Lectins and legume chemotaxonomy

Characterisation of the *N*-acetyl-D-galactosamine specific lectin of *Bauhinia purpurea*

N. Martin Young, David C. Watson and Ross E. Williams

Division of Biological Sciences, National Research Council of Canada, Ottawa K1A OR6, Canada

Received 10 December 1984; revised version received 30 January 1985

The composition, circular dichroism (CD) and N-terminal amino acid sequence of the lectin from *Bauhinia purpurea*, a member of the Caesalpinioideae subfamily, are reported. The lectin has one cysteine residue and the CD in the aromatic region is unusually high for plant lectins. The N-terminal sequence shows much stronger homology to that of another Caesalpinioideae lectin *Griffonia simplicifolia* GSI than to sequences of Papilinoideae lectins. However, its composition and CD differ from those of GSI, which more closely resemble those of Papilinoideae lectins of similar specificity.

Bauhinia purpurea *Circular dichroism* *Lectin* *N-terminal amino acid sequence*

1. INTRODUCTION

The increasing body of data on lectins from leguminous plants has led to proposals that these proteins may be useful in chemotaxonomy [1,2] particularly as N-terminal amino acid sequences are accumulated [3,4]. A simple test of their chemotaxonomic value is whether the division of the Leguminosae into the subfamilies Caesalpinioideae, Mimosoideae and Papilinoideae [5] can be discerned from the structures of their lectins. However, nearly all the structurally well-characterised lectins are from the Papilinoideae, and only the *Griffonia* (*Bandeiraea*) *simplicifolia* lectins [6] come from the Caesalpinioideae. To provide more information on Caesalpinioideae lectins, we investigated the structural properties of the *N*-acetyl-D-galactosamine specific lectin from *Bauhinia purpurea* which was partially characterised by Irimura and Osawa [7].

2. MATERIALS AND METHODS

B. purpurea lectin was purified from a crude commercial preparation (Pharmacia P-L Biochemicals) by affinity chromatography on an *N*-acetyl-D-galactosamine derivative [8] of Affi-Gel 10 (BioRad Laboratories), using 0.1 M lactose as the specific eluant. The gel electrophoresis and circular dichroism (CD) properties of this preparation were very similar to those of a commercially purified sample (E.Y. Laboratories, San Mateo, CA).

The molecular mass of the subunits was determined by SDS gel electrophoresis [9], and of the native protein by chromatography on a calibrated Sephadex G-200 column run in 0.15 M NaCl, 0.01 M Na phosphate (pH 7.3). Amino acid compositions were determined with a Durrum D500 analyser on samples hydrolysed for 20 h at 110°C in 6 N HCl, with or without prior oxidation with performic acid. Amino sugars were determined on the D500 analyser with samples hydrolysed for 6 h at 110°C in 4 N HCl.

CD spectra were measured with a Cary 61 spectropolarimeter. The samples were in 0.1 M NaCl, 0.01 M Na acetate buffer (pH 5.4) and the sample temperature was approximately 27°C. The data are expressed as mean residue molar ellipticities in

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

deg·cm^2·dmol^{-1} assuming a mean residue weight of 110. The extinction coefficient, $A_{280nm}^{1mg/ml}$, was 1.43, as determined by differential refractometry.

N-terminal amino acid sequence data were obtained using a Beckman 890D sequencer with a 1.0 M Quadrol program. The 2-anilino-5-thiazolinone derivatives were converted to phenylthiohydantoin derivatives by treatment with 20% trifluoroacetic acid at 65°C for 20 min. These derivatives were identified by reverse-phase HPLC using a Beckman PTH-Ultrasphere ODS column on a Varian Vista 55 chromatograph.

3. RESULTS

The composition of the *B. purpurea* lectin (table 1) was in reasonable agreement with the data of Irimura and Osawa [7] except for tryptophan and cysteine, for which their data indicate 4 and 7 residues, respectively, and the carbohydrate content was higher than their value of 11.1%. The presence of a single cysteine residue was confirmed by radio-alkylation of fully reduced protein with iodo[2-^{14}C]acetic acid. The considerable content of neutral carbohydrate (18%) is at least partially responsible for the apparent subunit molecular mass of 32 kDa as determined by SDS gel electrophoresis, being greater than that of concanavalin A and other lectins [9]. The peptide molecular mass is approx. 26 kDa. The molecular mass of the native lectin was estimated at 100 kDa by Sephadex G-200 gel filtration comparable with the value of 120 kDa obtained by size-exclusion HPLC [9].

The CD spectrum in the far UV region (fig.1) is of low intensity and analysis by the Contin procedure [12] indicated 65% β-sheet, 19% β-turn and no α-helix. In the near UV region, the CD spectrum consisted of a complex set of positive bands that showed small decreases when the ligand *N*-acetyl-D-galactosamine was added.

