Biochem. J. (1992) 284, 95–101 (Printed in Great Britain)

Primary structure of a Thomsen–Friedenreich-antigen-specific lectin, jacalin [*Artocarpus integrifolia* (jack fruit) agglutinin]

Evidence for the presence of an internal repeat

Sanjeev K. MAHANTA, S. SANKER, N. V. S. A. V. PRASAD RAO, M. J. SWAMY and Avadhesha SUROLIA*
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Jacalin [*Artocarpus integrifolia* (jack fruit) agglutinin] is made up of two types of chains, heavy and light, with M_r values of 16200 ± 1200 and 2090 ± 300 respectively (on the basis of gel-permeation chromatography under denaturing conditions). Its complete amino acid sequence was determined by manual degradation using a 4-dimethylaminoazobenzene 4′-isothiocyanate double-coupling method. Peptide fragments for sequence analysis were obtained by chemical cleavages of the heavy chain with CNBr, hydroxylamine hydrochloride and iodosobenzoic acid and enzymic cleavage with *Staphylococcus aureus* proteinase. The peptides were purified by a combination of gel-permeation and reverse-phase chromatography. The light chains, being only 20 residues long, could be sequenced without fragmentation. Amino acid analyses and carboxypeptidase-Y-digestion *C*-terminal analyses of the subunits provided supportive evidence for their sequence. Computer-assisted alignment of the jacalin heavy-chain sequence failed to show sequence similarity to that of any lectin for which the complete sequence is known. Analyses of the sequence showed the presence of an internal repeat spanning residues 7–64 and 76–130. The internal repeat was found to be statistically significant.

INTRODUCTION

The lectin from the seeds of *Artocarpus integrifolia* (jack fruit) agglutinin, jacalin, exhibits interesting biological activities and, as a consequence, has attracted considerable attention during the last few years [1–3]. The lectin has, for example, been shown to be a potent and selective stimulant of distinct T- and B-cell functions and shows a unique ability to specifically recognize IgA_1 from human serum [4,5]. This property of specific recognition of IgA_1 was found to have significant clinical promise, as has been shown in a fairly recent paper in which jacalin binding to IgA_1 was demonstrated to enhance phagocytosis of normally resistant type II group B streptococci [6]. The lectin has also been shown to react specifically with CD4-bearing T-lymphocytes unlike phytohaemagglutinin (PHA), another well-known T-cell-stimulating lectin that induces proliferation of all T-cell subsets [5,7]. Thermodynamic studies of binding with a variety of naturally occurring and synthetic saccharides, together with their conformations, highlighted the high degree of specificity of the lectin for the immunodominant portion of Thomsen–Friedenreich antigen, a tumour-associated antigen with a proven link with malignancy in man [8]. The lectin, unlike peanut (*Arachis hypogaea*) agglutinin, did not react with sugars such as lactose, *N*-acetyl-lactosamine etc. that are conformationally akin to the T-antigenic disaccharide. Subsequently we showed that the lectin reacts selectively with Thomsen–Friedenreich antigen, Galβ1,3GalNAcαSer [9]. The lectin reacted with Galβ1,3GalNAcαMe and, as expected, failed to bind a Galβ1,3GalNAcβMe and asialo-G_{M1} oligosaccharide. Interestingly enough, the specificity of the lectin in discriminating between Galβ1,3GalNacαR and Galβ1,3GalNAcβR, where R is an aglycone or a sugar chain, was far superior to that of anti-T monoclonal antibodies [10,11]. Recently another lectin, from *Amaranthus caudatus*, amaranthin, has been described that has similar ligand-binding properties [12,13]. By using specifically enriched sugars the mechanism of sugar binding by the lectin has been studied in detail by ^{13}C- and ^{19}F-n.m.r. spectroscopy [14]. Ligand-induced changes in the molar ellipticities of the lectin in its tertiary-structural region have allowed a definition of the spatial arrangement of sugars in the lectin-combining site [9]. The lectin has been crystallized with [15,16] and without sugar [16,17], but a solution of its structure by X-ray crystallography has yet to appear. To complement and extend these studies, and to obtain information on its evolutionary relationships, we have determined the complete primary structure of the protein. Primary structures of a large number of lectins are now available and reveal the presence of patterns of conserved features in regions associated with carbohydrate binding [18,19]. This strongly indicates that these proteins have been conserved throughout evolution and hence must have important functions. Jacalin did not show sequence similarity to any lectin for which the complete amino acid sequence is available. In an earlier study, however, based on the analyses of *N*-terminal amino acid sequence, Young and co-workers demonstrated a strong sequence similarity between jacalin and *Maclura pomifera* (osage orange) agglutinin, another lectin of the Moracae family with similar carbohydrate-binding properties [20]. Therefore, in conjunction with the above, the complete amino acid sequence of jacalin reported here makes it the first member with known primary structure of what is likely to emerge as yet another evolutionarily conserved family of carbohydrate-binding proteins. The lectin shows a region of internal sequence similarity spanning residues 7–64 and 76–130 which may have important bearing on the structure of the protein vis-à-vis its carbohydrate-binding properties.

Abbreviations used: MeUmb-α-Gal, 4-methylumbelliferyl α-D-galactopyranoside; MeUmb-α-GalNAc, 4-methylumbelliferyl-α-*N*-acetyl-D-galactosamine; PHA, phytohaemagglutinin; DABITC, 4-dimethylaminoazobenzene 4′-isothiocyanate; PITC, phenyl isothiocyanate; IBA, iodosobenzoic acid; SAP, *Staphylococcus* V8 endoproteinase; TFA, trifluoroacetic acid; OPA, *o*-phthalaldehyde; r.p., reverse-phase.

* To whom correspondence should be sent.

