



**HAL**  
open science

## Insights into the structural basis of the pH-dependent dimer-tetramer equilibrium through crystallographic analysis of recombinant Diocleinae lectins

Celso S. Nagano, Juan J. Calvete, Domingo Baretino, Alicia Pérez, Benildo S. Cavada, L Sanz

### ► To cite this version:

Celso S. Nagano, Juan J. Calvete, Domingo Baretino, Alicia Pérez, Benildo S. Cavada, et al.. Insights into the structural basis of the pH-dependent dimer-tetramer equilibrium through crystallographic analysis of recombinant Diocleinae lectins. *Biochemical Journal*, 2007, 409 (2), pp.417-428. 10.1042/BJ20070942 . hal-00478846

**HAL Id: hal-00478846**

**<https://hal.science/hal-00478846>**

Submitted on 30 Apr 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Insights into the structural basis of the pH-dependent dimer-tetramer equilibrium through crystallographic analysis of recombinant *Diocleinae* lectins**

Celso S. NAGANO<sup>\*</sup>, Juan J. CALVETE<sup>\*, ‡</sup>, Domingo BARETTINO<sup>\*</sup>, Alicia PÉREZ<sup>\*</sup>, Benildo S. CAVADA<sup>†</sup> & Libia SANZ<sup>\*, ‡</sup>

<sup>\*</sup>Instituto de Biomedicina de Valencia, C.S.I.C., Jaime Roig 11, 46010 Valencia, Spain.

<sup>†</sup>BioMol-Lab, Departamento de Bioquímica e Biología Molecular, Universidade Federal do Ceará, Fortaleza, CE 60451-970, Brazil

**Running Title:** Crystallographic analysis of recombinant *Diocleinae* lectins

<sup>‡</sup> **Address correspondence to:** Juan J. Calvete (for issues concerning X-ray crystallography) or Libia Sanz (for questions related to Molecular Biology), Instituto de Biomedicina de Valencia, C.S.I.C., Jaime Roig 11, 46010 Valencia (Spain). Phone: +34 96 339 1778, Fax: +34 96 369 0800, E-mail addresses: [jcalvete@ibv.csic.es](mailto:jcalvete@ibv.csic.es), [Libia.Sanz@ibv.csic.es](mailto:Libia.Sanz@ibv.csic.es).

## Synopsis

The structural ground underlying the pH-dependency of the dimer-tetramer transition of *Diocleinae* lectins was investigated by equilibrium sedimentation and X-ray crystal structure determination of wild-type and site-directed mutants of recombinant lectins. Synthetic genes coding for the full-length  $\alpha$ -chains of the seed lectins of *Dioclea guianensis* (r- $\alpha$ Dguia) and *Dioclea grandiflora* (r- $\alpha$ DGL) were designed and expressed in *E. coli*. This pioneer approach, which will be described in detail, yielded recombinant lectins displaying carbohydrate-binding activity, dimer-tetramer equilibria, and crystal structures indistinguishable from their natural homologues. Conversion of the pH-stable tetrameric r- $\alpha$ DGL into a structure exhibiting pH-dependent dimer-tetramer transition was accomplished through mutations that abolished the interdimeric interactions at the central cavity of the tetrameric lectins. Both, the central and the peripheral interacting regions bear structural information for formation of the canonical legume lectin tetramer. We hypothesize that the strength of the ionic contacts at these sites may be modulated by the pH, leading to dissociation of those lectin structures that are not locked into a pH-stable tetramer through interdimeric contacts networking the central cavity loops.

**Keywords:** Synthetic gene, recombinant lectin, *Diocleinae* lectin, dimer-tetramer equilibrium, site-directed mutagenesis, X-ray crystal structure

## INTRODUCTION

Lectins are a structurally heterogeneous group of proteins of non-immune origin possessing at least one non-catalytic domain that binds reversibly to a specific mono- or oligosaccharide [1-5]. Lectins are ubiquitous in all the kingdoms of life and play biological roles in many cellular processes, such as cell communication, host defense, fertilization and development, parasitic infection, tumor metastasis, plant defense against phyt predators and pathogens, etc. [3], by deciphering the glycodes encoded in the structures of glycans attached to soluble and integral cell membrane glycoconjugates [4]. Mechanisms for sugar recognition have evolved independently in a small number of protein domains [1,5], and lectins sharing a common sugar-binding fold may depart in their carbohydrate-recognition specificities [6]. Multivalence is a mechanism frequently used by lectins for increasing their glycan binding avidity and specificity [7,8]. In addition, most lectins are oligomeric or multidomain proteins, and variability in their quaternary associations represents a common evolutionary mechanism for creating structural and functional diversification among lectins exhibiting the same tertiary fold [5,9-13].

Although recent years have witnessed the elucidation of a large variety of novel structures of lectins [9,10,14], plant lectins from the Leguminosae family remain the most thoroughly studied group of sugar-binding proteins. In particular, lectins isolated from legume seeds have been regarded as paradigmatic molecules for investigating the structural basis and thermodynamics of selective sugar recognition [5-8,15-17]. Legume lectin subunits display a high degree of tertiary fold conservation. The architecture of the so-called "legume lectin fold" is structurally related to a jelly-roll motif characterized by sandwich formed by a curved seven-stranded front  $\beta$ -sheet and a nearly flat six-stranded back  $\beta$ -sheet, bridged on one side by a top five-stranded  $\beta$ -sheet [5]. The tetrameric assembly of seed lectins from the *Dioclea*, *Cratylia*, and *Canavalia* genera of the *Diocleinae* subtribe of the *Papilionoideae* subfamily of legume lectins involves a cross-wise back-to-back association of two dimers built by the side-by-side, antiparallel alignment of the subunit six-stranded back  $\beta$ -sheets [5], first characterized in the crystal structure of concanavalin A [18-20]. *Diocleinae* lectins exhibit pH-dependent dimer-tetramer equilibria [21-24]. The physiological relevance of this phenomenon remains elusive. However, it is worth to note that the 25 x 8 Å<sup>2</sup> water-filled central cavity of the tetramer of ConA and other leguminous lectins has been

reported to bind various non-polar compounds, such as the plant auxin  $\beta$ -indoleacetic acid and cytokinins [25, 26]. Auxins and cytokinins function as plant growth hormones and thus the lectins could use dimer-tetramer equilibrium as a mechanism for storage or transport of hormones.

Structural evidence gathered from comparison of the crystal structures of the pH-independent tetrameric *Dioclea grandiflora* lectin (1DGL) [27] and those of the lectins from *Dioclea guianensis* (1H9P, 1H9W) [28] and *Cratylia floribunda* (CFL, crystallized at acidic (2D3R) and basic (2D3P) pH) [29], which exhibit pH-dependent dimer-tetramer transition, revealed that interdimer contacts in the 'canonical ConA tetramer' are made between the homologous regions encompassing residues 53-78 of monomers A and D and B and C at the periphery of the dimers, and between the loop comprising residues 117-123 of each monomer located in the water-filled cavity at the centre of the tetramer. Interactions at the peripheral contact sites are essentially conserved in the structures of *Cratylia floribunda* seed lectin crystallized in conditions where it exist as a dimer (pH 4.6, aCFL) or as a tetramer (pH 8.5, bCFL) in solution [29]. These interface regions of aCFL and bCFL can be superimposed with an r.m.s.deviation of 0.42 Å. Similarly, the peripheral interdimeric sites of the seed lectins from *Dioclea guianensis* (pH-dependent tetramer) and *Dioclea grandiflora* (pH-independent tetramer) are essentially conserved [28]. The structural data also pointed to the participation of two histidine residues, His131 and H51, in stabilizing the conformation of the central cavity loops comprising residues 117-123 from each subunit. However, in contrast to loop ordering through a His51-dependent mechanism observed in CFL, which does not yield a pH-stable tetrameric association, His131 in DGL establishes a network of interdimer interactions bridging the four central loops into a pH-independent tetramer. Asn131 in CFL is unable to reproduce these intersubunit interactions. In addition, the number of interdimer contacts in the central cavity are drastically reduced in the aCFL versus the bCFL structure [29]. The reduced number of interdimer contacts yielding a "weakened" tetrameric association of aCFL is due by increased distances between atoms which are hydrogen-bonded in bCFL as a consequence of the distinct spatial orientation of residues His51, Ser66, Ser110, Thr117, Asn118, Ser119 and Thr120 in aCFL and bCFL [29].

The apparent paradox that a His residue (pKa 5.5-7.5) determines pH independence might be explained assuming that the network of intermonomer contacts centered at His131 can not be disrupted by altering the pH value of the solution in the range 4.5-

8.5. Current evidence suggests that although both, the central and the peripheral interacting regions may bear information for formation of the canonical legume lectin tetramer, the absence of the stabilizing intermonomeric interactions (i.e. when Asn occupies position 131) allows pH-dependent destabilization of the tetrameric lectin structure through a hitherto elusive mechanism. This view reinforces our hypothesis of the key role of the side chain of the residue at position 131 for determining the pH dependence or independence of the tetrameric association of the seed lectins from *Diocleinae*.

To gain a deeper insight into the structural determinants underlying *Diocleinae* lectin pH-dependent quaternary association, we have designed and expressed in *E. coli* synthetic genes coding for the full-length  $\alpha$ -chains of *Dioclea guianensis* and *Dioclea grandiflora* lectins. This pioneer approach, which will be described in detail, yielded recombinant proteins (r- $\alpha$ -Dguia and r- $\alpha$ -DGL), which are biochemically indistinguishable from the natural occurring lectins. Conversion by site-directed mutagenesis of residues at positions 123, 131, and 132 of the pH-independent tetrameric r-DGL for the corresponding residues of the *Dioclea guianensis* lectin endowed the triple mutant with pH-dependent dimer-tetramer equilibrium. The structural basis of this transition was investigated by solving the crystal structures of the wild type (r- $\alpha$ -Dguia, r- $\alpha$ -DGL) and mutated (r- $\alpha$ -Dguia S131H and r- $\alpha$ -DGL E123A/H131N/K132/Q) recombinant lectins.

## EXPERIMENTAL

### *PCR amplification of the genes and mRNAs coding for the precursors of Dioclea guianensis and Dioclea grandiflora seed lectins*

*Dioclea guianensis* and *Dioclea grandiflora* seeds were harvested from plants at the Federal University of Ceara campus, Fortaleza, Brazil. For germination (at the Instituto de Biomedicina de Valencia, Valencia, Spain), seeds were pretreated by immersion in concentrated sulphuric acid for 20 min followed by washing in running tap water for 5 min. The SO<sub>4</sub>H<sub>2</sub>-treated seeds were placed between two filter paper sheets moistened with distilled water, rolled-up, and kept in a water-saturated chamber in the dark. Genomic DNAs (gDNA) from *D. guianensis* and *D. grandiflora* were extracted from young leaves using the CTAB procedure [30]. For total RNA isolation, young buds in liquid nitrogen were immediately grounded to a powder with a pestle.

