Volume 282, number 2, 382–384 FEBS 09706 May 1991

ADONIS 001457939100415U

The amino acid sequences of jacalin and the *Maclura pomifera* agglutinin

N. Martin Young Rosemary A.Z. Johnston and David C. Watson

Institute for Biological Sciences, National Research Council of Canada, Ottawa, Ont. K1A 0R6, Canada

Received 14 February 1991; revised version received 18 March 1991

Amino acid sequences for the α-chains of the Moraceae lectins, jacalin and *Maclura pomifera* agglutinin, were determined by protein sequencing. Both are 133 residues long and contain several genetically variant positions; the overall homology is 85%. A possible site for the known glycopeptide of jacalin was located. The α-chains have a conserved tryptophan residue that may be part of the binding-site.

Jacalin; *Artocarpus integrifolia* lectin; *Maclura pomifera* agglutinin; Amino acid sequence

1. INTRODUCTION

The *Maclura pomifera* agglutinin and jacalin, the *Artocarpus integrifolia* lectin, are homologous seed lectins that are specific for the T-antigen structure Galβ1-3GalNAc [1]. This specificity has made jacalin useful for studies of *O*-glycosylated proteins, particularly human IgA [2]. The lectins are from plants of the Moraceae family and are not homologous to legume or other plant lectins. They have unusual quaternary structures, combining α-chains of apparent M_r 11–12 kDa with 20–21 residue β-chains [1]. Crystallographic analysis of MPA indicated an $\alpha_2\beta_2$ or $\alpha_4\beta_4$ structure [3]; the M_r value of ~48 kDa from gel-filtration is not conclusive for either form.

The amino acid sequences of the various forms of the β-chains and the N-terminal sequences of the α-chains were previously described [1]. We report the complete sequences of the jacalin and MPA α-chains, which will be required for the X-ray crystallographic studies underway on both proteins [3–5].

2. MATERIALS AND METHODS

Jacalin was obtained from Pierce Chemical Co. and MPA was prepared as previously described [1]. The α-chain of jacalin was separated from the β-chains by reverse-phase HPLC on a C_4 column [1]. Cleavage at methionine residues with CNBr was carried out in 88% formic acid under argon at room temperature for 24 h. For cleavage at aspartic acid residues, the proteins were digested in 0.25 M acetic acid at 110°C for 7 h [6]. Asn–Gly peptide bonds were cleaved in succinylated proteins with 2 M hydroxylamine in 6 M guanidinium chloride [7] at 45°C for 4 h. Trypsin, purified by HPLC to remove chymotryptic contaminants [8], was used at an enzyme/substrate ratio of 1:50 and digestion was at 37°C for 16 h, in 0.2 M ammonium bicarbonate. Chymotryptic digests were carried out similarly for 4 h. Digestions with pepsin were for 15 min at room temperature with a 1:50 enzyme/substrate ratio. The peptide products were separated by reverse-phase HPLC, using 0.1% trifluoroacetic acid buffer, pH 2.0, and acetonitrile gradients, and a C_4 column of 5 μm material (100 × 4.6 mm, Synchropak RP-4, Synchrom Inc. or Hypersil WP-300, Shandon Southern Products).

Amino acid analyses were performed with a Durrum D-500 analyser. The protein and peptide samples were hydrolysed in vacuo at 110°C in 6 M HCl with 1% phenol for 22 h, or, for tryptophan analysis, in 4 M methane/sulfonic acid. Automated gas-phase sequencing was performed on an Applied Biosystems model 475A sequencer with on-line identification of the phenylthiohydantoin derivatives. Molecular masses were determined for the intact proteins by B.N. Green (VG Analytical Ltd., Manchester) using a VG mass spectrometer equipped with a caesium ion gun [9].

3. RESULTS

The jacalin α-chain sequence was obtained from tryptic peptides supplemented with other sets (Fig. 1) together with N-terminal analysis of succinylated jacalin cleaved with hydroxylamine. The C-terminus was identified from a small CNBr peptide. It was apparent from the tryptic peptide map that there were more peptides than expected from the composition and estimated molecular weight. From minor peptides, genetic variants were identified at seven positions (Fig. 2). The calculated molecular mass for the main sequence is 14662.6 Da in good agreement with the mass spectrometry value of 14657 ± 7.