MATERIALS AND METHODS

Materials

4-Dimethylaminoazobenzene 4′-isothiocyanate (DABITC), phenyl isothiocyanate (PITC), iodosobenzoic acid (IBA), chloramine-T, carboxypeptidase Y, thiolactic acid, iodoacetic acid and guanidinium chloride (Gdm·Cl) were obtained from Sigma. DABITC was purified by recrystallization from acetone as reported [21]. Trifluoroacetic acid (TFA), *Staphylococcus* V8 endoproteinase (SAP) and *o*-phthalaldehyde (OPA) were from Pierce. Polybuffers 74 and 96 and Pharmalyte (6.5–9.0) were purchased from Pharmacia. Hydroxylamine hydrochloride was obtained from Merck. H.p.l.c.-grade acetonitrile and heptane were obtained from Spectrochem Co. (Bombay, India). CNBr was purchased from s.d. Fine Chemicals Ltd. (Boisar, India). Butyl acetate obtained from Loba-Chemie Indoaustranal Company (Bombay, India) was distilled before use. All other reagents were of the highest grade available from commercial sources.

Methods

Purification of Jacalin and its subunits. Jacalin was purified by affinity chromatography on cross-linked guar gum [8]. To purify its subunits, jacalin was dissolved in 6 M-Gdm·Cl, followed by incubation at 60 °C for 2 h. TFA was then added to a final concentration of 25% and the sample loaded on to a reverse-phase (r.p.) h.p.l.c. column (C4; Vydac; 4.6 mm × 25 cm) and eluted with a gradient of acetonitrile in 0.1% TFA. The major peaks thus obtained were numbered I, II, III and IV in the order of their elution from the column. In another method the lectin was heat-denatured and the precipitated protein dissolved in 80% (v/v) formic acid. It was then loaded on to a Sephadex G-75 column (1.15 cm × 115 cm) equilibrated and eluted with 40% formic acid. The column was calibrated with reduced and carboxymethylated insulin A-chain (M_r 2380), insulin B-chain (3420), cytochrome *c* (12300), lysozyme (14300) and apo-myoglobin (17 000). The peaks obtained upon gel filtration, JacI and JacII, were repurified on a C4 r.p.h.p.l.c. column as described above. JacI obtained from gel filtration was subjected to two successive reverse-phase purifications, and each time only the leading edge was collected. This was termed 'PIVa' and used for sequence determination of the heavy chain.

Amino acid analysis. Amino acid analyses of PIVa, PIVb and their derived peptides, as well as peptides I, II and III, were carried out using OPA as described by Cooper *et al.* [22]. The OPA-derivatized hydrolysate were analysed by r.p.h.p.l.c. using a Waters 420 AC fluorescence detector connected to a Waters 745 data module.

Sequence determination. Peptides were sequenced using the DABITC/PITC double-coupling method of Chang [23]. The thiohydantoin amino acids were detected by h.p.l.c. using a Waters NOVA PAK C_{18} (3.9 mm × 150 mm) column with a gradient elution and solvent system similar to that described by Chang [24].

Enzymic cleavage. H.p.l.c.-purified PIVa was digested with SAP under conditions specific for cleavage at glutamic acid residues [25] in the presence of 0.1% SDS and 120 mM-ammonium bicarbonate, pH 8.0. Enzyme was added at an enzyme/protein ratio of 1:30 (w/w), and digestion was allowed to proceed at 37 °C for 12 h. A second aliquot of the enzyme was then added to a final enzyme/protein ratio of 1:15, and the reaction was allowed to continue for another 12 h. The digest was then dissolved in 40% formic acid and purified with a reverse-phase column (PepRPC; 15 μm; 10/10; Pharmacia) and a linear gradient of acetonitrile containing 0.08% TFA.

CNBr cleavage. Dried Peak IVa was dissolved in 70% formic acid and an 80-fold molar excess (relative to methionine) of CNBr was added. The reaction was allowed to proceed in the dark for 24 h. The reaction mixture, after dilution and freeze-drying, was redissolved in 40% formic acid and fractionated on a Sephadex G-75 column (1.15 cm × 115 cm) equilibrated with 40% formic acid. The peaks obtained were numbered CNBr1, CNBr2 and CNBr3 as they were eluted from the column. All the peaks were subjected to sequence and amino acid analysis.

Blocking the *N*-terminus of PIVa was done by the method of Bhown *et al.* [26]. The protein was suspended in 0.4 M-potassium borate buffer, pH 9.6, and OPA reagent was added. After 1 min of incubation at room temperature the protein was precipitated and washed with acetone before cleavage with CNBr. Gel filtration was performed to separate the fragments under conditions identical with those used for the separation of CNBr1, CNBr2 and CNBr3. The peak eluted at the position of CNBr2 was subjected to sequence analysis.

Hydroxylamine hydrochloride cleavage. Cleavage at Asn–Gly bonds was carried out as described in [27]. A solution of 6 M-Gdm·Cl and 2 M-hydroxylamine was prepared at ambient temperature and 4 M-LiOH was added slowly until a pH of 9.0 was reached. To 2 ml of this solution, 10 mg of PIVa was added and the reaction allowed to continue for 4 h at 45 °C. The reaction was stopped by lowering the pH to 2.0 using formic acid. The product was fractionated on a Sephadex G-25 column (1.4 cm × 150 cm) in the presence of 10% formic acid.

IBA cleavage. IBA cleavage of PIVa was performed by the method of Fontana *et al.* [28]. A 10 mg/ml solution of IBA made in 4 M-Gdm·Cl/80% acetic acid was left in the dark for 2 h before the addition of 10 mg of PIVa/ml of the above solution. The mixture was incubated at room temperature for 20 h. After diluting and drying the mixture, it was redissolved in 40% formic acid and purified on an r.p.f.p.l.c. column (Pep RPC; 15 μm; 10/10; Pharmacia) using a linear gradient of acetonitrile containing 0.1% TFA.

Carboxypeptidase Y digestion. Carboxypeptidase Y digestion was performed as described in [29], except for the inclusion of Gdm·Cl. Polypeptides (0.5–1 nmol) were dissolved in 0.05 M-sodium acetate buffer, pH 5.5, and 2 M-Gdm·Cl and enzyme were added at an enzyme-to-peptide ratio of 1:250. Portions were removed after various times for amino acid analyses.