Total cellular RNA from *D. guianensis* was isolated with Concert Plant RNA reagent (Invitrogen) according to the manufacturer's protocol. The first strand of cDNA of *D. guianensis* was reverse-transcribed using a standard method. Briefly, 12  $\mu$ l of 0.1% diethylpyrocarbonate (DEPC)-treated water containing 0.1  $\mu$ g of 5'-CCA GTG AGC AGA GTG ACG AGG ACT CGA GCT CAA GCT (T)<sub>16</sub>-3' (Qt) (Sigma-Genosys, Haverhill, UK) and 1  $\mu$ g of total RNA were heated to 70 °C for 10 min to denature any possible secondary RNA structure, cooled at ice temperature, and mixed with 4  $\mu$ l of 5x First Strand Buffer (Promega), 2  $\mu$ l of 0.1 M DTT (Promega), 1  $\mu$ l of the ribonuclease inhibitor RNasin (40u/ $\mu$ l) (Promega), 1  $\mu$ l of dNTP (10 mM each) (Eppendorf), and 200 units of MMLV RT (Moloney Murine Leukemia Virus Reverse Transcriptase) (RNase H Minus, Promega). The final volume was adjusted to 20  $\mu$ l and the reaction mixture was incubated at 42 °C for 1 h, followed by 10 min at 50 °C to inactivate the enzyme. DNA was PCR-amplified using one-tenth of the reverse transcriptase reaction as template. The forward primer, 5'-CCNGTSCACATTTGGGA-3', and the reverse primer, 5'-WACYCTHACCCAYTCNGG-3' (in IUPAC code for degeneracies) were designed according to the conserved amino acid sequences of legume lectins, PVHIWE primer sense and PEWVRV primer antisense, respectively. The PCR amplification protocol included an initial denaturation step at 95 °C for 10 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 50 °C for 30 s, and extension at 72 °C for 2 min. The recovered PCR products (500bp) were cloned into pGEM-T vector (Promega), and transformed into *Escherichia coli* strain DH5 $\alpha$  (Novagen, Madison, WI, USA). The white transformants were screened by PCR and the positive clones were confirmed by sequencing using an Applied Biosystems model 377 DNA sequencer. The DNA sequences obtained were used for designing specific oligonucleotides for 3'RACE [31]. The reaction was carried out with the sense primer GSP-3'RACE (5'-TGG GAA GGT AGG AAC TGC ACA CA -3') and the antisense primer Qt-anchor (5'-CCA GTG AGC AGA GTG ACG-3'). To determine the upstream sequence of the *Dioclea guianensis* lectin gene, 1  $\mu$ g of genomic DNA was digested with 10 units of *Rsa*I, ligated to an adaptor [32] with 10 units of T4-DNA ligase (Invitrogen) for 12 hours at 13 °C and amplified by nested PCR using the sense primers IN (5'-AGG GCG TGG TGC GGA GGG CGG T-3') or OU (5'-TGT AGC GTG AAG ACG ACA GAA-3') and anti-sense specific primers GSP1-5'RACE (5'-AAG CTT GCG ACC ACA GCA GAT-3') or GSP2-5'RACE (5'-CCA CCA GAA CCA CTA GGG ATT GAA-3'). Based on the amplified sequences, we designed the specific forward primer DguiaFW (5'- CCA



TGG ATG GGT ATT TCA AAA AAA TC-3', which includes a *NcoI* restriction site, underlined) and the reverse primer Dguia RV (5'-TCA GAC GAC GGA TGC AAT-3') for amplification of the complete Open Read Frame (using either cDNA or gDNA). For PCR amplification with specific primers, the reaction protocol was similar to the cDNA amplification with degenerate primers, but the annealing step was performed at 60°C for 30s. The amplified fragments were separated by 2% agarose gel electrophoresis, purified using the Perfect Pre Gel Clean Up kit (Eppendorf, Hamburg, Germany), and cloned in a pGEM-T vector.

For PCR-amplification of the DGL precursor, genomic DNA was used as template with the forward primer 5'-CTG ATA GTA GTG AGC AGG GTG AGC-3', corresponding to the amino acid sequence <sup>20</sup>LIVVSRVS<sup>27</sup> of the Dguia precursor protein (Fig.1), and the reverse primer 5'- ATT CGT CTT TAA CTT AGA AG-3' that corresponds to the last six amino acids of the  $\beta$ -chain (<sup>275</sup>SKLKTN<sup>280</sup>) (Fig.1). The reaction protocol was similar to the cDNA amplification with semi-degenerated primers. The obtained PCR product was cloned into the pGEM-T vector.

### ***Construction of synthetic genes encoding the $\alpha$ -chains of the *Dioclea guianensis* and the *Dioclea grandiflora* lectins***

To produce synthetic genes coding for mature-like  $\alpha$  ( $=\beta + \gamma$ ) -chains of the *Dioclea guianensis* and the *Dioclea grandiflora* seed lectins, DNAs encoding the beta and the gamma chains were amplified separately using primers containing restriction sites.

Amplification of the  $\beta$  chains of both, Dguia and DGL, was performed with the forward primer  $\beta$  5'- gcc atg GCC GAT ACT ATC GTT GCT G-3' which contains a *NcoI* restriction site (underlined) and a sequence coding for the first six amino acid of the polypeptide chain (ADTIVA), and the reverse primer  $\beta$  5'- cccgga ATT CGT CTT TAA CTT AGA AG-3', which includes an *EcoRI* restriction site and the last seven amino acids of the  $\beta$ -chain (SKLKTN).

The  $\gamma$ -chains were PCR-amplified with the sense primer  $\gamma$  5'- ccggaat TCA ATA GCA GAT G(C/A)A AAT-3' (C for Dguia and A for DGL) that possesses the *EcoRI* restriction site and the first six amino acid of the gamma chain (Dguia, SIADAN; DGL, SIADEN) and the anti-sense primer  $\gamma$  5'- aag cg gcc gc TCA ATT TGC ATC AGG AAA GAG which includes a restriction site for *NotI* (underlined), a stop codon (in italics and bold), and the last six C-terminal residues of the mature lectins (LFPDAN).



The PCR protocol included denaturation at 94 °C for 2 min, followed by 25 cycles of denaturation (30 s at 94 °C), annealing (30 s at 60 °C), and extension (30 s at 72 °C), and a final extension for 7 min at 72 °C. The amplified fragments were purified using the Perfect Prep Gel Clean Up kit (Eppendorf, Hamburg, Germany) and cloned in a pGEM-T. *E. coli* DH5 $\alpha$  cells (Novagen, Madison, WI) were transformed by electroporation using an Eppendorf 2510 electroporator following the manufacturer's instructions. Positive clones, selected by growing the transformed cells in Luria broth (LB) medium containing 100  $\mu$ g/ml ampicillin, were confirmed by PCR amplification using the above primers, and the PCR-amplified fragments were sequenced (using an Applied Biosystems model 377DNAsequencer) to check the correctness of the sequences and direction of inserts.

For generation of expression vectors comprising  $\alpha$ -chain-thioredoxin-His<sub>6</sub> fusion proteins, the pGEM-T plasmids containing beta-chains ( $\beta$ -chain/pGEM-T), gamma-chains ( $\gamma$ -chain/pGEM-T), and the pET32a vector (Novagen) were serially digested with *NcoI/EcoRI*, *EcoRI/Not I*, and *NcoI/NotI*, respectively, for 12 h at 37 °C. The 300-bp beta and gamma fragments and the pET32a vector were purified from 1.5% agarose gels using the Eppendorf Perfect Pre Gel Clean Up kit. The DNA fragment coding for the  $\gamma$ -chain was dephosphorylated at 37 °C for 1 hour with 1 unit of CIP alkaline phosphatase (Roche). After this reaction, the beta and the gamma fragments, and the open pET32a vector were ligated with 10 units of T4-DNAIigase (Invitrogen) overnight at 13 °C. These constructs were used to transform electrocompetent *E. coli* DH5 $\alpha$  cells. The plasmid was sequenced to confirm the correctness of the construct. The plasmidic DNAs from positive clones were used to transform (by electroporation) *E. coli* BL21 (DE3) cells (Novagen). Another pool of BL21(DE3) cells were transformed with mock pET32a plasmid and used as negative control for the recombinant expression of  $\alpha$ -chain-thioredoxin-His<sub>6</sub> fusion proteins.

The nucleotide sequences of the synthetic gene coding for the  $\alpha$ -chains of the lectins from *Dioclea guianensis* (r- $\alpha$ Dguia) and *Dioclea grandiflora* (r- $\alpha$ DGL) have been deposited with the EMBL Nucleotide Sequence Database and are accessible under accession codes **AM701772** and **AM701773**, respectively.

#### ***Generation of r- $\alpha$ Dguia and r- $\alpha$ DGL mutants***

Site-directed mutagenesis was performed essentially as described in the QuickChange Site-Directed Mutagenesis Kit of Stratagene (La Jolla, CA). To this end, the pET-32a plasmids containing the r- $\alpha$ Dguia and r- $\alpha$ DGL sequences flanked by *NcoI* and *NotI* restriction sites were used as templates for PCR amplification (initial denaturation at 94 °C for 2 min, followed by 12 cycles of denaturation (30 s at 94 °C), annealing (60 s at 55 °C), and extension (12 min at 68 °C), and a final extension for 10 min at 68 °C) using the primers described in Table 1. The mutated plasmids were sequenced to confirm the absence of undesired mutations.

### ***Recombinant expression of r- $\alpha$ -thioredoxin-His<sub>6</sub> fusion proteins***

Positive *E. coli* BL21 (DE3) clones, shown by PCR to contain the r- $\alpha$ -chain-thioredoxin-His<sub>6</sub> fusion constructs, were grown overnight at 37 °C in LB medium containing 100  $\mu$ g/ml of ampicillin. For inducing the expression of the recombinant fusion proteins, the cell cultures were diluted 1:10 (v/v) with medium, isopropyl- $\beta$ -D-thiogalactosidase (IPTG) was added to a final concentration of 1 mM, and the cell suspensions were incubated for another 24 h at 14°C. Thereafter, the cells were pelleted by centrifugation, resuspended in the same volume of 20 mM sodium phosphate, 150 mM NaCl, pH 7.4, washed three times with the same buffer, and resuspended in 100 ml/liter of initial cell culture of 20 mM sodium phosphate, 250 mM NaCl, 10 mM imidazole, pH 7.4. The cells were lysed by sonication (15 cycles of 15 s sonication followed by 1 min resting) in an ice bath. The lysates were centrifuged at 10,000g for 30 min at 4 °C, and the soluble and the insoluble fractions were analyzed by SDS-15% polyacrylamide gel electrophoresis.

### ***Purification of recombinant alpha chains of Dioclea guianensis and Dioclea grandiflora seed lectins***

The  $\alpha$ -chain-thioredoxin- His<sub>6</sub> fusion proteins (r- $\alpha$ Dguia or r- $\alpha$ DGL) were purified from the soluble fraction of positive *E. coli* BL21 (DE3) clones by affinity chromatography using an ÄKTA Basic chromatographer equipped with a 5-ml HisTrap HP column (Amersham Biosciences) equilibrated in 20mM sodium phosphate, 250mM NaCl, 10 mM imidazole, pH 7.4. After the absorbance at 280 nm of the flowthrough fraction reached baseline, the bound material was eluted at a flow rate of 1.5 ml/min with a linear gradient of 10–500 mM imidazole for 60 min. The purified protein

fractions (checked by SDS-PAGE) were pooled, dialyzed against 50mM Tris/HCl, pH 7.4, and digested with 0.5 units of enterokinase (Invitrogen) per mg of recombinant protein. The reaction mixture was freed from enterokinase by chromatography on a 0.5-ml column of agarose-trypsin inhibitor (Sigma) equilibrated and eluted with 50 mM Tris-HCl, pH 7.4. The r- $\alpha$ -chain was separated from thioredoxin-His6 by affinity chromatography on a Sephadex G-75 column (2.0 x 7.0 cm) equilibrated in 50 mM Tris-HCl, pH 7.4 containing 150 mM NaCl, 5 mM CaCl<sub>2</sub> and 5 mM MnCl<sub>2</sub>. After exhaustive washing with equilibration buffer, the bound material (carbohydrate-binding r- $\alpha$ -chain) was eluted with equilibration buffer containing 200mM D-glucose, dialyzed for 30 minutes against 0.1M acetic acid, overnight against Milli-Q water, and lyophilized. The purity of the isolated proteins was assessed by SDS-PAGE.