The SDS-PAGE of jacalin shows a minor glycosylated species of apparent M_r 15 kDa [10]. Its N-terminal sequence, obtained by transferring the band to a membrane for sequencing [11], was identical to that of the α-chain to residue 20.

The α-chain of MPA could not be obtained in good yield by HPLC, so peptide fragmentations were carried

Abbreviations: HPLC, high-performance liquid chromatography; MPA, *M. pomifera* agglutinin; SDS-PAGE, sodium dodecylsulphate polyacrylamide gel electrophoresis

Correspondence address: N.M. Young, Institute for Biological Sciences, National Research Council of Canada, 100 Sussex Drive, Ottawa, Ont., K1A 0R6, Canada

Published by Elsevier Science Publishers B.V.

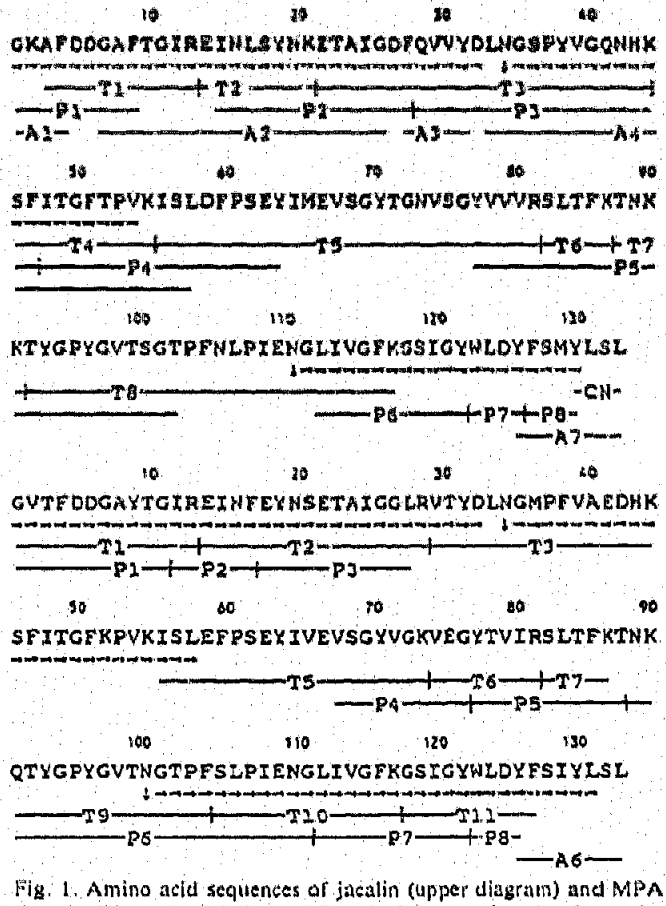

Fig. 1. Amino acid sequences of jacalin (upper diagram) and MPA (lower diagram) α-chains. N-Terminal sequencing is shown by arrows, and the trypsin, pepsin, chymotrypsin, acid cleavage, and CNBr cleavage peptide products are identified by T, P, Ch, A and CN, respectively. The hydroxylamine cleavage points are indicated with vertical arrows.