Computer analysis. The internal-repeat regions in the sequence were found by using the programs PLFASTA [30] and LFASTA [31]. The regions shown to be similar by the above methods were subjected to IALIGN algorithm [32] for quantification of the significance of similarity.

Chromatofocusing and isoelectric focusing

Jacalin in water was loaded on an f.p.l.c. column (Mono P HR 5/20; Pharmacia) equilibriated with 25 mM-Tris/acetate, pH 8.3. Elution was carried out with a mixture of Polybuffer 96 and Polybuffer 74 (3:7, v/v), diluted 15-fold and the pH adjusted to 4.9 with acetic acid. Isoelectric focusing was performed on Agarose gels using Pharmalyte (6.5–9.0) according to the supplier's instructions.

RESULTS

Overall procedure for obtaining the primary structure

Jacalin polypeptides, upon r.p. purification, were eluted as a bunch of three early-eluted peaks (I, II and III) and a late-eluted peak (IV) as shown in Fig. 1. Similar results have been reported in a previous study by Young *et al.* [20], where the sequence from *N*-terminus was shown for each, the early- as well as the late-eluted peaks. However, the reported r.p.h.p.l.c. purification procedure proved unsatisfactory in achieving complete dissocia-

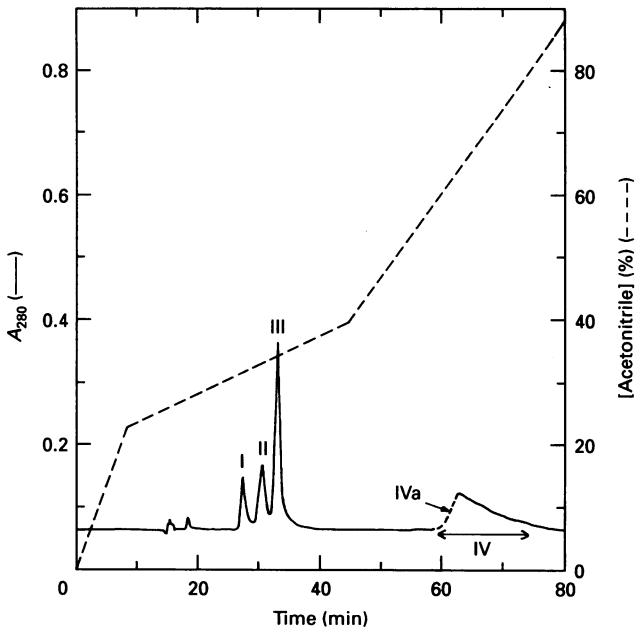

Fig. 1. R.p. separation of jacalin subunits

Jacalin was denatured with 6 M-Gdm·Cl at 60 °C for 2 h. TFA was added to a final concentration of 25% and the mixture applied on to a C4 column (Vydac; 4.6 mm × 25 cm) and eluted with a gradient of acetonitrile in 0.1% TFA as shown. The peaks obtained were designated I, II, III and IV in the order of their elution from the column. The leading edge of peak IV collected separately is designated 'PIVa'.

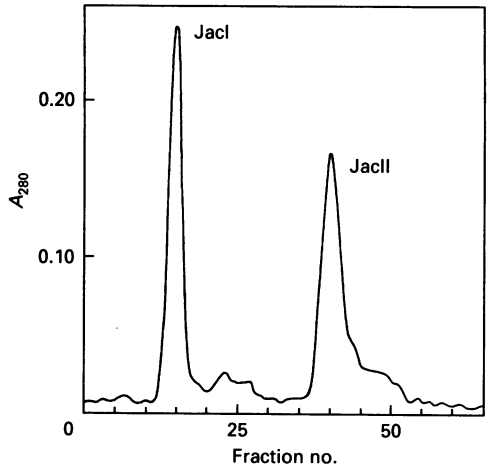

Fig. 2. Gel-permeation chromatography of jacalin

The dissociation of jacalin subunits and their separation on Sephadex G-75 was carried out as described in the Materials and methods section. A flow rate of 10 ml/h was used, and fractions (1.95 ml each) were collected. Both peaks, JacI and JacII, were concentrated and subjected to r.p. purification under conditions identical with those described in Fig. 1. JacI was eluted in the same position as PIV, whereas JacII gave a three-peak pattern similar to peaks I, II and III of Fig. 1.

tion of the early- and the late-eluted peaks. The presence of 6M-Gdm·Cl and 25% TFA in the sample enabled the complete dissociation and separation of the two chain types on rpHPLC (Fig. 1).

Gel-permeation chromatography of jacalin on Sephadex G-75 under denaturing conditions yielded two peaks, JacI and JacII, with approximate M_r values of 16200 ± 1200 and 2090 ± 300 respectively (Fig. 2). *N*-Terminal sequence analyses and r.p.h.p.l.c. patterns revealed JacI to be identical with PIV and JacII a mixture of the peaks I, II and III (see Fig. 1). This consistency of results obtained from two independent methods, gel-filtration and r.p. separation, establishes that jacalin, as indicated by the sequence data, is indeed made up of heavy and light chains. It may be noted that earlier size-exclusion-based attempts had reported failure to observe the light- and heavy-chain constituents, presumably owing to the inability, under the conditions used, to dissociate the two chain types [33]. JacI as well as PIV eluted as a broad peak on r.p.h.p.l.c., possibly owing to heterogeneity of the polypeptide chains. Hence only the leading edge of JacI purified twice successively on r.p.h.p.l.c. was collected separately. This was designated as PIVa and used for sequence determination. PIV without additional purification showed a more heterogenous sequence, as explained below.

The sequences of the three light chains obtained by us, I, II, and III, were, as shown below, consistent with the results of Young *et al.* [20], except for one discrepancy at the *C*-terminus of the light-chain subtype II.

