimerization. From these experimental data, we can sensibly conclude that this sequence motif from position 97 to 101 is responsible for the conversion between monomeric and dimeric states of monocot mannose-binding lectins.

*Putative carbohydrate binding sites*—Although the complex structure of gastrodianin with its saccharide ligands was not available in our studies, some binding features can also be anticipated by comparison with its homologous sequences and structures. The consensus sequence motif QXDXNXVXY (X is one of any amino acid residues) (5) occurs in all three putative carbohydrate-recognition domains (CRDs) of the orchid protein (Fig. 1), indicating the same binding mode as other bulb lectins. The conserved side chains in the binding pockets of gastrodianin consist well with those in GNA (r.m.s.d. 0.17 Å), strengthening the view that binding carbohydrate does not cause any noticeable conformational change of the protein itself. This also
confirms its strong preference of the axial hydroxyl at position 2 in the ligand, which is the common property shared with all known members in this protein superfamily (10,13,14). However, such a deduction may lead to a variance from previous biochemical observations suggesting another possibility of chitin-binding activity of the orchid lectin (23).

In spite of the similar geometries at binding sites in both the orchid and snowdrop lectins, their preferences to complex glycans may vary considerably. Previous observations has suggested that the variability in quaternary structures may be an alternative means of diversifying ligand preferences among lectins (42). A good example for bulb lectins has been well documented. Some tetramers such as GNA or NPL display an inhibitory activity against retrovirus resulting from their strong affinity to GP120, the major glycoprotein of HIV (6,7), while the garlic agglutinin, a dimer, does not (9). The structure of the snowdrop lectin complex with a branched mannopentaose revealed two distinct binding modes, both cross-linking two subunits to form extended binding sites (43). Gastrodianin, however, obviously cannot dock the mannopentaose in a similar fashion considering its monomeric state. In contrast, one can hardly believe that a monomer like gastrodianin is able to bind polysaccharide with complicated branches as oligomers do, due to lacking contacts from neighboring subunits. This may lead one to surmise that the orchid lectin possesses a distinct choice for its ligands, which are most probably different from what we know for oligomers.

**Biological implications**—Like other lectin families, the physiological role of the monocot mannose-binding lectins remains poorly understood thus far (2), even though this question has been addressed since the first discovery of GNA (44). As evidences has been accumulated in recent years, bulb lectins are gradually believed to serve as a device of plant defense system
against the ravage of rodent pests, insects and microorganisms (1,8). For instance, the presumed protective role of GNA has been well-established from experiments in which the detrimental effect on sucking invertebrates was demonstrated when they were fed with artificial diets containing the lectin or transgenic plants expressing its gene (45,46). Likewise, gastrodianin has been thought to be involved in microorganism defense and was firstly defined as an antifungal protein (20,22). The orchid *G. elata*, unlike green plants, lives on the released nutrients from harvested haphae rather than lead an autotrophic life. Gastrodianin, in this circumstance, seems to exist as a long-term evolutonal product of such specialization, which helps the plant fight with its symbiotic partner, the fungi. From evolutonal point of view, it possibly follows that nature has evolved monocot mannose-binding lectins with different quaternary states for their adaptation in various situations. Whereas the oligomers, such as GNA and NPL, play a defensive role against insects or other invertebrates, the monomers function as antifungal proteins.

During the past decades have formidable arrays of plant proteins been identified as antifungal proteins and a number of them were hence biotechnologically applied in crop production, including chitinases (47), hevein-type proteins (48,49), thionins, plant defensins (50) and type-1 ribosome-inactivating proteins (51). Exclusively, gastrodianin belongs to none of the known classes of antifungal proteins in terms of sequence homology or structure similarity, although its inhibitory activity against the fungus *Valsa ambiens* is on a par with that of another established antifungal lectin, *Urtica dioica* agglutinin (UDA) (49). Despite the fact that most plant lectins with antifungal properties are found to be chitin-binding proteins, such as hevein or hevein-like proteins, we can hardly believe that the orchid lectin works in the same manner because its binding specificity seems to be fixed on mannose only from our structure studies, as a
typical monocot mannose-binding lectins. Noticeably, however, the most important feature of the orchid protein mainly lies in its particular quaternary state distinctive from other known members in the superfamily of bulb lectins, and in this way, the antifungal activity of gastrodianin appears to arise more from its overall structure as a monomer than from any local structure variations. This may arrive at a conclusion that it is the monomeric form that plays the dominant role in its inhibitory ability. The view is supported by an observation that a homologous merolectin from the orchid *Epipactis helleborine* also retards hyphal growth *in vitro*, whereas related dimeric or tetrameric lectins, including GNA, do not display any antifungal activities (25). In principle, smaller molecular size as a result of its monomeric form definitely favors the orchid lectin to penetrate through fungal cell walls, since the size exclusion limit for a typical antifungal protein keeps around 15 – 20 KD (52). An evidence comes from the bioassay of the MBP-gastrodianin fusion protein that shows no inhibitory activity *in vitro*, as opposed to the protease-cleaved protein (25).