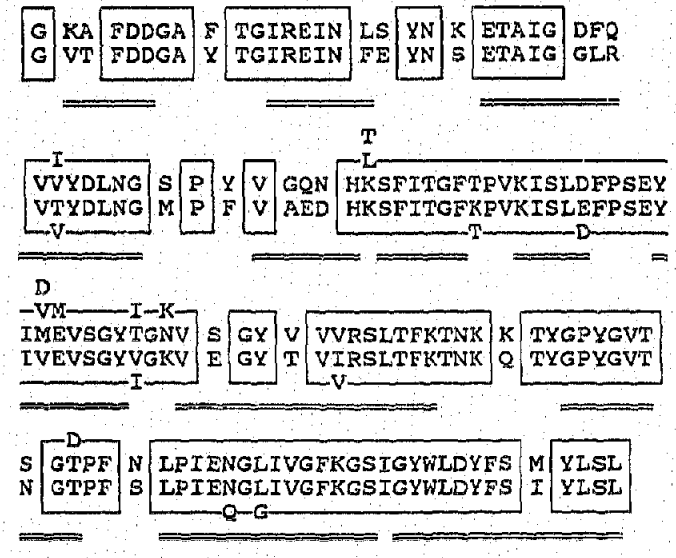

Fig. 2. Sequence alignment of jacalin α-chain (upper sequence) and MPA α-chain (lower sequence). Sequence variants are shown above or below the respective primary sequences. Predicted β-sheet segments [14] for jacalin are shown with horizontal lines.

Table I

Amino acid compositions of jacalin α-chain and MPA

Amino acid	Amount (mol/mol)[a]	
	Jacalin α-chain	MPA $\alpha+\beta_1$
Asp	13.92 (14)	15.34 (15)
Thr	10.20 (10)	11.94 (12)
Ser	11.40 (12)	11.62 (12)
Glu	7.18 (7)	11.22 (11)
Pro	5.87 (6)	6.16 (8)
Gly	16.45 (17)	19.71 (21)
Ala	3.05 (3)	3.29 (3)
Val	10.24 (11)	13.03 (13)
Met	1.53 (2)	0.86 (1)
Ile	8.18 (9)	11.73 (13)
Leu	9.13 (9)	8.76 (9)
Tyr	10.89 (11)	10.51 (11)
Phe	9.98 (10)	9.70 (10)
His	1.16 (1)	1.17 (1)
Lys	8.15 (8)	8.17 (8)
Arg	1.91 (2)	4.21 (4)
Trp	1.06 (1)	1.73 (2)

[a] Values in parentheses are the number of residues deduced from the major sequences, MPA including its β_1 sequence [1]. Thr, Ser and Val values are from extrapolations of 24, 48 and 66 h hydrolyses

out on whole MPA. The sequence (Fig. 1) is also 133 residues long, with genetic variants at seven positions (Fig. 2), two of which were variable in jacalin. Mass spectrometry indicated the molecular mass of the MPA α-chain was 14758 ± 10 Da, and the main sequence shown corresponds to 14757.9. The correspondence of the compositions derived from the MPA and jacalin sequences to those from amino acid analyses are shown in Table I. An alignment of the two sequences (Fig. 2) showed 85% overall homology.

4. DISCUSSION

The 133-residue sequences and mass spectrometric data give molecular masses considerably above the SDS-PAGE values of 11–12 kDa [1]. There is no obvious feature in the sequence that might account for the anomalous behaviour. The glycine contents of the proteins are a little higher than usual, but these residues are distributed throughout the sequences rather than forming a collagen-like region or other unusual structure. Both proteins have genetic variants at several positions, detected from minor peptides. A minimum of three genes for jacalin and two for MPA are indicated by the α- and β-chain [1] data. It is remarkable that both proteins crystallise well [3–5] despite the variant residues.

Jacalin has a minor glycosylated form [10] and it has potential *N*-glycosylation sites at Asn-16, Asn-35 and Asn-74 (in one variant). The structure of the jacalin glycopeptide has been reported [12] and its ^{1}H NMR spectrum contains threonine resonances, indicating

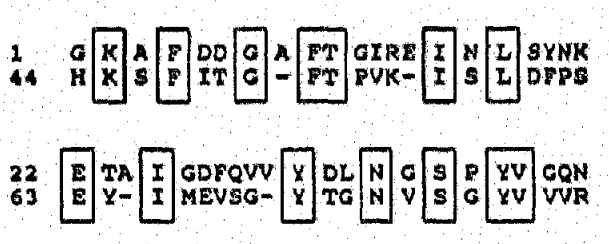

Fig. 3. Possible internal duplication in the first 82 residues of the jacalin α-chain.

that Asn-74 is the substitution site since it has a threonine nearby, Thr-72. MPA has only one potential glycosylation site, at Asn-100.