I NEQSGKSQTVIVGSWGAKVS
II NEQSGKSQTVIVGPWGAQVST
III DEQSGISQTVIVGPWGAKVS

Cleavage with SAP resulted in six peptides, two of which contained Asn–Gly cleavage sites. Therefore cleavage with hydroxylamine hydrochloride was performed, which yielded three fragments. CNBr cleavage of PIVa also yielded three fragments. Additional fragments for peptide overlap were obtained from IBA cleavage. The sequence-derived amino acid composition of jacalin subunits and all the fragments matched those obtained from amino acid analysis (results not shown). A higher estimate of the M_r of the jacalin heavy chain obtained from gel filtration compared with that obtained from the sequence is possibly due to glycosylation, as jacalin is a glycoprotein with defined carbohydrate structure [34].

Sequence analysis of peptides obtained from an SAP digest

Peptides obtained by cleavage at the glutamic acid residues of PIVa were SAP-1, SAP-3, SAP-4, SAP-5, SAP-6 and SAP-7 (Fig. 3). SAP-1, SAP-3 and SAP-6 were completely sequenced. *C*-Terminal sequence analyses were also performed for the larger peptides, SAP-4, SAP-5 and SAP-7. The presence of glutamic acid as the last residue in all the fragments confirmed the specificity of the cleavage.

Sequence analysis of CNBr-cleavage-derived peptides

Sephadex G-75 gel-filtration chromatography of the CNBr digest of PIVa resulted in three fragments, two of which, of similar size, were co-eluted as one peak, CNBr2. Attempts to purify CNBr2 further by r.p.h.p.l.c. proved unsuccessful. As one sequence in the mixture appeared to come from the *N*-terminus, PIVa was blocked with OPA at the *N*-terminus and cleavage was performed again. Sequence analysis of the peak (CNBr2b) obtained at the same position as that of CNBr2 now showed a single amino acid residue at each cycle. An unambiguous sequence up to 23 residues was obtained for this peptide (Fig. 3). The other peptide which in the mixture showed the sequence Gly-Lys-Ala-Phe-Asp-Asp... was therefore concluded to be the *N*-terminal fragment and was designated 'CNBr2a'. CNBr3 upon sequencing was found to be a tetrapeptide (Fig. 3).

Sequence analysis of the hydroxylamine-derived peptides

Gel-filtration chromatography on Sephadex G-25 of hydroxylamine-cleaved PIVa gave three peaks, HA3, HA4 and HA5, corresponding to unique peptides. HA3 was sequenced through

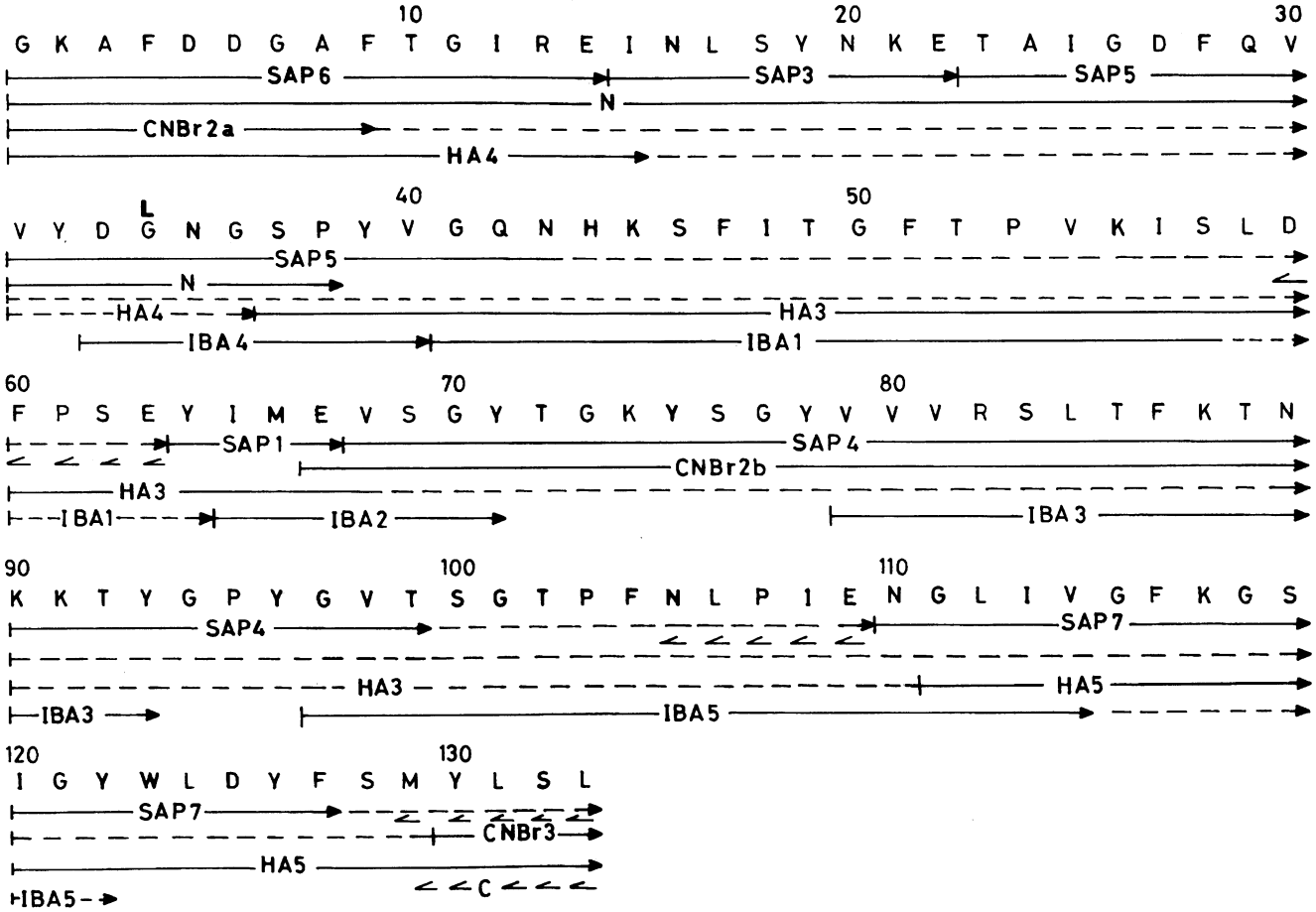

Fig. 3. Complete sequence of jacalin heavy chain (PIVa)

The sequence of the peptide is represented by the single-letter code. N denotes the sequence determined by *N*-terminal sequence analysis. SAP, CNBr, HA and IBA denote peptides derived from SAP, CNBr, hydroxylamine hydrochloride and IBA digestion of PIVa respectively. Continuous lines indicate the amino acid sequence obtained by manual sequencing, and the broken lines indicate the sequence which was not determined by sequencing but was compatible with the amino acid analysis of the peptides. Lines with reverse arrows indicate the sequence obtained from the *C*-terminus by digestion with carboxypeptidase Y, for the peptides under which they occur. The result of *C*-terminal analyses of uncleaved PIVa is shown by the reverse arrows at the extreme end. Leu, observed at position 34 along with Gly, is also shown.