### **Analytical ultracentrifugation**

The apparent molecular mass of the recombinant lectins in solutions of different pH were determined by analytical ultracentrifugation at 20 °C using a Beckman XL-A centrifuge with UV absorption scanner optics using an AN-50 Ti 8-hole rotor and charcoal filled eppon 6-channel centrepieces. This setup allows the simultaneous analysis of 21 different samples. The lectins were dissolved at 0.6-1.0 mg/ml in the following buffers containing 1 mM CaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub> and 0.1 M NaCl: Tris-HCl (for pH 7.5 and 8.5), 20 mM sodium citrate (for pHs 2.5, 3.5, and 4.5), and 20 mM MES (pH 5.5 or 6.5). Molar masses were determined by sedimentation-diffusion equilibrium experiments using short (approx. 3 mm) sedimentation columns. To avoid differences in apparent molecular masses due to rotor speed-dependent weighting of apparent molecular masses, all experiments were carried out at the same speed (15.000 rpm). Sedimentation equilibrium was assumed to be attained when the measured concentration profile remained unchanged for at least 12 h. The equilibrium concentration gradient for a single species is described by equation:

$$\ln c(r) - \ln c(m) = [M(1 - \bar{v}\rho) / 2RT] \omega^2 \times (r^2 - m^2),$$

where  $c(r)$  and  $c(m)$  are the concentrations at radius  $r$  and at the meniscus (radius  $m$ ), respectively;  $M$  is the molar mass of the solute;  $\bar{v}$ , partial specific volume of the solute (assumed to be  $7.35 \times 10^{-3} \text{ m}^3 \times \text{kg}$ );  $\omega$ , angular speed of the rotor (radians  $\times \text{s}^{-1}$ ), and  $R$  and  $T$ , are the gas constant ( $8.314 \text{ J mol}^{-1} \text{ K}^{-1}$ ) and the temperature (in K), respectively.

Apparent molar masses were determined by fitting this function to the measured radial

distribution of the concentration gradient at equilibrium using the program EQASSOC [33] provided by the manufacturer. Blank buffer absorption was determined after overspeeding (50,000 rpm) to sediment all material to the bottom of the cell.

### ***Crystallization, data collection and model building***

Crystals of r- $\alpha$ Dguia were grown at 22 °C by the vapour diffusion method using hanging drops composed of equal volumes of protein solution (10-15 mg/ml in 0.1 M MES, pH 6.0, containing 10 mM CaCl<sub>2</sub>, 10 mM MnCl<sub>2</sub>, 3mM 5-bromo-4-chloro-3-indolyl- $\alpha$ -D-Manose (X-Man, stock solution, 12 mM in DMSO)), and of the reservoir solution (30% polyethylene glycol 400, 100 mM MES, pH 6.4, 100 mM CdCl<sub>2</sub>) used as precipitant.

Crystals of r- $\alpha$ Dguia S131H were obtained as above except that the reservoir solution was 600 mM NaCl, 100 mM HEPES, pH 7.0.

For crystallization of r- $\alpha$ DGL and its triple mutant E123A-H131N-K132N, the lectins were dissolved at about 10-12 mg/ml in 50 mM HEPES, pH 7.5, containing 3 mM X-Man and 10 mM of each MnCl<sub>2</sub> and CaCl<sub>2</sub>. Crystals of r- $\alpha$ DGL were grown at constant room temperature (20 °C) by the hanging drop vapour diffusion method using 15% polyethylene glycol 4000, 100 mM HEPES (pH 7.5), and 100mM lithium sulfate as precipitant. Crystals of triple mutant were obtained with the same method but using 15% PEG 8000, 100 mM cacodylate, pH 6.5, containing 200 mM zinc acetate as precipitant.

Reservoir buffer containing 20-25% glycerol proved to be a suitable cryoprotectant. The crystals were transferred directly from the drop to the cryoprotectant solution and were allowed to equilibrate for approximately 10 seconds. Thereafter, single crystals were mounted in cryo-loops and rapidly transferred to the cryostream. Cryoprotected crystals were analyzed using Synchrotron Radiation Source at the ESRF (Grenoble, France) on beam lines ID23-1 (r- $\alpha$ Dguia), BM14 (r- $\alpha$ Dguia S131H), ID23-2 (r- $\alpha$ DGL), and BM16 (r- $\alpha$ DGL triple mutant E123A/H131N/K132N).

Crystallographic data were indexed, integrated and scaled using MOSFLM [34] and SCALA [35]. The crystal structures were determined by molecular replacement using the MOLREP program [36]. The search models used were *D. guianensis* seed lectin structures (PDB code 1H9P for r- $\alpha$ Dguia and 1H9W for r- $\alpha$ Dguia S131H) [28] and *D. grandiflora* seed lectin (1DGL) [27] for r- $\alpha$ DGL. The crystal structure of the triple

mutant r- $\alpha$ DGL E123A-H131N-K132N was determined using the refined structure of r- $\alpha$ DGL. The position and orientation of the molecules as a single rigid body entity were refined for 20 cycles with REFMAC [CCP4] [35]. Several steps of rebuilding, interspersed with restrained refinement using REFMAC, yielded the final models at resolutions ranging from 1.65 to 2.1 Å (Table 2). The X-Man molecule was placed by inspection of the  $F_o - F_c$  maps. Finally, water molecules were placed in the model over several steps of refinement with Arp/Warp and inspected manually. The atomic coordinates, fitted with COOT [37], are accessible from the Protein DataBank (<http://www.rcsb.org/pdb/>) under codes **2JDZ** (r- $\alpha$ Dguia), **2JE7** (r- $\alpha$ Dguia S131H), **2JE9** (r- $\alpha$ DGL) and **2JEC** (r- $\alpha$ DGL E123A-H131N-K132N).

## RESULTS AND DISCUSSION

In this paper we have primarily addressed the question of the role of interdimeric interactions at the central cavity for conferring pH-dependent dimer-tetramer transition upon Diocleinae lectins. To this end it was clearly necessary to produce recombinant lectins. To circumvent the inability of *E. coli* of post-translationally processing pre-pro Diocleinae lectins as in the plant system (discussed below), we have designed and cloned synthetic genes coding for mature wild-type and mutated single-polypeptide  $\alpha$ -chains. This novel approach, which may inspire molecular biology studies of other lectins, was successfully employed for producing recombinant wild-type and mutated proteins biochemically indistinguishable from the natural occurring seed lectins, and is described in the following sections.

### **The *Dioclea guianensis* and *Dioclea grandiflora* lectin precursors are coded for by intronless genes**

PCR using primers designed from the cDNA sequences of the highly-related concanavalin A [38] and ConBr [39] lectin precursors failed to amplify any product using cDNA or genomic DNA of either *D. guianensis* and *D. grandiflora*. However, using degenerated primers based on highly conserved amino acid sequences among legume lectins, a 500 bp product was PCR-amplified (Fig.1). This cDNA-amplified nucleotide sequence was then used to design specific primers to extend the *D. guianensis* transcript. the Dguia cDNA sequence was extended in the 3' direction by



3'RACE. The upstream Dguia gene sequence was determined by nested PCR using RsaI-digested genomic DNA as template. The 3'RACE- (cDNA) and 5'-extension (gDNA) gathered nucleotide sequences were employed to construct specific primers, which were then used to amplify a fragment of 876 bps encoding the full-length *Dioclea guianensis* lectin open reading frame (Fig.1A). Identical sequence was amplified using cDNA and genomic DNA templates, clearly showing that the gene coding for the *Dioclea guianensis* seed lectin does not possess introns. Similarly, the seed lectins from *Canavalia ensiformis* [38], *Canavalia brasiliensis* [39], and *Canavalia gladiata* [40] have been previously reported to be coded for by intronless genes.

The DNA-deduced amino acid sequence of the Dguia precursor (Fig.1) departs at eight positions from the primary structure of the lectin isolated from *Dioclea guianensis* seeds [24, 28]. The eight conservative changes (Dguia(DNA)/Dguia(seed), Ser41/Asn, Gly54/Ser, Gln65/Glu, Arg68/Lys, Asp72/Ser, Ser74/Asp, Ser230/Thr, and Thr232/Ser in Fig.1A) were found in every clone sequenced, ruling out a DNA sequencing error. None of these residues participates in interdimeric contacts in the crystal structure of the Dguia seed lectin [28]. Noteworthy, the isotope-averaged molecular masses calculated for the full-length Dguia  $\alpha$ -chain (25395 Da) and its derived beta (12832 Da) and gamma (12584 Da) fragments are in excellent agreement with those determined by electrospray-ionization mass spectrometry for the Dguia seed lectin [24], suggesting that the cloned and the seed-isolated lectins may be isobaric isolectins.

The DNA sequence coding for the *Dioclea grandiflora* seed  $\alpha$ -chain lectin (Fig.1B) was PCR-amplified using genomic DNA as template and forward and reverse primers designed from the DNA sequence of the Dguia lectin precursor. The same 783 bp product was amplified using cDNA as template, indicating that, like the Dguia lectin, the *D. grandiflora* lectin gene does not include any intronic sequence. The calculated and the experimentally-determined masses of the DGL alpha- (25607 Da), beta- (12872 Da) and gamma- (12752 Da) chains [24] coincide within the experimental error limit ( $\pm 2$  Da).

### **Design and cloning of synthetic genes coding for the $\alpha$ -chains of the seed lectins of *Dioclea guianensis* and *Dioclea grandiflora***

Maturation of *Diocleinae* seed lectins involves a complex mechanism of post-translational processing that includes deglycosylation-dependent proteolytic cleavage of



the precursor at the carboxyl site of internal Asn residues, and religation of the resulting gamma and beta fragments in inversed DNA-coded order [38, 41]. Thus, the resulting active  $\alpha$ -chain ( $\alpha = \beta\text{-}\gamma$ ) is circularly permuted in primary sequence relative to its own inactive  $\gamma\text{-}\beta$  precursor. Proteolytic cleavage and religation of the fragments to form the  $\alpha$ -chain are both catalyzed by an asparaginyl endopeptidase [42]. The nature of the splicing and religation events has been controversial, with the report of Yamauchi and Minamikawa [40] that pre-pro-ConA from *Canavalia gladiata* expressed in *Escherichia coli* underwent peptide cleavage and ligation in the same way as the lectin synthesized during seed maturation, contrasting with a lack of evidence for any processing of pre-pro-ConA in the cytoplasm [43, 44] or of pre-ConA in the periplasm of *E. coli* [45]. Indeed, *E. coli* does not express asparaginyl endopeptidase and is therefore unable to process the legume lectin precursor as in the plant [43]. Thus, to produce native-like and active *Dioclea guianensis* and *Dioclea grandiflora* seed lectins, we have designed and cloned synthetic genes coding for the single-polypeptide  $\alpha$ -chains. The cloning strategy employed for constructing synthetic genes for the  $\alpha$ -chains from the DNA-amplified precursors of the *D. guianensis* and *D. grandiflora* seed lectins, schematically displayed in Fig.2, included PCR engineering of *EcoRI* restriction sites at the 3' of the  $\beta$ -chain and at the 5' of the  $\gamma$ -chain. The  $\beta$ /pGEM-T and  $\gamma$ /pGEM-T constructs were then serially digested with *EcoRI/NcoI* and *EcoRI/NotI*, respectively, generating complementary cohesive ends that were subsequently ligated into a full-length  $\alpha$ -chain flanked by 5'-*NcoI* and 3'-*NotI* restriction sites. The synthetic  $\alpha$ -chain was cloned into an open pET-32a expression vector digested with *NcoI* and *NotI* creating an  $\alpha$ -chain-thioredoxin-His<sub>6</sub> fusion protein.