Even with such a speculation, the details of how gastrodianin resists fungal invasion stay unclear on account of our meager knowledge about the cell wall structure of the fungus *Armillaria mellea*, the symbiotic partner with the orchid *Gastrodia elata*. A scenario that one can figure out concerns the carbohydrate binding preference of the orchid lectin. The lectin may work in such a way, like chitin-binding proteins, that it binds mannose-rich glycans, which are used by phytopathogenic fungi as building blocks in their cell wall constructions, and consequently blocks their hyphal growth. Alternatively, gastrodianin can also bind mannose-containing glycoconjugates on fungal membranes so as to affect their cellular metabolism thereafter, which is implied by homology modelling of the core domain of the endogenous lectin comitin from *Dictyostelium*
discoideum using GNA as the template, showing its mannose-binding specificity onto mannose residues in membrane glycans (53).

Acknowledgement

The work was supported by the NSFC (30070162), the Ministry of Sciences and Technology of China (863-2002BA711A13, 973-G1999075064), and the Chinese Academy of Sciences (KSCX1-SW-17, KSCX2-SW-322). The X-ray data collection was supported by KEK of Japan (2002G310). We thank Prof. N. Sakabe for his helps during data collection.

PDB entries

Both coordinates of gastrodianin have been deposited in PDB and assigned with IDs of 1XD5 (gastrodianin-1) and 1XD6 (gastrodianin-4) respectively.
Reference


Legends of figures

Figure 1. Sequence alignment among gastrodianin-1, gastrodianin-4 and their homologies inclusive of LOMBP, EHMBP (19), GNA (10), NPL (13), subunits A and D from ASAI (15), SCAman (12) and domains 1 and 2 of SCAfet (16). All identities and similarities are shaded with dark background and the β-barrel related strands are labeled with arrows beneath the sequence. All three mannose-binding sites with the consensus sequence motif are marked out in the alignment, and the two residues at position 97 and 98 (corresponding to 98 and 99 for GNA or NPL), where a trans peptide bond presents in between for monomers instead of the cis peptide bond generally existing in oligomers, are denoted with up-triangles.

Figure 2. Stereo view of the C-terminus of molecule A in crystal structure of gastrodianin-1, with the corresponding electron density $(2F_o - F_c)$ map contoured at 1.0 σ level. The carboxyl of Asn112 at C end of the main chain can be clearly defined.

Figure 3. The overall structure of gastrodianin-1 in crystals. (a) The four independent molecules A, B, C and D in the asymmetric unit projected down the y axis. The pseudo dimer AD and BC are related by a two-fold non-crystallographic axis approximately parallel to the crystallographic y axis. (b) Interactions between monomers A and D. (c) Contacts between A and B.

Figure 4. Monomeric gastrodianin fold. (a) superimpositions between four gastrodianin monomers (green) and gastrodianin-4 monomer (red). (b) The consensus general structure of gastrodianins distinct with three subdomains, I, II and III, consists of
four-stranded β-sheet, to form β-prism II fold.

Figure 5. Structural organization of subdomain I in monomeric gastrodianin through the C-terminal self-assembly in comparison with that of GNA via the C-terminal exchange. (a) Unique fold pattern of gastrodianin. Subdomain I composed of four-stranded β-sheet from a single chain is the structural basis for its monomeric form. Hydrophobic residues that form the hydrophobic core inside the barrel are colored pink, yellow, purple and cyan, according to their strand location. The single disulfide bridge is shown in green. Residues in a reverse turn from 97 to 101 that determine the fold of C-terminal strand are highlighted with magenta backgrounds. (b) The homologous (left) and hybrid (right) four-stranded β-sheets appeared in monomeric gastrodianin and dimeric GNA subunit, respectively. (c) Architectural comparison of the protomer between gastrodianin (left) and GNA (right) whose subdomain I were built in mode of C-terminal self-assembly and C-terminal exchange, respectively. For the C-terminal self-assembly, all the four β-strands composing the subdomain I come from the same chain (left), whereas for the C-terminal exchange (right) the last β-strand (red) with the residue Trp102 is located on the external chain from the opposite subunit. The GNA models was drawn from PDB entry 1JPC.