20 cycles and proline observed at the position 18 was used to block all the primary amino groups after the cycle 17 and the sequence carried through an additional 15 cycles. The HA3 sequence could thus be deduced up to 33 residues. HA4 showed a sequence corresponding to the *N*-terminus. HA5 was sequenced completely. The sequences derived from HA3, HA4 and HA5 are shown in Fig. 3.

Sequence analysis of the IBA-derived peptides

Reverse-phase separation of IBA-digested PIVa yielded several peptides, out of which five, namely IBA1, IBA2, IBA3, IBA4 and IBA5, could be sequenced. Whereas peptides IBA2, IBA3 and IBA4 were sequenced completely, IBA1 and IBA5 were each sequenced through 18 cycles (Fig. 3).

Determination of the *C*-terminal residue

Alignment of all the relevant peptides showed the tetrapeptide CNBr3 to be farthest from the *N*-terminus. Leucine, being the last residue of CNBr3, was established as the *C*-terminal residue of PIVa.

Carboxypeptidase Y digestion of PIVa showed leucine to be the *C*-terminal residue, followed by serine (Fig. 4). Although tyrosine apparently follows serine, the biphasic release pattern of leucine indicates the presence of an additional leucine residue. Consideration of this leucine residue as the third one makes tyrosine and methionine the fourth and fifth residues respectively. This is consistent with the *C*-terminal sequence Met-Tyr-Leu-Ser-Leu for PIVa obtained from sequence analyses of peptide fragments HA5 and CNBr3.

Carboxypeptidase digestion of peaks I, II and III confirmed the sequences obtained from Edman degradation, except for peak II, which had an additional threonine residue as compared with the sequence of II reported by Young and co-workers [20].

Heterogeneity of the jacalin heavy-chain sequence

PIV without additional purification showed heterogeneity at several places in the sequence (Table 1). Though not fully characterized, seven positions could be identified where more than one residue were seen in the sequence. These positions and the amino acids observed at each position are as follows. Gly and Leu at position 34, Lys and Thr at position 45, Met, Asp and Val at position 66, Ile and Thr at position 72, Asn and Lys at position 74, Thr and Asp at position 102 and Leu and Asn at position 131.

The sequence of PIVa did not exhibit such heterogeneity. However, at position 34, two residues, Gly and Leu, were both observed in substantial amounts. The yield of Gly being comparatively higher, it was assigned to this position. Leucine probably derives from a heavy-chain isoform present in significant amounts in PIVa (approx. 30% of the total).

DISCUSSION

Alignment of peptides and the complete sequence of jacalin heavy chain

A minimum of three residues overlap was taken for any alignment of peptides. Over 70% of the sequence was deduced from the peptides obtained from SAP cleavage. Additional peptides obtained by cleavage with the chemicals CNBr, IBA and hydroxylamine hydrochloride were used for complete alignment, as shown in Fig. 3. Additional pieces of evidence suggesting that Fig. 3 represents the complete sequence of jacalin heavy chain are as follows: the sequence-derived amino acid composition of the peptide chain is in agreement with its amino acid analysis for a polypeptide chain of M_r 14668, the excess M_r obtained by gel filtration (M_r 16200±1200) being probably accounted for by the presence of carbohydrates. No peptides were found from any digest that were not accounted for in the sequence (Fig. 3). Carboxypeptidase-Y-digestion *C*-terminal sequence analyses of PIVa and SAP7 provided further confirmatory evidence for the completeness of the sequence.

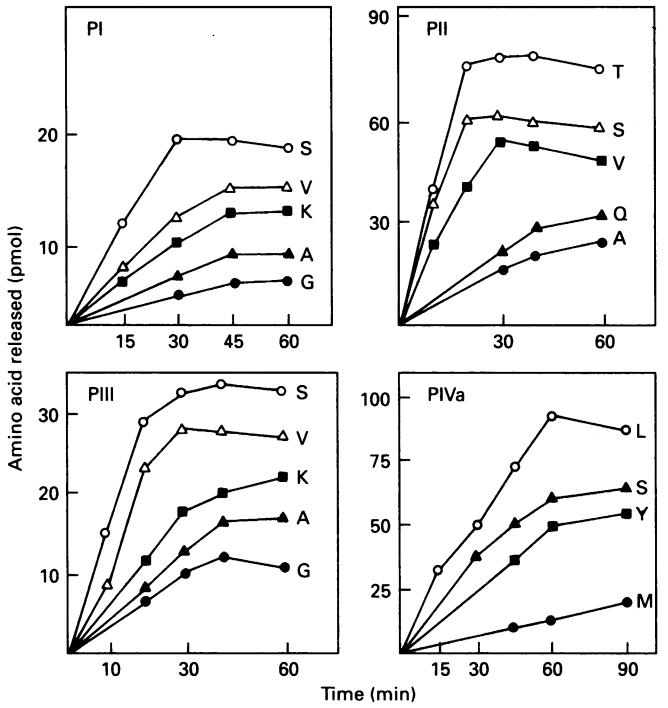

Fig. 4. Release of amino acids upon digestion of polypeptides with carboxypeptidase Y

Digestion with the enzyme and the amino acid analyses was carried out as described in the Materials and methods section. (*a*), (*b*), (*c*) and (*d*) represent the amino-acid-release patterns observed for the polypeptides I, II, III and PIVa respectively.