### **Recombinant expression in *E. coli* of synthetic and active $\alpha$ -Dguia and $\alpha$ -DGL**

Recombinant synthetic  $\alpha$ -chains of *Dioclea guianensis* and *Dioclea grandiflora* lectins were expressed as thioredoxin-His<sub>6</sub>-tagged fusion proteins in *E. coli* BL21 (DE3) cells. Biosynthesis of the recombinant construct was independent of the addition of the Lac operon inducer, isopropyl  $\beta$ -D-thiogalactosidase, and ~90% of the fusion protein was recovered in the soluble cell lysate (Fig.3). Following affinity chromatography in HisTrap and digestion with enterokinase, r- $\alpha$ Dguia and r- $\alpha$ DGL were purified to apparent homogeneity by affinity chromatography on Sephadex G-75. The average protein yield was similar for both lectins, about 20 mg/L. The recombinant Dguia and

DGL  $\alpha$ -chains were quantitatively retained in the cross-linked dextran gel column and were desorbed with D-glucose, providing strong evidence that the lectins displayed native-like, carbohydrate-binding activity. In line with this assumption, r- $\alpha$ Dguia and r- $\alpha$ DGL also exhibited identical hemagglutinating activity than the corresponding lectins isolated from seeds (not shown).

As judged by sedimentation equilibrium, r- $\alpha$ Dguia exhibited similar pH-dependent dimer-tetramer transition than the *D. guianensis* seed lectin (Fig.4). On the other hand, r- $\alpha$ DGL, like the lectin isolated from *D. grandiflora* seeds, behave as a pH-independent tetrameric protein (Table 3). Furthermore, r- $\alpha$ Dguia and r- $\alpha$ DGL were crystallized and their structures, solved to 2.1 Å resolution (Table 2), can be superimposed with those of their natural homologues with overall r.m.s. deviations for all C $_{\alpha}$  atoms of 0.33/0.45 Å (monomer/tetramer, r- $\alpha$ DGL/1DGL) and 0.35/0.56 Å (monomer/tetramer, r- $\alpha$ Dguia/1H9P).

Similarly to what has been reported for the seed lectin from *Dioclea guianensis* [28], the r- $\alpha$ Dguia crystal structure displayed poor electronic density for the central cavity loop residues 117-123. This observation indicates that disorder of the corresponding loop in the natural lectin is not due to the incomplete ligation of the C-terminal residue of the  $\beta$ -chain and the N-terminal residue of the  $\gamma$ -fragment that leads to the mixture of  $\alpha$ ,  $\beta$ , and  $\gamma$  chains.

The architecture of the primary monosaccharide recognition site and the interactions established between the protein and the mannose residue of X-Man (Fig.5 panels A-C, Table 4) are essentially the same as in other *Diocleinae* lectin-sugar complexes. The quality of the electron density maps around the X-Man moiety is shown in Fig. 5C. Though the apo (1H9P, 1H9W) and the sugar-bound (r- $\alpha$ Dguia, 2JDZ) wild-type *D. guianensis* lectins possess identical preformed carbohydrate-recognition motifs, the average B-factor for the main chain of the mannose-binding loops of apo-Dguia (52 Å<sup>2</sup>) is significantly higher than that for the r- $\alpha$ Dguia/X-Man complex (34 Å<sup>2</sup>) (Fig.5, panels D and E), suggesting that binding of the brominated monosaccharide derivative restrains the conformational flexibility of the sugar-binding, solvent-exposed loops. The low average B-factors of the mannose-binding loops of r- $\alpha$ Dguia S131H/X-Man (20.9 Å<sup>2</sup>), r-DGL/X-Man (26.0 Å<sup>2</sup>) and r- $\alpha$ Dguia E123A-H131N-K132Q/X-Man (19.0 Å<sup>2</sup>) (Table 2) further support this conclusion. Similar observation has been reported for the

seed lectin from *Parkia platycephala* when the apo- and the X-Man-complexed structures were compared [46].

Taken together, our data indicate that the recombinant lectins are structurally and biochemically indistinguishable from their natural homologues.

### **Conversion by site-directed mutagenesis of r- $\alpha$ DGL into a lectin exhibiting pH-dependent dimer-tetramer transition**

The amino acid sequences of the recombinant seed lectins from *Dioclea guianensis* and *Dioclea grandiflora* depart in only 11 residues, of which 3 are located in the 123-132 region: rDgua, <sup>121</sup>ADANSLHFSFSQFS<sup>134</sup>; rDGL, <sup>121</sup>ADENSLHFSFHKFS<sup>134</sup>. Previous studies [28] have suggested that residues within this region of 1DGL participate the interdimeric contacts that yield a pH-independent tetrameric structure. To gain a closer insight into the mechanism governing the pH-dependent dimer-tetramer transition, we have produced singly, doubly, and triply mutated recombinant lectins and have assessed their structure-activity correlations by combination of equilibrium sedimentation analytical ultracentrifugation and X-ray crystallography.

The crystal structures of r- $\alpha$ Dgua, r- $\alpha$ Dgua S131H, r- $\alpha$ DGL, and r- $\alpha$ DGL E123A/H131N/H132K can be superimposed with r.m.s. deviations for C $\alpha$  atoms of 0.3-0.7 Å, being interdimeric hydrogen bonding contacts between R60 and D78 of opposed monomers at the peripheral sites essentially conserved (Fig.6), whereas the major differences appear to be associated with the central cavity loops (Fig.7, panel A).

Mutant r- $\alpha$ Dgua S131H exhibited dimer-tetramer equilibrium as the wild-type (Fig.4). This result was unexpected since a previous work strongly suggested that a histidine at position 131 was responsible for stabilizing the pH-independent tetrameric structure of the *Dioclea grandiflora* lectin [28]. However, inspection of the crystal structure of the S131H mutant showed that H131 established contact with D122 from an adjacent monomer of the canonical dimer (Fig.7, panel B). The absence of interactions between H131 and the central cavity loop residues, as in r- $\alpha$ DGL (Fig.7, panel C) from an opposed monomer explains why the mutation did not locked r- $\alpha$ Dgua S131H into a pH-independent tetramer.

Also unexpected, the single mutant r- $\alpha$ DGL H131N and the double mutants r- $\alpha$ DGL E123A/H131N and r- $\alpha$ DGL H131N/K132Q, all exhibited the same equilibrium sedimentation behavior than wild-type r- $\alpha$ DGL (Fig.4), indicating that mutating H131

to asparagine, *per se*, does generate a pH-dependent dissociable tetramer. On the other hand, pH-dependent tetramer dissociation was achieved with the double mutant r- $\alpha$ DGL E123A/K132Q and the triple mutant r- $\alpha$ DGL E123A/H131N/K132Q (Fig.4). The r- $\alpha$ DGL E123A/K132Q mutant is equivalent to r- $\alpha$ Dguia S131H discussed above.

The crystal structure of r- $\alpha$ DGL E123A/H131N/K132Q was solved to a resolution of 2.0 Å (Table 2). This structure and that of the *Dioclea grandiflora* seed lectin (1DGL) [27] can be superimposed with r.m.s. deviation of 0.56 Å for all C $\alpha$  atoms, excepting for the central cavity loops (residues 117-123), where significant conformational differences were noticed (Fig.7, panel D). Thus, in the triple mutant structure, the polypeptide stretch 117-123 folds into a well-defined loop, which adopts a less extended projection towards the center of the central cavity than the corresponding loop in 1DGL (Fig.7, panel D). The reason for this different loop 117-123 conformation in the triple mutant and in 1DGL lie in the distinct intradimeric interactions established by residues 123, 131, and 132 in each structure (Fig.7, panels E and F). In 1DGL, O $\epsilon$ 2 of E123 (from monomers D and A) interacts with NZ of K132 (C/B), and the main-chain oxygen of E123 is hydrogen-bonded to the main-chain nitrogens of residues H131(C/B) and K132(C/B). In the triple mutant, the C $\alpha$  residues of D122 and A123 are displaced 3.34 Å and 3.44 Å, respectively, from the position occupied in r- $\alpha$ DGL, the amide oxygen of Q132 and N131(C/B) contact, respectively, ND2 and OD1 of N124(D/A), and ND2 of N131(C/B) makes a hydrogen bond with the main-chain oxygen of D122(D/A). It is worth to notice that the conformation of residues <sup>122</sup>DAN<sup>124</sup> in the r- $\alpha$ DGL E123A/H131N/K132Q structure mirrors that adopted by the same residues in the crystal structure of the *Dioclea guianensis* seed lectin (1H9P) [28]. However, the remaining of the central loop (residues 117-121) is disordered in hindering any structural comparison between the triply mutated r- $\alpha$ DGL and 1H9P. Nevertheless, the consequence of the structural alteration in the triply mutated loop structure renders it unable to establish the interdimeric tetramer-stabilizing interactions observed in the wild-type *Dioclea grandiflora* seed lectin.

### **Concluding remarks**

Evidence gathered from comparison of the crystal structures of lectins isolated from seeds of *Diocleinae*, which exhibit (ConA [1NLS], ConBr [1AZD], Dguia [1H9P], CFL [2D3P, 2D3R], *Cratylia mollis* lectin [1MVQ], CanGla [1WUV], CanMar [2CWM]) or

not (DGL [1DGL], *Dioclea violacea* lectin [47]) dimer-tetramer transition in the pH range 4.5-8.5, have indicated that the occurrence of interdimeric interactions between residues at positions 131 and 120 of opposite monomers locks the quaternary structure of the lectin into a pH-independent tetramer. Similarly, conversion of the pH-stable tetrameric r- $\alpha$ DGL into a structure exhibiting pH-dependent dimer-tetramer equilibrium was accomplished through mutations that abolished the interdimeric interactions at the central cavity. In addition to these "central" loop contacts, other interdimer interactions are established between H51 of monomers A and B and residues at positions 116 and 117 of monomers D and C, respectively, and between homologous regions encompassing residues 53-78 of monomers A and D and B and C at the periphery of the dimers. Both, the central and the peripheral interacting regions may bear information for formation of the canonical legume lectin tetramer (Fig.4). The "peripheral" interfases include a network of interactions orchestrated by residues N55, R60, D78, which are absolutely conserved in legume lectins that adopt the cross-wise tetrameric association classified as type II by Brinda and coworkers [48]. Our working hypothesis is that the strength of the ionic contacts established by H51, R60 and D78 may be modulated by the pH, leading to protonation-dependent tetramer dissociation of those lectin structures that are not stabilized through interdimeric interactions networking the central cavity loops of the two dimers.