Figure 6. Structural switch for the monomer/dimer conversion. (a) The structural comparison between gastrodianin (in blue) and the counterpart of GNA (in red). An inflection point of the dramatic structure change appeared at position 97 (corresponding to 98 in GNA) from which the C-terminal polypeptide chains go into complete different
directions with a deviation of 131°. (b) The five-residue reverse turn (97-101) formed in gastrodianin. In the turn the peptide bond (97-98) adopts an usual trans form and the reversed conformation is well stabilized with a hydrogen bond network, which directed the C-terminal segment to turn into the intrinsic subdomain I to form the homolougous β-sheet through C-terminal self-assembly. (c) The extended conformation in the corresponding region of GNA caused by an unusual cis peptide bond between the corresponding residues 98 and 99, which induced the C-terminal peptide to go into subdomain I of its partner subunit to build up a hybrid β-sheet with C-terminal exchange. (d) The two modes schematically showing the structural mechanisms by which the quaternary states of monocot mannose-binding lectins could be governed.

Figure 7. The elution profiles from size exclusion chromatography and SDS-PAGE results of the native (A, B) and mutant fusion proteins (C, D, E). (A) Superdex 75 column for the wild type fusion protein. The native fusion protein appeared at peak w with the elution volume of 10.1 ml (expected time of 20.2 min) corresponding to an apparent molecular weight of 53 KD. (B) SDS-PAGE result of the recombinant native fusion protein. Line I is the standard protein marker; Line V represents the elution peak after affinity chromatography; Line II ~ IV are the fractions of whole peak w in Fig. 7A, which show that the elution of Superdex 75 column has only one fraction with a molecular weight of 55 KD estimated from BandScan 4.5 software. (C) Superdex 75 column for the mutant fusion protein. The mutant protein in question was eluted in the flow-through fraction with the elusion volume of 8.2 ml (expected time of 16.4
min) (Fig. 7C), indicating its molecular size must be greater than 70 kDa. (D) SDS-PAGE result of the mutant fusion protein. Line I is the standard protein marker; Line VI represents the elution peak after affinity chromatography; Line II ~ V are the fractions of whole peak m in Fig. 7C, which shows a molecular weight of about 55 KD (estimated from BandScan 4.5 software). (E) Elution profile of the mutant protein from Superdex 200 column. The mutant fusion protein appeared at peak m with the elution volume of 13.4 ml (expected time of 26.8 min) corresponding to an apparent molecular weight of about 130 KD, which is roughly coincident to the molecular size of a dimeric form of the muted fusion protein. The result indicated that the mutant fusion protein should consist of two identical mutant protomers linked by non-covalent interactions. These results confirm the change of the protein quaternary states caused by mutation, and most likely, the association form of the mutant fusion protein can be assumed as dimerization.
Table 1. Data processing and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>Gastrodianin-1</th>
<th>Gastrodianin-4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2;2;2 (No. 18)</td>
<td>C222; (No. 20)</td>
</tr>
<tr>
<td>Unit cell dimensions (Å)</td>
<td>a = 61.1, b = 91.5, c = 81.1</td>
<td>a = 78.1, b = 97.5, c = 36.1</td>
</tr>
<tr>
<td>Contents of the asymmetric unit</td>
<td>4 monomers</td>
<td>1 monomer (110 residues)</td>
</tr>
<tr>
<td></td>
<td>(A and B: 112 residues; C: 111 residues; D: 110 residues)</td>
<td></td>
</tr>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>31377</td>
<td>9216</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.23</td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å) (last shell)</td>
<td>24.69 – 2.0</td>
<td>18.93 - 2.0 (2.07 –2.0)</td>
</tr>
<tr>
<td>Completeness (%) (last shell)</td>
<td>99.8 (99.8)</td>
<td>95.3 (91.2)</td>
</tr>
<tr>
<td>1/σ (last shell)</td>
<td>6.7 (2.5)</td>
<td>6.3 (2.9)</td>
</tr>
<tr>
<td>Rmerge (%) (last shell)</td>
<td>9.7 (29.6)</td>
<td>6.7</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflections used in refinement</td>
<td>31339</td>
<td>9092</td>
</tr>
<tr>
<td>Size of test set (%)</td>
<td>5.0</td>
<td>7.5</td>
</tr>
<tr>
<td>Rcryst / Rfree (%)</td>
<td>16.6 / 20.6</td>
<td>19.9 / 23.8</td>
</tr>
<tr>
<td><strong>Ramachandran plot</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core region (%)</td>
<td>84.9</td>
<td>88.5</td>
</tr>
<tr>
<td>Allowed region (%)</td>
<td>15.1</td>
<td>10.5</td>
</tr>
<tr>
<td><strong>Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>3442</td>
<td>847</td>
</tr>
<tr>
<td>No. of hetero atoms</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>No. of solvent molecules</td>
<td>265</td>
<td>43</td>
</tr>
<tr>
<td>rmsd of bond lengths (Å)</td>
<td>0.008</td>
<td>0.010</td>
</tr>
<tr>
<td>rmsd of bond angles (deg)</td>
<td>1.29</td>
<td>1.45</td>
</tr>
<tr>
<td>rmsd of dihedral angles (deg)</td>
<td>27.11</td>
<td>27.66</td>
</tr>
<tr>
<td>rmsd of improper angles (deg)</td>
<td>1.18</td>
<td>1.31</td>
</tr>
<tr>
<td>Estimated coordinate error (Å)</td>
<td>0.189 / 0.183</td>
<td>0.244 / 0.194</td>
</tr>
<tr>
<td><strong>Averaged B-factors (Å²)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monomer A</td>
<td>13.6</td>
<td>24.7</td>
</tr>
<tr>
<td>Monomer B</td>
<td>14.1</td>
<td></td>
</tr>
<tr>
<td>Monomer C</td>
<td>22.5</td>
<td></td>
</tr>
<tr>
<td>Monomer D</td>
<td>27.0</td>
<td></td>
</tr>
<tr>
<td>Hetero atoms</td>
<td>17.1</td>
<td>33.0</td>
</tr>
<tr>
<td>Solvent atoms</td>
<td>31.1</td>
<td>38.2</td>
</tr>
</tbody>
</table>