As mentioned before, unfractionated jacalin heavy-chain (PIV) sequence shows heterogeneity at many positions. Several sequences of the heavy chain are therefore theoretically possible, corresponding to various combinations of amino acids at these positions, which may or may not all be realized. However, taken together with the three major light chains, permutations and combinations of these light and heavy chains would conceivably give rise to a large number of isoforms of the oligomeric protein, as is evident from the results of chromatofocusing of jacalin (Fig. 5), which shows the presence of at least 29 discernible fractions. None of these fractions, even after repeated chromatofocusing, yielded any single isoform that was homogeneous by isoelectric focusing and in quantity enough for sequence determination. Thus the complexity observed with jacalin is higher than that usually observed in lectins. The five isolectins of PHA for example, are separated with relative ease by ion-exchange chromatography [35]. Also, in contrast with the marked differences in the biological activity observed with the isolectins of PHA [35], which arises as a result of varying proportions of the lymphocyte- and the erythrocyte-reacting subunits with distinct carbohydrate specificities, the chromatofocused fractions of jacalin did not show any difference in carbohydrate specifity (S. K. Mahanta, M. J. Swamy & A. Surolia, unpublished work).

Persistence of heterogeneity to an extent in PIVa points towards the presence of more than one heavy-chain isoform in it. Nevertheless, in view of the enrichment achieved by repeated fractionations to obtain PIVa, we believe that the sequence

Table 1. Amino acid composition of r.p.h.p.l.c.-purified jacalin subunits

Values in parentheses indicate residues from sequence analysis. Abbreviation: ND, not detected.

Amino acid	Subunit ...	I	II	III	PIVa	PIV
P		– (–)	0.76 (1)	0.79 (1)	5.4 (6)	5.64 (6)
D		1.17 (–)	0.89 (–)	1.32 (1)	12.76 (6)	15.31 (6)
E		2.73 (1)	3.32 (–)	2.87 (1)	7.05 (5)	7.25 (5)
N		ND (1)	ND (1)	ND (–)	ND (7)	ND (7)
Q		ND (2)	ND (3)	ND (2)	ND (2)	ND (2)
S		3.81 (4)	3.26 (3)	3.08 (3)	11.90 (12)	11.83 (12)
H		– (–)	– (–)	– (–)	1.12 (1)	0.96 (1)
G		2.89 (3)	3.12 (3)	2.65 (3)	17.61 (18)	18.41 (18)
T		1.28 (1)	1.88 (2)	0.79 (1)	10.16 (10)	10.96 (10)
R		– (–)	– (–)	– (–)	1.89 (2)	2.09 (2)
A		0.96 (1)	0.93 (1)	1.28 (1)	3.11 (3)	2.81 (3)
Y		– (–)	– (–)	– (–)	12.07 (12)	11.88 (12)
M		– (–)	– (–)	– (–)	1.79 (2)	1.93 (2)
V		2.97 (3)	2.86 (3)	3.41 (3)	9.63 (10)	10.78 (10)
W		ND (1)	ND (1)	ND (1)	ND (1)	ND (1)
F		– (–)	– (–)	– (–)	9.86 (10)	9.94 (10)
I		1.12 (1)	0.78 (1)	2.12 (2)	8.71 (9)	10.60 (9)
L		– (–)	– (–)	– (–)	9.18 (9)	11.22 (9)
K		2.14 (2)	2.02 (2)	0.87 (1)	9.03 (9)	8.66 (9)

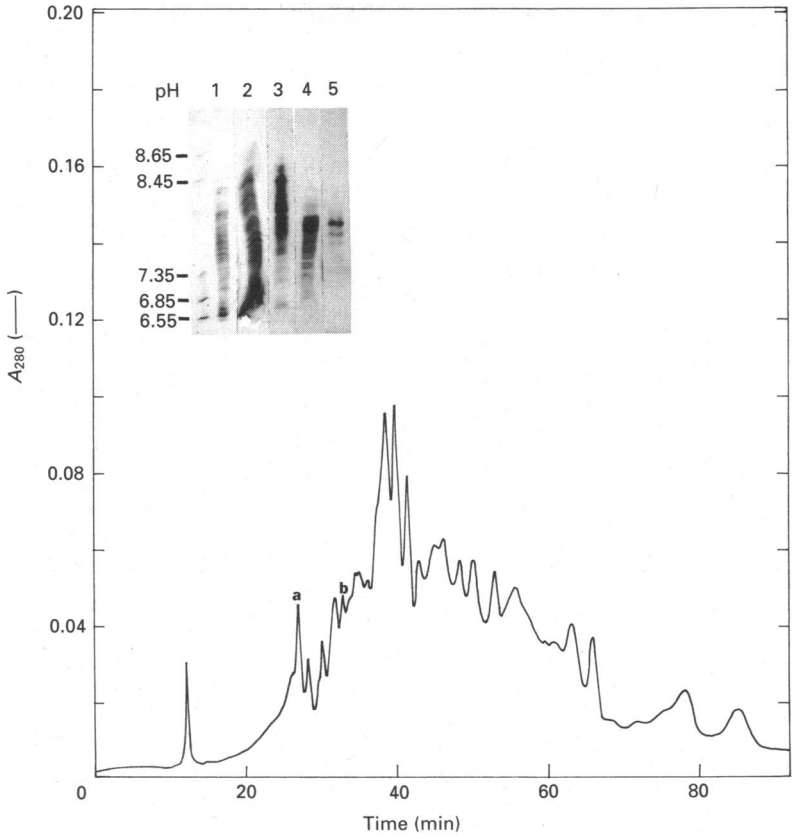

Fig. 5. Chromatofocusing and isoelectric focusing

Chromatofocusing of jacalin was performed as described in the Materials and methods section. The inset shows the isoelectric-focusing patterns: 1 and 2 represent two different concentrations of jacalin; 3 and 4 are the isoelectric-focusing patterns of the chromatofocusing-derived fractions a and b; 5 is obtained upon rechromatofocusing b twice. The markers from anode to cathode are: human carbonic anhydrase B (pI 6.55); horse myoglobin acidic (pI 6.85); horse myoglobin basic (pI 7.35); lentil middle (pI 8.45); and lentil basic (pI 8.65).

reported here represents that of a single heavy-chain isoform, specifically the one predominant in our preparation.