#### ACKNOWLEDGEMENTS

This work has been financed in part by grant BFU2004-01432/BMC from the Ministerio de Educación y Ciencia, Madrid, Spain (to J.J.C.). C.S.N. is recipient of a Ph.D. fellowship from the Coordenação Aperfeiçoamento de Pessoal de Nivel Superior (CAPES/MEC, Brazil). B.S.C is a senior investigator of CNPq (Brazil).

## REFERENCES

1. Van Damme, E.J.M., Peumans, W.J., Barre, A. and Rougé, P. (1998). Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit. Rev. Plant Sci.* **17**, 575-692.
2. Peumans, W.J., Van Damme, E.J., Barre, A. and Rougé, P. (2001). Classification of plant lectins in families of structurally and evolutionary related proteins. *Adv. Exp. Med. Biol.* **491**, 27-54.
3. Gabius, H-J. and Gabius, S. (eds.) (1997). *Glycoscience. Status and Perspectives.* Chapman & Hall, Weinheim (Germany).
4. Gabius, H-J., Siebert, H.C., André, S., Jiménez-Barbero, J. and Rüdiger, H. (2004). Chemical biology of the sugar code. *Chembiochem.* **5**, 740-764.
5. Loris, R. (2002) Principles of structures of animal and plant lectins. *Biochim. Biophys. Acta* **1572**, 198-208.
6. Elgavish, S. and Shaanan, B. (1997). Lectin-carbohydrate interactions: different folds, common recognition principles. *Trends Biochem. Sci.* **22**, 462-467.
7. Drickamer, K. (1995). Multiplicity of lectin-carbohydrate interactions. *Nature Struct. Biol.* **2**, 437-439.
8. Rini, J.M. (1995). Lectin structure. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551-577.
9. Bouckaert, J., Hamelryck, T., Wyns, L. and Loris, R. (1999) Novel structures of plant lectins and their complexes with carbohydrates. *Curr. Op. Struct. Biol.* **9**, 572-577.
10. Vijayan, M. and Chandra, N. (1999). Lectins. *Curr. Op. Struct. Biol.* **9**, 707-714.
11. Prabu, M.M., Suguna, K. and Vijayan, M. (1999). Variability in quaternary association of proteins with the same tertiary fold: a case study and rationalization involving legume lectins. *Struct. Funct. Genet.* **35**, 58-69.
12. Chandra, N., Prabu, M.M., Suguna, K. and Vijayan, M. (2001). Structural similarity and functional diversity in proteins containing the legume fold. *Prot. Engineering* **14**, 857-866.
13. Srinivas, V.R., Reddy, G.B., Ahmad, N., Swaminathan, C.P., Mitra, N. and Surolia, A. (2001). Legume lectin family, the 'natural mutants of the quaternary state', provide insights into the relationship between protein stability and oligomerization. *Biochim. Biophys Acta* **1527**, 102-111.
14. The 3D Lectin Database at <http://www.cermav.cnrs.fr/lectines>
15. Loris, R., Hamelryck, T., Bouckaert, J. and Wyns, L. (1998). Legume lectin structure. *Biochim. Biophys. Acta* **1383**, 9-36.



16. Chervenak, M.C. and Toone, E.J. (1995). Calorimetric analysis of the binding of lectins with overlapping carbohydrate-binding ligand specificities. *Biochemistry* **34**, 5685-5695.
17. Dam, T.K. and Brewer, C.F. (2002). Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem. Rev.* **102**, 387-429.
18. Edelman, G.M., Cunningham, B.A., Reece, G.N., Becker, J.W., Waxdal, J.M. and Wang, J.L. (1972). The covalent and three-dimensional structure of concanavalin A. *Proc. Natl. Acad. Sci.* **69**, 2580-2584.
19. Hardman, K.D. and Ainsworth, C.F. (1972). Structure of concanavalin A at 2.4 Å resolution. *Biochemistry* **11**, 4910-4919.
20. Reece, G.N., Becker, J.W. and Edelman, G.M. (1975). The covalent and three-dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding, and quaternary structure. *J. Biol. Chem.* **250**, 1525-1547.
21. Huet, M. and Claverie, J.M. (1978). Sedimentation studies of the reversible dimer-tetramer transition kinetics of concanavalin A. *Biochemistry* **17**, 236-241.
22. Senear, D.F. and Teller, D.C. (1981). Thermodynamics of concanavalin A dimer-tetramer self-association: sedimentation equilibrium studies. *Biochemistry* **20**, 3076-3083.
23. Senear, D.F. and Teller, D.C. (1981). Effects of saccharide and salt binding on dimer-tetramer equilibrium of concanavalin A. *Biochemistry* **20**, 3083-3091.
24. Calvete, J.J., Thole, H.H., Raida, M., Urbanke, C., Romero, A., Grangeiro, T.B., Ramos, M.V., Rocha, I.M.A., Guimarães, F.N. and Cavada, B.S. (1999). Molecular characterization and crystallization of *Diocleinae* lectins. *Biochim. Biophys Acta* **1430**, 367-375.
25. Edelman, G.M. and Wang, J.L. (1978). Binding and functional properties of concanavalin A and its derivatives. III. Interactions with indoleacetic acid and other hydrophobic ligands. *J. Biol. Chem.* **253**, 3016-3022.
26. Roberts, D.D. and Goldstein, I.J. (1983). Adenine binding sites of the lectin from lima beans (*Phaseolus lunatus*). *J. Biol. Chem.* **258**, 13820-13824.
27. Rozwarski, D.A., Swami, B.M., Brewer, C.F. and Sacchettini, J.C. (1998). Crystal structure of the lectin from *Dioclea grandiflora* complexed with core trimannoside of asparagine-linked carbohydrates. *J. Biol. Chem.* **273**, 32818-32825.
28. Wah, D.A., Romero, A., Gallego del Sol, F., Cavada, B.S., Ramos, M.V., Grangeiro, T.B., Sampaio, A.H. and Calvete, J.J. (2001). Crystal structure of native and Cd/Cd-

- substituted *Dioclea guianensis* seed lectin. A novel manganese-binding site and structural basis of dimer-tetramer association. *J. Mol. Biol.* **310**, 885-894.
29. Gallego del Sol, F., Cavada, B.S. and Calvete, J.J. (2007). Crystal structures of *Cratylia floribunda* seed lectin at acidic and basic pHs. Insights into the structural basis of the pH-dependent dimer-tetramer transition. *J. Struct. Biol.* **158**, 1-9.
  30. Steenkamp, J., Wiid, I., Lourens, A. and van Helden, P. (1994). Improved method for DNA extraction from *Vitis vinifera*. *Am. J. Enol. Vitic.* **45**, 102-106.
  31. Frohman, M.A. and Martin, G.R. (1989). Rapid Amplification of cDNA ends using nested primers. *Techniques* **1**, 165-170.
  32. Siebert P.D., Chenchik A., Kellogg D.E., Lukyanov K.A. and Lukyanov S.A. (1995). An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* **23**, 1087-1088.
  33. Minton, A. P. (1994). *Modern Analytical Ultracentrifugation* (Schuster, M. and Laue, T.M., eds.) Birkhauser, Boston (USA), 81-93.
  34. Leslie, A.G.W. (1997) *Mosflm User's Guide: Mosflm Version 5.50*, MRC Laboratory of Molecular Biology, Cambridge, UK.
  35. Collaborative Computational Project, Number 4. (1994) *Acta Cryst.* **D50**, 760-763.
  36. Vagin, A. and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022-1025.
  37. Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr.* **D60**, 2126-2132.
  38. Carrington, D.M., Auffret, A. and Hanke, D.E. (1985). Polypeptide ligation occurs during post-translational modification of concanavalin A. *Nature* **313**, 64-67.
  39. Grangeiro T.B., Schriefer, A., Calvete, J.J., Raida, M., Urbanke, C., Barral-Netto, M. and Cavada, B.S. (1997). Molecular cloning and characterization of ConBr, the lectin of *Canavalia brasiliensis* seeds. *Eur. J. Biochem.* **248**, 43-48.
  40. Yamauchi, D. and Minamikawa, T. (1990). Structure of the gene encoding concanavalin A from *Canavalia gladiata* and its expression in *Escherichia coli* cells. *FEBS Lett.* **260**, 127-130.
  41. Chrispeels, M.J., Hartl, P.M., Sturm, A. and Faye, L. (1986). Characterization of the endoplasmic reticulum-associated precursor of concanavalin A. Partial amino acid sequence and lectin activity. *J. Biol. Chem.* **261**, 10021-10024.
  42. Min, W. and Jones, D.H. (1994). *In vitro* splicing of concanavalin A is catalyzed by asparaginyl endopeptidase. *Nature Struct Biol.* **1**, 502-504.

43. Min, W. and Jones, D.H. (1992). Stability and detection of recombinant pre-pro-concanavalin A after cytoplasmic expression in *Escherichia coli*. FEBS Lett. **301**, 315-318.
44. Dinkturk, H.B., Dunn, A.J. and Jones, D.H. (2001) Recombinant pre-pro-Concanavalin A (jack bean) is stable but ay low solubility. J. Biosci. **26**, 635-640.
45. Min, W., Dunn, A.J. and Jones, D.H. (1992) Non-glycosylated recombinant pro-concanavalin A is active without polypeptide cleavage. EMBO J. **11**, 1303-1307.
46. Gallego del Sol, F., Nagano, C.S., Cavada, B.S. and Calvete, J.J. (2005). The first crystal structure of a Mimosoideae lectin reveals a novel quaternary arrangement of a widespread domain. J. Mol. Biol. **353**, 574-583.
47. Gallego del Sol, F. (2006) Estudios Estructurales de Lectinas Vegetales. Ph.D. Thesis, University of Valencia.
48. Brinda, K.V., Mitra, N., Surolia, A. and Vishveshwara, S. (2004). Determinants of quaternary association in legume lectins. Protein Sci. **13**, 1735-1749.

## FIGURE LEGENDS

**Fig.1.** (A) Nucleotide and deduced amino acid sequence of the full-length ORF encoding the precursor of the *Dioclea guianensis* seed lectin. The signal peptide sequence (residues 1-28) is underlined. The amino acid sequences of the gamma- (29-147) and the beta- (163-280) chains are highlighted. The eight codons coding for different amino acid residues in Dguia DNA compared to the seed-isolated lectin are highlighted in grey boxes. The degenerated nucleotide sequences of the forward (F) and reverse (R) primers, based on highly conserved amino acid sequences among legume lectins, used to amplify a specific 500 bp DNA product are shown. Nucleotide sequence (837-840) used to engineer an *EcoRI* restriction site is underlined. (B) Nucleotide and deduced amino acid sequence of the precursor of the *Dioclea grandiflora* seed lectin. The forward (F) and reverse (R) primers used for amplifying the full-length DNA coding for the full-length  $\alpha$ -chain are shown.