A single tryptophan residue was found in both α-chains, Trp-123. This or the tryptophan in the β-chain must be near the carbohydrate-binding site since there is a large increase in intrinsic fluorescence upon ligand binding [1]. Tyrosine residues are also implicated in the site [13], hence the α-chain Trp-123 appears more likely, since Tyr-122 and Tyr-126 are close to it. Though the carbohydrate specificities of the two proteins are very similar, there are some differences in their binding-site properties. MPA has a K_a for methyl α-D-galactoside that is half the jacalin value and the change in intrinsic fluorescence on sugar binding is twice as large [1]. The sequence differences between the proteins therefore have functional consequences.

The CD spectra of jacalin and MPA showed that they are β-sheet proteins [1]. Structure prediction [14] for the α-chains (Fig. 2) agreed with this assignment, suggesting 9–11 β-strands. Several of the intervening segments have dipeptide segments often found in β-turns, such as Asn–Gly, common in Type I′ turns, and Gly–Ser, common in Type II′ turns [15].

One interesting feature of the jacalin sequence is a possible internal repeat. Alignment of residues 1–43 against residues 44–83 (Fig. 3) gives 33% identity, with several conservative changes. Each subunit may therefore be divided into two segments of 40–43 residues, each containing four β-strands, and a third segment of 50 residues from the α-chain, together with one β-chain. This structural arrangement is similar to that of the B-chain of ricin, which has a peptide fold of 35 residues that is repeated three times in each of its two domains, together with one additional peptide segment [16]. However, the lectins show no sequence similarity to the ricin B-chain, nor to any other reported sequence. The X-ray studies underway on jacalin [4,5] and MPA [3] should give definitive structural information on these proteins.

Acknowledgements: We thank Dr M. Yaguchi for amino acid analyses, Dr M. Zuker for computer analyses of the sequences, and B.N. Green of VG Analytical, Manchester, for the mass spectrometry.

REFERENCES

[1] Young, N.M., Johnston, R.A.Z., Szabo, A.G. and Watson, D.C. (1989) Arch. Biochem. Biophys. 270, 596–603.
[2] Roque-Barreira, M.C. and Campos-Neto, A. (1985) J. Immunol. 134, 1740–1745.
[3] Lee, X., Johnston, R.A.Z., Rose, D.R. and Young, N.M. (1989) J. Mol. Biol. 210, 685–686.
[4] Basu, D., Delucas, L., Parks, E.M. and Suddath, F.L. (1988) J. Mol. Biol. 201, 661–662.
[5] Dhanaraj, V., Patanjali, S.R., Surolia, A. and Vijayan, M. (1988) J. Mol. Biol. 203, 1135–1136.
[6] Schultz, J. (1967) Methods Enzymol. 11, 255–263.
[7] Bornstein, P. and Balian, G. (1977) Methods Enzymol. 47, 132–145.
[8] Titani, K., Sasagawa, T., Resing, R. and Walsh, K.A. (1982) Anal. Biochem. 123, 408–412.
[9] Barber, M. and Green, B.N. (1987) Rapid Commun. Mass Spectrosc. 1, 80–83.
[10] Aucouturier, P., Mihaesco, E., Mihaesco, M. and Preud'homme, J.-L. (1987) Mol. Immunol. 24, 503–511.
[11] Matsudaira, P. (1987) J. Biol. Chem. 262, 10035–10038.
[12] Capon, C., Piller, F., Wieruszeski, J.-M., Leroy, Y. and Fournet, B. (1990) Carbohyd. Res. 199, 121–127.
[13] Mahanta, S.K., Sastry, M.V.K. and Surolia, A. (1990) Biochem. J. 265, 831–840.
[14] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[15] Sibanda, B.L. and Thornton, J.M. (1985) Nature 316, 170–174.
[16] Rutenbar, E., Ready, M. and Robertus, J.D. (1987) Nature 326, 624–626.