Sequence comparison analyses

The amino acid sequence of jacalin did not show similarity with any lectin for which the complete sequence was available. However, the results of Young and co-workers [20], based on *N*-terminal sequence analyses, clearly brings out the sequence similarity of jacalin to *Maclura pomifera* agglutinin, also a lectin from the Moracae family. Moreover, the two lectins also have overlapping carbohydrate-binding specificity. This reveals the existence of a new evolutionarily conserved family of carbohydrate-binding proteins.

Interestingly, jacalin heavy-chain sequence, when analysed with the comparison programs PLFASTA and LFASTA [30,31] showed regions of internal repeat. These alignment programs are based on the algorithm of Pearson & Lipman [30] and are suitable for identifying local sequence similarities or duplications. Two segments, namely residues 7–64 and 76–130, were found to overlap with each other (Fig. 6), with an identity of 25.4%, as shown below:

```
GAFTGIREINLSYNKETAIGDFQVVYDGNGSPYVGQNHKSFITGF-TPVKISLDFPSEY
.... .:..... ::.: :. : ..:.:. . ....:.:: ... ::. : :
SGYVVVRSLTFKTNKKT-YGPYGV---TSGTPFNLPIENGLIVGFKGSIGYWLDYFSMY
```

where : indicates complete identity and . conservative replacement.

The significance of the similarity was tested by the program IALIGN, which gave a score of 5.40 S.D. with a gap penalty of 6 and 100 random runs, implying that the probability of obtaining this score by chance alone is less than 1 in 10^6. Therefore the similarity between the repeat regions appear significant, though conclusions regarding gene duplication as being the cause would be premature in the absence of the genomic sequence and because the repeat regions are not in tandem.

The internal repeat of jacalin heavy chain needs to be examined in the context of its subunit composition. Earlier reports described jacalin as a protein of M_r 40000, contributed equally by four identical subunits [36,37] and offering two binding sites for sugar [8,37]. That this is not correct is evident from the sequence reported here and the findings of Young and co-workers [20]. Redetermination of the M_r by gel-permeation chromatography on Sephadex G-100 in the presence of galactose and Ferguson analyses of jacalin electrophoretic mobilities yielded a value of approx. 66000 [16]. Correction for the M_r in the earlier estimate of the number of binding sites results in a change of this value from 2.2 to 3.63, implying that the lectin offers four binding sites instead of only two, as believed currently. Our recent Scatchard analyses for the binding of sugars such as 4-methylumbelliferyl α-D-galactopryanoside and 4-methylumbelliferyl-α-*N*-acetyl-D-galactosamine with the lectin also provide evidence for the presence of four binding sites (A. Surolia & D. Gupta, unpublished work). Therefore the internal-repeat structure of jacalin cannot be envisaged as giving rise to the formation of two independent sugar-binding sites per heavy chain, as that would lead to a total of eight binding sites per molecule. The possibilities

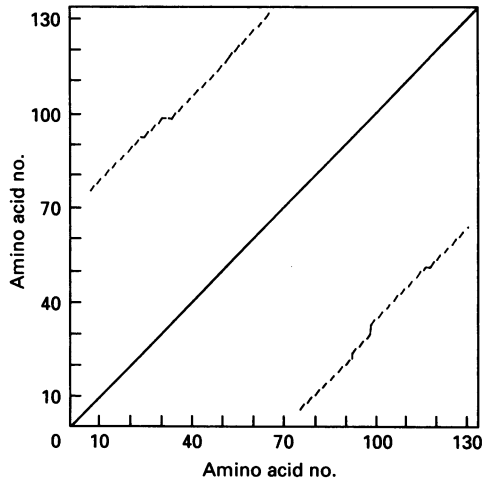

Fig. 6. Internal sequence similarities in jacalin heavy chain

Identification of the regions of internal repeats was carried out by using the program PLFASTA, which gives a graphic-matrix representation of regions showing similarity. A value of 1.0 was used as the *ktup* parameter. The two lines on either side of the diagonal and parallel with it show the regions of internal homology. The amino acid segment 7–64 is seen to overlap with the segment 76–130.

that remain are as follows: (*a*) The two identical repeat regions combine together to form one binding site, as is observed in the structures of aspartic proteinases [39] and (*b*) the two internal regions are contributed one from each heavy subunit to form the combining site at the protomer–protomer interface as observed in wheat-germ agglutinin [40].

The earlier study of Young *et al.* [20] had proposed that the early-eluted peaks on r.p.h.p.l.c. were 20 residues in length, though no additional evidence was provided. The amino acid compositions reported here (Table 1) support the sequences shown for I, II and III. Conclusive evidence for the length of these peptides was also provided by carboxypeptidase-Y-digestion *C*-terminal sequence analyses (Fig. 4).

In conclusion, Jacalin is a protein of considerable heterogeneity, which presumably arises as a result of the numerous forms of its constituent polypeptide chains. Its heavy chain shows the presence of an internal repeat region spanning over 92.0% of the sequence, which in all likelihood also encompasses the active-site residues. Therefore the internal repeat is an important structural feature and is possibly of direct relevance to its carbohydrate binding. Information on its subunit structure, characterization of its heterogeneity, location of the active site and, finally, its three-dimensional structure, would be necessary to fully understand the structure–function relationship in this protein.

This work was supported by a grant to A.S. from the Department of Science and Technology, Government of India. N.V.S.A.V.P.R., S.S. and M.J.S. are Research Associates funded by the above grant. S.K.M. is a senior Research Fellow of the University Grants Commission, Government of India. We are grateful to the Distributed Information Centre, Department of Physics, Indian Institute of Science, Bangalore 560 012, a facility funded by the Department of Biotechnology, Government of India, for their help in computation with regard to comparison of protein sequences. We thank Dr. M. R. N. Murthy and Mr. Kamaldeep Puri for useful discussions.