**Fig.2.** Strategy followed for construction of synthetic *Dioclea guianensis* and *Dioclea grandiflora* lectin  $\alpha$ -chains. After PCR-engineering of *EcoRI* restriction sites at the 3' of the  $\beta$ -chain and at the 5' of the  $\gamma$ -chain, the  $\beta$ - and the  $\gamma$ -fragments were separately cloned into pGEM-T vectors containing, respectively, *NcoI* and *NotI* restriction sites. The vectors were serially digested, the fragments were purified by agarose gel electrophoresis, the  $\gamma$ -chain was dephosphorylated, and the  $\beta$  and  $\gamma$  chains were ligated through their *EcoRI*-created cohesive ends. Synthetic  $\alpha$ -chains were cloned into open expression vectors pET-32a, and the constructs were expressed as  $\alpha$ -chain-thioredoxin-His<sub>6</sub> fusion proteins in *E. coli*.

**Fig.3.** SDS-(12.5%)-polyacrylamide gel electrophoretic analysis of the recombinant expression and the purification of the synthetic  $\alpha$ -chain of *Dioclea guianensis* lectin. The same result was obtained with the synthetic *Dioclea grandiflora* lectin construct. Lane **a**, cell lysates of *E. coli* BL21 (DE3) cells transformed with the mock pET32a plasmid. Lanes **b** and **c**, 100  $\mu$ g of total proteins of the soluble and insoluble fractions, respectively, of the cell lysate of BL21 (DE3) cells expressing the  $\alpha$ Dguia-thioredoxin-His<sub>6</sub> fusion protein. Lane **d**, HisTrap affinity-purified  $\alpha$ Dguia-thioredoxin-His<sub>6</sub> construct. Lane **e**, protein mixture generated by digestion with enterokinase of the  $\alpha$ Dguia-thioredoxin-His<sub>6</sub> fusion protein. Lane **f**, r- $\alpha$ Dguia purified by affinity

chromatography on Sephadex G75. Lane **g**, Mature lectin isolated from *Dioclea guianensis* seeds exhibiting the typical  $\alpha$  (25 kDa),  $\beta$  (14 kDa), and  $\gamma$  (12 kDa) mixture. Lane **s**, molecular mass markers (Mark12, Invitrogen), from top to bottom: bovine serum albumin (66.3 kDa), glutamic dehydrogenase (55.4 kDa), lactate dehydrogenase (36.5 kDa), carbonic anhydrase (31.0 kDa), soybean trypsin inhibitor (21.5 kDa), lysozyme (14.4 kDa), and aprotinin (6 kDa).

**Fig.4.** Determination by equilibrium sedimentation analytical centrifugation of the apparent molecular masses of the recombinant lectins as a function of the pH of the solution. Apparent molecular masses ( $M_{app}$ ) between a pure dimer (D, 51 kDa) and a pure tetramer (T, 102 kDa) correspond to  $M_{app} = (\%D \times 51 + \%T \times 102)$  kDa. The location of the peripheral (P) and central (C) interdimeric regions are indicated in the structures of the canonical dimer and tetramer of r- $\alpha$ Dguia.

**Fig.5.** Comparison of the mannose binding-sites of (A) the r- $\alpha$ DGL/X-Man complex and (B) the *Dioclea grandiflora* seed lectin complexed with a trimannoside (1DGL) [27]. Broken lines represent hydrogen bonds. For detailed interactions between lectin and sugar atoms see Table 4. Panel C shows a portion of the 2Fo-Fc electron density map contoured at 1.0  $\sigma$  showing the X-Man bound to the carbohydrate-binding site of monomer A of the Dguia S131H mutant. Panels D and E display, respectively, close-up views of the carbohydrate-recognition domains of r- $\alpha$ Dguia with bound X-Man (2JDZ) [this work] and apo-Dguia (1H9P) [28]. B-factors are color-coded from 25  $\text{\AA}^2$  (green) to 60  $\text{\AA}^2$  (red).

**Fig.6.** Comparison of the interactions (hydrogen bonds, dashed lines) at the peripheral interdimeric regions between monomers A and D (see Fig.7A) of natural (1DGL) (A) and recombinant wild type (rDGL) (B) and triply mutated (rDGL E123A/H131N/H132K) (C) *Dioclea grandiflora* lectins. The latter exhibits pH-dependent dimer-tetramer transition, whereas 1DGL and rDGL are pH (4.5-8.5)-independent tetrameric structures (see Fig.4). Interactions between monomers B and C are essentially the same.

**Fig.7.** Crystal structures of recombinant lectins. Panel **A**, Superposition of the C $\alpha$  traces of recombinant lectins r- $\alpha$ Dguia (black), r- $\alpha$ Dguia S131H (red), r- $\alpha$ DGL (blue), and r- $\alpha$ DGL E123A/H131N/H132K (green). Panel **B**, detail of the interdimeric interactions at the central cavity of r- $\alpha$ DGL between H131 and I120 from an opposite monomer that render a pH-stable tetrameric structure. Panel **C**, detail of the intradimeric contacts at the central cavity of r- $\alpha$ Dguia S131H. Panel **D**, close-up view of the interdimeric region at the central cavity of r- $\alpha$ DGL (blue) and its triple mutant E123A/H131N/H132K (green). Panels **E** and **F**, comparison of the intradimeric hydrogen bond network in the crystal structures of r- $\alpha$ DGL (blue) and its triple mutant (green) showing the structural changes caused by the E123A/H131N/H132K mutations, which generate a 117-123 loop conformation less extended towards the center of the central cavity than the corresponding structure in r- $\alpha$ DGL (see also panel **D**).



**Table 1**

Forward (F) and reverse (R) primers used for generating the site-directed mutants of the  $\alpha$ -chains of the seed lectins of *Dioclea guianensis* (r- $\alpha$ -Dguia) and *Dioclea grandiflora* (r- $\alpha$ -DGL). Codons producing the mutations are in boldface and underlined.

Mutant	Template used	Primers (5'-3')
r- $\alpha$ -Dguia S131H	r- $\alpha$ -Dguia	F: AAT TCA CTC CAT TTC AGC TTC <b>CAC</b> CAG TTT AGC CAA AAC CCA AAG G R: GG GTT TTG GCT AAA CTG <b>GTG</b> GAA GCT GAA ATG GAG TGA ATT TGC
r- $\alpha$ -DGL H131N	r- $\alpha$ -DGL	F: TCA CTC CAT TTC AGC TTC <b>AAC</b> AAA TTT AGC CAA AAC CCA AAG G R: GG GTT TTG GCT AAA TTT <b>GTT</b> GAA GCT GAA ATG GAG TGA ATT TT
r- $\alpha$ -DGL E123A/H131N	r- $\alpha$ -DGL H131N	F: G ACG AAT TCA ATA GCA GAT <b>GCA</b> AAT TCA CTC CAT TTC AGC R: GCT GAA ATG GAG TGA ATT <b>TGC</b> ATC TAT TCA TAT TGA ATT CGT C
r- $\alpha$ -DGL H131N/K132Q	r- $\alpha$ -DGL H131N	F: CTC CAT TTC AGC TTC AAC <b>CAA</b> TTT AGC CAA AAC CCA AAG G R: CTT TGG GTT TTG GCT AAA <b>TTG</b> GTT GAA GCT GAA ATG GAG
r- $\alpha$ -DGL E123A/K132Q	r- $\alpha$ -DGL E123A/ H131N/K132Q	F: AAT TCA CTC CAT TTC AGC TTC <b>CAC</b> CAG TTT AGC CAA AAC CCA AAG G R: GG GTT TTG GCT AAA CTG <b>GTG</b> GAA GCT GAA ATG GAG TGA ATT TGC
r- $\alpha$ -DGL E123A/H131N/K132Q	r- $\alpha$ -DGL E123A/H131N	F: CTC CAT TTC AGC TTC AAC <b>CAA</b> TTT AGC CAA AAC CCA AAG G R: CTT TGG GTT TTG GCT AAA <b>TTG</b> GTT GAA GCT GAA ATG GAG

**Table 2**  
Summary of data collection and refinement statistics

	r- $\alpha$ Dguia	r- $\alpha$ Dguia S131H	r- $\alpha$ DGL	r- $\alpha$ DGL- E123A-H131N-K132Q
<i>A. Data collection statistics</i>				
Space group	I4 <sub>1</sub> 22	I222	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell constants				
a (Å)	88.2	65.6	72.2	72.2
b (Å)	88.2	88.0	85.2	84.5
c (Å)	106.1	91.5	175.7	176.2
$\alpha=\beta=\gamma$	90°	90°	90°	90°
Resolution(Å)	2.1	1.65	2.1	2.0
Total reflections	133,461	98,371	295,146	278,161
Unique reflections	12,834	31,757	63,973	73,352
Data coverage (%)	99.9 (100.0)*	98.8 (99.3)*	99.9 (99.7)*	99.6 (100.00)*
R <sub>merge</sub> (%) <sup>a</sup>	8.2 (30.3)	4.9 (36.2)	7.9 (30.3)	7.2 (25.2)
Multiplicity	10.4 (10.6)	3.1 (3.0)	4.6 (4.0)	3.8 (3.8)
<i>I</i> / $\sigma$ < <i>I</i> >	5.6 (2.1)	8.4 (2.1)	6.9 (2.1)	12.9 (3.0)
Molecules in asymmetric unit	monomer	monomer	tetramer	tetramer
<i>B. Refinement statistics</i>				
Resolution range (Å)	34.1-2.1	44.0-1.65	72.2-2.1	28.9-2.0
Reflections	11,848	30,142	60,684	69,591

THIS IS NOT THE FINAL VERSION - see doi:10.1042/BJ20070942

R <sub>factor</sub> (%) <sup>b</sup>	19.6	15.4	18.7	18.0
R <sub>free</sub> (%)	24.9 <sup>c</sup>	20.1 <sup>d</sup>	23.9 <sup>d</sup>	23.0 <sup>d</sup>
rms deviations				
bond lengths (Å)	0.014	0.012	0.015	0.015
bond angles(deg.)	1.6	1.5	1.6	1.5
<i>C. Temperature factors (Å<sup>2</sup>)</i>				
Main chain atoms (average)	31.0	18.7	24.0	18.2
X-Man (average)	42.2	21.0	37.7	24.9
X-Man Binding Loop atoms	34.0	20.9	26.0	19.0
* Data in parentheses are for the highest resolution shell.				
<sup>a</sup> R <sub>merge</sub> = $ I - \langle I \rangle  / \langle I \rangle$ , where $I$ is the intensity and $\langle I \rangle$ is the average $I$ for all observations of equivalent reflections.				
<sup>b</sup> R-factor = $ F_{\text{obs}} - F_{\text{calc}}  /  F_{\text{obs}} $ .				
<sup>c</sup> Calculated with 8% or <sup>d</sup> 5% of the reflections omitted from refinement				

THIS IS NOT THE FINAL VERSION - see doi:10.1042/BJ20070942

Stage 2(a) POST-PRINT

Table 3

Protein-carbohydrate interactions in the mannose-binding site of the crystal structures of wild-type and mutated *Dioclea guianensis* and *Dioclea grandiflora* lectin/X-Man complexes. For comparison, the interactions between the seed lectin from *Dioclea grandiflora* and the mannose residue occupying the monosaccharide-binding site in the crystal structure of the lectin-trimannoside complex (1DGL) are listed.