§ Estimated according to Luzzati plot (54) and the Cruickshank method (55).
<table>
<thead>
<tr>
<th>Protein</th>
<th>Mannose-binding site I</th>
<th>Mannose-binding site II</th>
<th>Mannose-binding site III</th>
</tr>
</thead>
<tbody>
<tr>
<td>gastrodianin-1</td>
<td>S D R N S - - - - - - - - G H Q G - - - - A G G S G S A E G P - - - - L F I M E N D C N L V L Y N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gastrodianin-2</td>
<td>S D R N A - - - - - - - - G K S G - - - - A G G S G S A E G P - - - - L F I M E N D C N L V L Y N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOMB</td>
<td>S D R N A - - - - - - - - G Q S G - - - - A G G S G S A E G P - - - - L F I M E N D C N L V L Y N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EHMBP</td>
<td>A D R N A - - - - - - - - G Q S G - - - - A G G S G S A E G P - - - - L F I M E N D C N L V L Y N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GNA</td>
<td>D N I Y S - - - - - - - - G E T S G T E F N G S - - - - V F I M E D E C N L V L Y N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPL</td>
<td>D N I Y S - - - - - - - - G E T S G T E F N G S - - - - V F I M E D E C N L V L Y N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASAI A</td>
<td>R N L I T - - - - - - - - G E T S G T E F N G S - - - - V F I M E D E C N L V L Y N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASAI D</td>
<td>R N L I T - - - - - - - - G E T S G T E F N G S - - - - V F I M E D E C N L V L Y N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCAman</td>
<td>N N I - - - - - - - - G E T S G T E F N G S - - - - V F I M E D E C N L V L Y N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCAfet-DOM1</td>
<td>R N L I T - - - - - - - - G E T S G T E F N G S - - - - V F I M E D E C N L V L Y N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCAfet-DOM2</td>
<td>R N L I T - - - - - - - - G E T S G T E F N G S - - - - V F I M E D E C N L V L Y N</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1
Fig. 5
Fig. 6
Figure 7
Structural mechanism governing the quaternary organization of monocot mannose-binding lectin revealed by the novel monomeric structure of an orchid lectin

Wei Liu, Na Yang, Jingjin Ding, Ren-huai Huang, Zhong Hu and Da-Cheng Wang

J. Biol. Chem. published online January 13, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M411634200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/early/2005/01/13/jbc.M411634200.citation.full.html#ref-list-1