REFERENCES

1. Bunn-Moreno, M. M. & Campos-Neto, A. (1981) J. Immunol. **127**, 427–429
2. Hiemstra, P. S., Garter, A., Stuurman, M. E., Van, Es. L. A. & Dhar, M. R. (1987) Scand. J. Immunol. **26**, 111–117
3. Crane, I., Leung, H., Barwick, S., Parti, S. & Meager, A. (1984) Immunology **53**, 855–859
4. Kondoh, H., Kobayashi, K., Hagiwara, K. & Kaji, T. (1986) J. Immunol. Methods **88**, 171–173
5. Aucouturier, P., Pineau, N., Brugier, J. C., Mihaesco, E., Duarte, F., Skvaril, F. & Preud'homme, J. L. (1989) J. Clin. Lab. Anal. **3**, 244–251
6. Payne, N. R., Concepcion, N. F. & Anthony, B. F. (1990) Infect. Immunol. **58**, 3663–3770
7. Goldstein, I. J. & Poretz, R. D. (1986) in The Lectins: Properties, Functions and Applications in Biology and Medicine (Leiner, I. E., Sharon, N. & Goldstein, I. J., eds.), p. 173, Academic Press, Orlando
8. Krishna Sastry, M. V., Banerjee, P., Patanjali, S. R., Swamy, M. J., Swarnalatha, G. V. & Surolia, A. (1986) J. Biol. Chem. **261**, 11726–11733
9. Mahanta, S. K., Krishna Sastry, M. V. & Surolia, A. (1990) Biochem. J. **265**, 831–840
10. Clausen, H., Levery, S. B., Kannagi, R. & Hakomori, S.-i. (1986) J. Biol. Chem. **261**, 831–840
11. Clausen, H., Stroud, M., Parker, J., Springer, G. & Hakomori, S.-i. (1988) Mol. Immunol. **25**, 199–204
12. Rinderle, S. J., Goldstein, I. J., Matta, K. L. & Ratcliffe, R. M. (1989) J. Biol. Chem. **264**, 16123–16131
13. Swamy, M. J., Gupta, D., Mahanta, S. K. & Surolia, A. (1991) Carbohydr. Res. **213**, 59–67
14. Krishna Sastry, M. V., Swamy, M. J. & Surolia, A. (1988) J. Biol. Chem. **263**, 14826–14831
15. Dhanraj, V., Patanjali, S. R., Surolia, A. & Vijayan, M. (1988) J. Mol. Biol. **203**, 1135–1136
16. Banerjee, R., Dhanaraj, V., Mahanta, S. K., Surolia, A. & Vijayan, M. (1991) J. Mol. Biol. **221**, 773–776
17. Basu, D., Delucas, I., Parks, E. H. & Sudath, F. L. (1988) J. Mol. Biol. **201**, 661–662
18. Drickamer, K. (1988) J. Biol. Chem. **263**, 9557–9560
19. Sharon, N. & Lis, H. (1990) FASEB J. **4**, 3198–3214
20. Young, M. N., Johnston, R. A. Z., Szabo, A. G. & Watson, D. C. (1989) Arch. Biochem. Biophys. **270**, 596–603
21. Chang, J. Y. (1983) Methods Enzymol. **91**, 455–466
22. Cooper, J. D. H., Lewis, M. T. & Turnell, D. C. (1984) J. Chromatogr. **285**, 490–494
23. Chang, J. Y., Brauer, D. & Wittmann-Liebold, B. (1978) FEBS Lett. **93**, 205–214
24. Chang, J. Y. (1988) Anal. Biochem. **170**, 542–556
25. Drapeau, G. R. (1977) Methods Enzymol. **47**, 189–191
26. Bhown, A. S., Cornelius, T. W., Volanakis, J. E. & Bennet, J. C. (1983) Anal. Biochem. **131**, 337–340
27. Bornstein, P. & Balian, G. (1977) Methods. Enzymol. **47**, 132–145
28. Fontana, A., Dalzoppo, D., Grandi, C. & Zambonin, M. (1983) Methods. Enzymol. **91**, 311–318
29. Jones, B. N. (1986) in Methods of Protein Microcharacterization (Shively, J. E., ed.), pp. 344–348, Humana, Clifton, NJ
30. Pearson, W. R. & Lipman, D. J. (1988) Proc. Natl. Acad. Sci. **85**, 2444–2448
31. Pearson, W. R. (1990) Methods Enzymol. **183**, 88–91
32. Dayhoff, M. O. (1979) in Atlas of Protein Sequence and Structure (Dayhoff, M. O., ed.), vol. 5, suppl. 3, pp. 345–358, National Biomedical Research Foundation, Washington, DC
33. Roque-Barraeira, M. C., Praz, F., Halbwachs-Mercarelli, L., Greene, L. J. & Campos-Neto, A. (1986) Brazil J. Med. Biol. Res. **19**, 149–157
34. Capon, C., Piller, F., Wiermzeksi, J. M., Lesoy, Y. & Fomnet, B. (1990) Carbohydr. Res. **199**, 121–127
35. Leavitt, R. D., Felsted, R. L. & Bachur, N. R. (1977) J. Biol. Chem. **252**, 2961–2966
36. Sureshkumar, G., Appukuttan, P. S. & Basu, D. (1982) J. Biosci. **4**, 257–261
37. Ahmed, H. & Chaterjee, B. P. (1989) J. Biol. Chem. **264**, 9365–9372
38. Appukuttan, P. S. & Basu, D. (1985) FEBS Lett. **180**, 331–334
39. Blundell, T. L., Jenkins, J. A., Sewell, B. T., Pearl, L. H., Cooper, J. B., Tickle, I. J., Veerapandian, B. & Wood, S. P. (1990) J. Mol. Biol. **211**, 919–941
40. Wright, C. S. (1989) J. Mol. Biol. **209**, 475–487

Received 1 July 1991/28 October 1991; accepted 4 November 1991