Sugar atom	Amino acid residue atom	Interacting distance (Å)				
		r- $\alpha$ Dguia [2JDZ]	r- $\alpha$ Dguia S131/H [2JE7]	r- $\alpha$ DGL [2JE9]	r- $\alpha$ DGL E131N/K132Q/E123A [2JEC]	1DGL
O3	R <sup>228</sup> N	2.9	2.9	2.9	2.9	2.9
O4	N <sup>14</sup> ND2	2.8	2.9	2.9	2.8	2.8
	D <sup>208</sup> OD2	2.5	2.5	2.6	2.6	2.7
O5	L <sup>99</sup> N	3.4	3.1	3.3	3.2	2.9
O6	Y <sup>100</sup> O	3.1	3.3	3.5	3.4	3.0
	L <sup>99</sup> N	3.3	3.1	3.2	3.3	2.9
	Y <sup>100</sup> N	3.1	3.0	3.1	3.1	3.0
	D <sup>208</sup> OD1	2.7	2.7	2.0	2.9	3.3

Figure 1A

```

M G I S K K S Q L V P L L A F I T M F L 20
ATG GGT ATT TCA AAA AAA TCT CAG TTG GTT CCT TTA CTT GCG TTC ATA ACC ATG TTC CTG 60

I V V S R V S S S I A D A N S L H F S F 40
ATA GTA GTG AGC AGG GTG AGC TCA TCA ATA GCA GAT GCA AAT TCA CTC CAT TTC AGC TTC 120

-----|----- γ-chain -----|
S Q F S Q N P K D L I L Q G D A T T D S 60
AGT CAG TTT AGC CAA AAC CCA AAG GAT TTG ATC CTT CAA GGT GAC GCC ACC ACC GAT TCA 180
-----|-----
D G N L Q L T R V S S D G S P Q G S S V 80
GAT GGT AAT TTA CAA CTC ACT AGG GTT TCC AGT GAT GGT AGT CCA CAG GGA AGC AGT GTG 240
-----|-----
G R A L F Y A P V H I W E K S A V V A S 100
GGC CGG GCT TTG TTC TAC GCC CCA GTC CAC ATT TGG GAA AAA TCT GCT GTG GTC GCA AGC 300
F: CCN GTS CAC ATT TGG GA-3'

-----|-----
F D A T F T F L I K S P D R D P A D G I 120
TTC GAT GCT ACA TTT ACA TTT CTG ATT AAA TCA CCT GAC CGT GAT CCA GCT GAT GGA ATT 360
-----|-----
T F F I A N T D T S I P S G S G G R L L 140
ACC TTC TTC ATT GCA AAT ACT GAC ACT TCA ATC CCT AGT GGT TCT GGT GGA AGA CTC CTT 420
-----|-----
G L F P D A N I I K N S T N L D F N A A 160
GGA CTC TTT CCT GAT GCA AAT ATC ATT AAA AAT TCA ACT AAT CTT GAT TTC AAT GCT GCT 480
-----|----- γ-chain -----|
Y N A D T I V A V E L D S Y P N T D I G 180
TAT AAT GCC GAT ACT ATC GTT GCT GTT GAA TTG GAC TCT TAC CCT AAT ACT GAT ATT GGA 540
-----|----- β-chain -----|
D P S Y P H I G I D I K S I R S K S T A 200
GAT CCA AGT TAT CCA CAC ATT GGA ATC GAT ATA AAG TCT ATT CGT TCC AAG TCA ACT GCC 600
-----|-----
R W N M Q T G K V G T A H I S Y N S V A 220
AGA TGG AAC ATG CAA ACT GGG AAG GTA GGA ACT GCA CAC ATA AGC TAC AAC TCT GTC GCT 660
-----|-----
K R L S A V V S Y S G T S S T T V S Y D 240
AAG AGA CTA AGT GCT GTT GTT TCT TAT TCT GGT ACT AGT TCT ACC ACT GTA TCT TAC GAT 720
-----|-----
V D L N N V L P E W V R V G L S A T T G 260
GTT GAC CTC AAC AAT GTC CTT CCT GAA TGG GTT AGG GTT GGT CTT TCT GCT ACA ACC GGT 780
R: GGN CTY ACC CAH TCY CAW-5'

-----|-----
L Y K E T N T I L S W S F T S K L K T N 280
CTT TAC AAA GAA ACC AAT ACC ATT CTC TCT TGG TCT TTT ACT TCT AAG TTA AAG ACC AAC 840
-----|----- β-chain -----|
Q L Q D L R I A S V V * 292
CAG CTC CAA GAC TTG AGA ATT GCA TCC GTC GTC TGA 876

```

Figure 1B

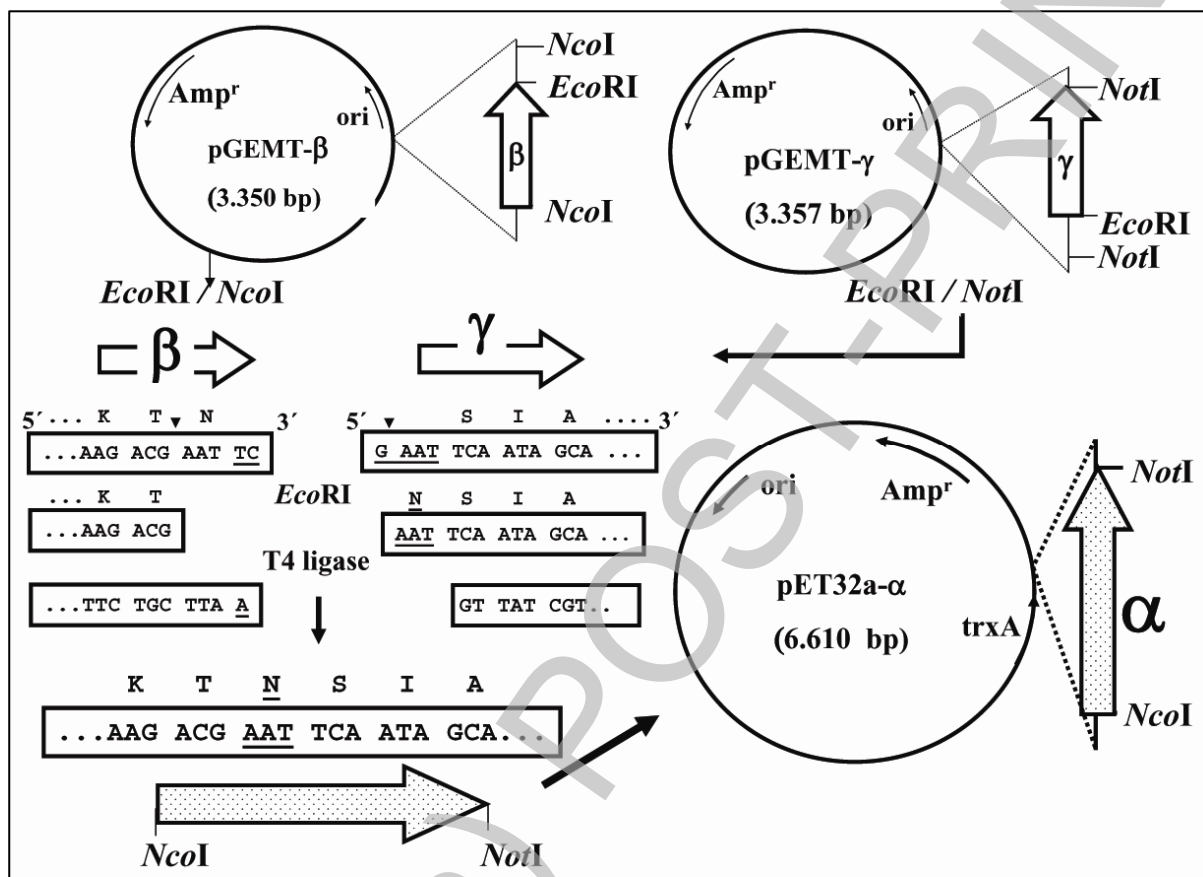
```

L I V V S R V S S S I A D E N S L H F S      20
CTG ATA GTA GTG AGC AGG GTG AGC TCA TCA ATA GCA GAT GAA AAT TCA CTC CAT TTC AGC
CTG ATA GTA GTG AGC AGG GTG AGC TCA |----- γ-chain -----
F H K F S Q N P K D L I L Q G D A F T D      40
TTC CAC AAA TTT AGC CAA AAC CCA AAG GAT TTG ATC CTT CAA GGT GAC GCC TTC ACC GAT
-----
S D G N L Q L T K V S S S G D P Q G N S      60
TCA GAT GGT AAT TTA CAA CTC ACT AAG GTT TCC AGT AGT GGT GAT CCA CAG GGA AAT AGT
-----
V G R A L F Y A P V H I W E K S A V V A      80
GTG GGC CGG GCT TTG TTC TAC GCC CCA GTT CAC ATT TGG GAA AAA TCT GCT GTG GTC GCA
-----
S F D A T F T F L I K S P D R E P A D G      100
AGC TTC GAT GCT ACC TTT ACA TTT CTG ATT AAA TCA CCT GAC CGT GAA CCA GCT GAT GGA
-----
I T F F I A N T D T S I P S G S G G R L      120
ATT ACC TTC TTC ATT GCA AAT ACT GAC ACT TCC ATC CCT AGT GGT TCT GGT GGA AGG CTC
-----
L G L F P D A N I V K N S T T L D F N A      140
CTT GGA CTC TTT CCT GAT GCA AAT ATC GTA AAA AAT TCC ACT ACC CTT GAT TTC AAT GCT
----- γ-chain -----|
A Y N A D T I V A V E L D S Y P N T D I      160
GCT TAT AAT GCC GAC ACT ATC GTT GCT GTT GAA TTG GAC TCT TAC CCT AAT ACT GAT ATT
-----
G D P N Y P H I G I D I K S I R S K S T      180
GGA GAT CCA AAT TAT CCA CAC ATT GGA ATT GAT ATC AAG TCT ATT CGT TCC AAG TCA ACC
-----
A R W N M Q T G K V G T V H I S Y N S V      200
GCC AGG TGG AAC ATG CAA ACA GGG AAG GTA GGA ACT GTT CAC ATA AGC TAC AAC TCT GTC
-----
A K R L S A V V S Y S G S S S T T V S Y      220
GCT AAG AGA CTA AGT GCT GTT GTT TCT TAT TCT GGA AGT AGT TCT ACA ACT GTA TCT TAC
-----
D V D L N N V L P E W V R V G L S A T T      240
GAC GTT GAC CTC AAC AAT GTC CTT CCT GAA TGG GTT AGA GTT GGT CTT TCT GCT ACA ACC
-----
G L Y K E T N T I L S W S F T S K L K T      260
GGT CTT TAC AAA GAA ACC AAT ACC ATT CTC TCT TGG TCT TTT ACT TCT AAG TTA AAG ACG
R: 3'-GA AGA TTC AAT TTC TGC
----- β-chain -----
N                                             261
AAT
TTA-5'
---|                                         783

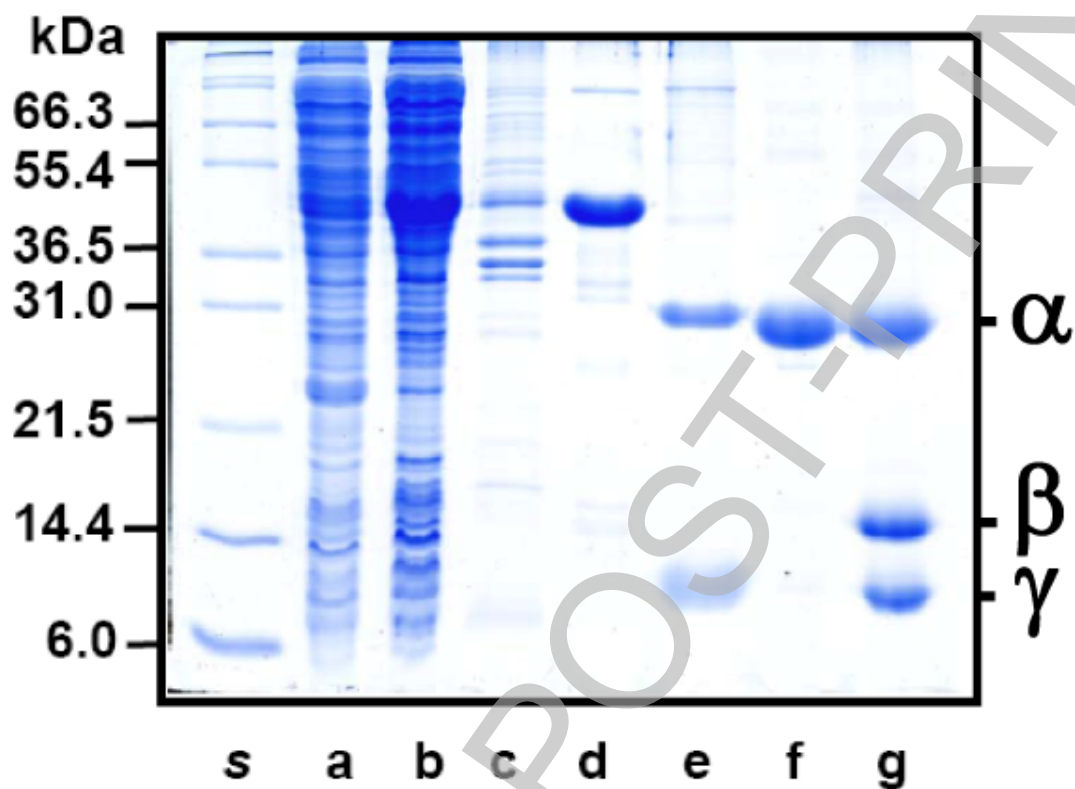
```



Figure 2



**Figure 3**



THIS IS NOT THE FINAL VERSION - see doi:10.1042/BJ20070942

Figure 4

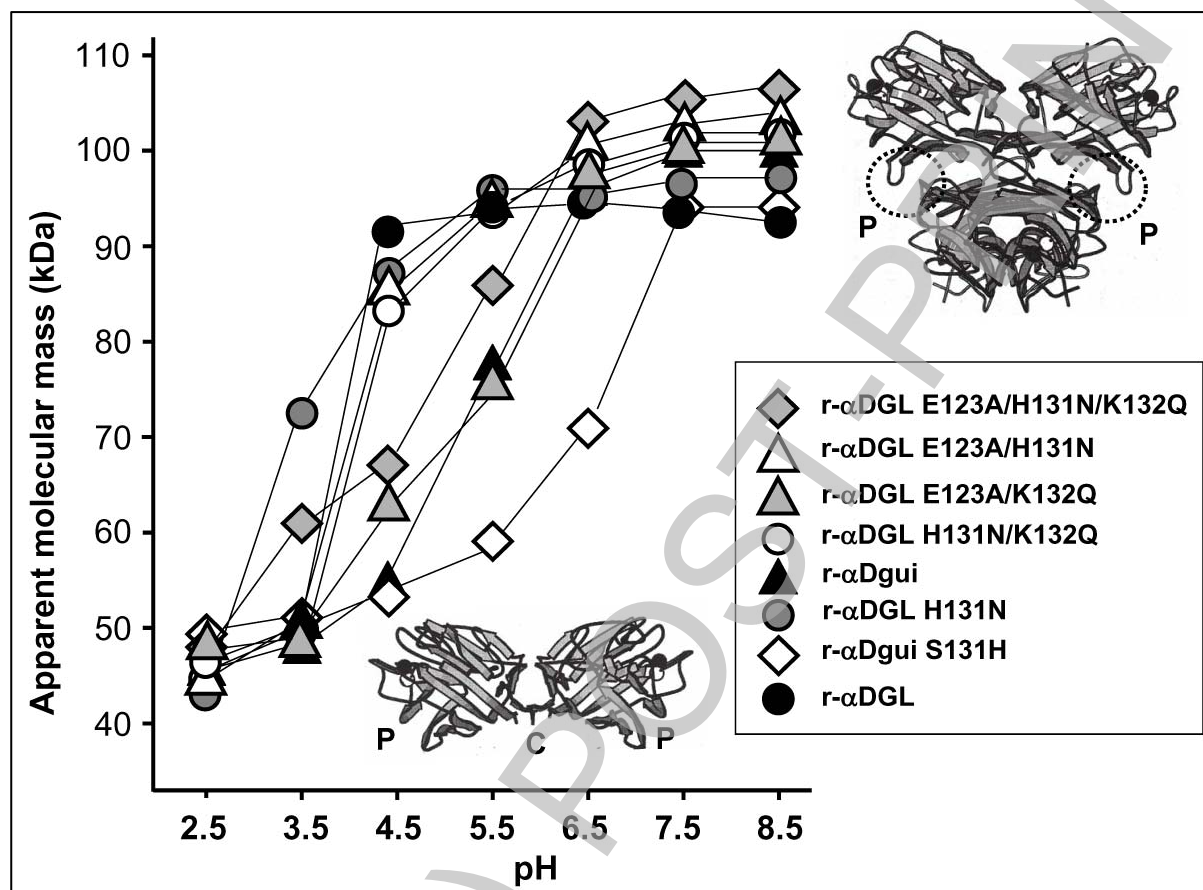


Figure 5

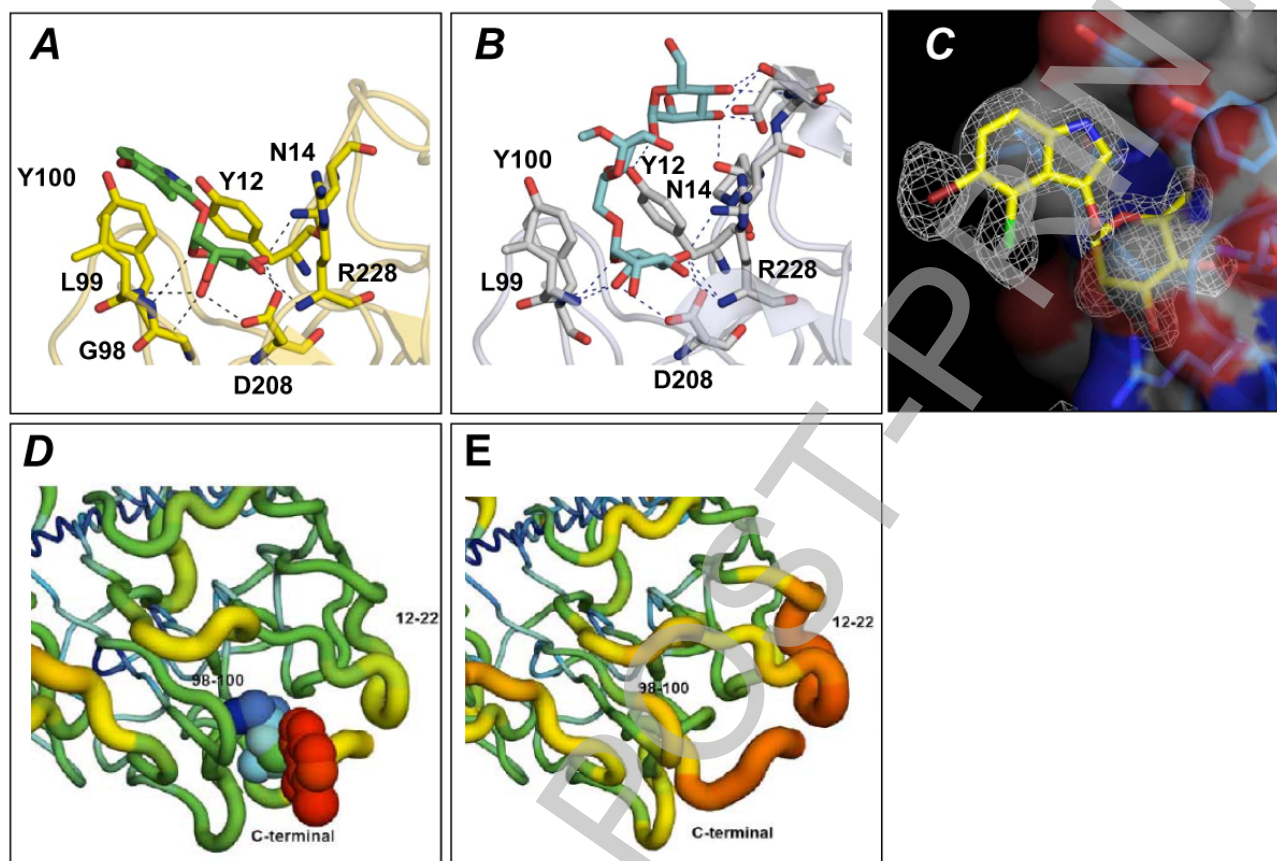


Figure 6

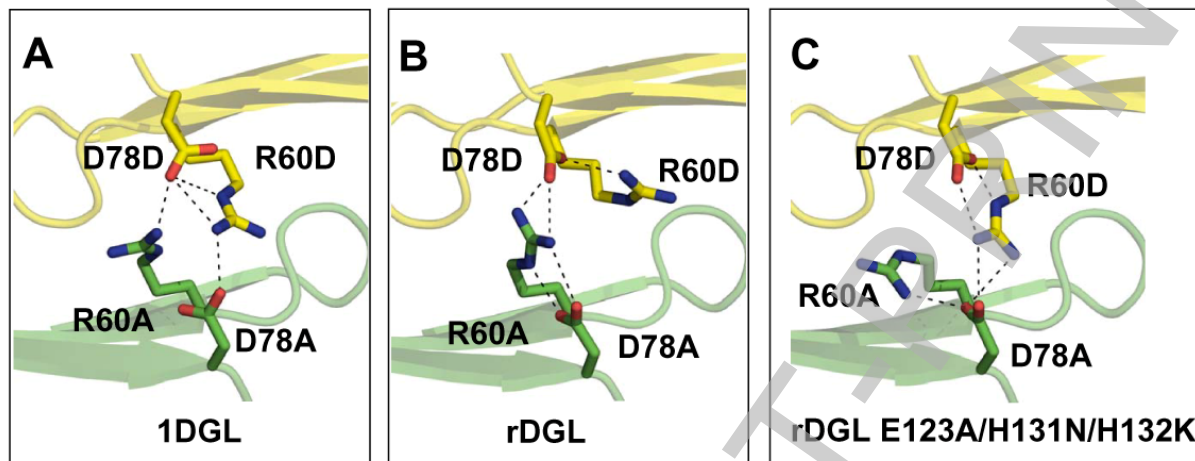




Figure 7