Amino acid sequencing of native and radio-alkylated lectin gave the N-terminal sequence to residue 34. This sequence is shown in fig.2 along with previously reported sequences for the two isolectin chains of the Caesalpinioideae lectin, *G. simplicifolia* GSI [6], and of two Papilinoideae lectins, the peanut [3] and soybean agglutinins [13]. No derivatives could be identified for residues 20 and 29. Since residues 22 and 31 are both threonine and the residues in GSIA corresponding to 20 and 29 are, respectively, asparagine and aspartic acid, it is possible that residues 20 and 29 are positions of carbohydrate attachment. At position 27 both glutamine and leucine were found indicating that isolectin forms are present; the protein shows two close bands in SDS gel electrophoresis [9]. No significant radioactivity was detected in any of the derivatives obtained from the sequencing of a

Table 1

Composition of *B. purpurea* lectin

Component	Residues per 32 kDa	Component	Residue per 32 kDa
Cys[a]	0.8	Ile	12.5
Asx	26.8	Leu	16.8
Thr	24.5	Tyr	11.5
Ser	22.5	Phe	16.1
Glx	15.9	His	8.5
Pro	9.8	Lys	7.6
Gly	20.6	Arg	7.6
Ala	13.2	Trp[b]	5.1
Val	13.9	Glucosamine	10.7
Met	0.6	Hexose[c]	22.3

[a] Determined as cysteic acid
[b] Determined spectrophotometrically [10]
[c] Determined by the orcinol-H_2SO_4 assay [11] with D-mannose as standard

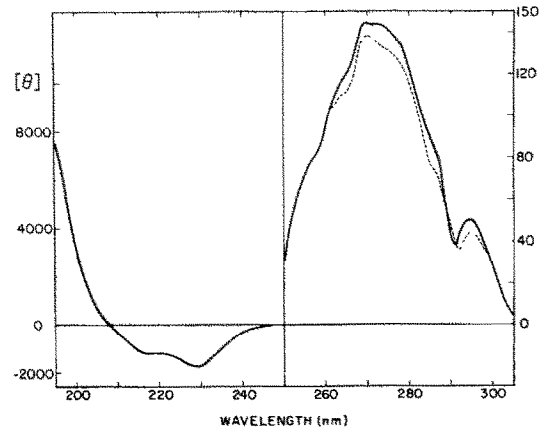

Fig.1. CD spectra of native *B. purpurea* lectin (——) and in the presence of 5 mM *N*-acetyl-D-galactosamine (– – –).

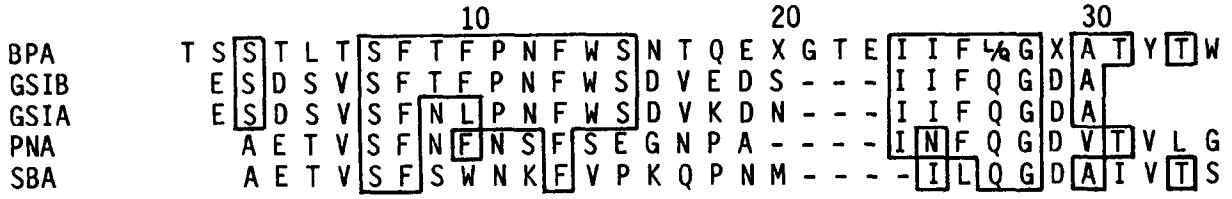

Fig.2. N-terminal amino acid sequences of the *B. purpurea* lectin (BPA); *G. simplicifolia* GSI isolectins (GSIA and GSIB [6]); peanut agglutinin (PNA [3]) and soybean agglutinin (SBA [13]). The dashes indicate gaps introduced to maximise homology and the residues X at positions 20 and 29 are discussed in the text.

[^{14}C]carboxymethylated sample, hence the cysteine residue lies outside the region sequenced. A sequence to residue 27, previously reported without experimental details [4], differs from the present sequence in assigning leucine to residue 20, serine to residue 23, and only glutamine to residue 27.

4. DISCUSSION

The molecular mass data indicate that the *B. purpurea* lectin has a tetrameric structure, as do nearly all lectins of similar specificity. Probably due to its carbohydrate content, its subunit weight is a little higher than that of other lectins [9]. However, a molecular mass of 195 kDa was obtained by ultracentrifugation [7] suggesting that the lectin undergoes aggregation. Its metal analysis [14] showed an Mn^{2+} content below that of the other lectins, but close to that of the *G. simplicifolia* GSI lectin. The carbohydrate content is much greater than that of other legume lectins, and the hexose:glucosamine ratio indicates the structure(s) will be different from that of the carbohydrate unit of soybean agglutinin [15].

The *B. purpurea* lectin has a single cysteine residue, a relatively rare feature in legume lectins. The other Caesalpinioideae lectin, *G. simplicifolia* GSI, also has one cysteine per subunit which chemical modification experiments indicate to be not essential for carbohydrate binding [16], in contrast to the cysteine of the lima bean lectin [17]. More sequence data are required to establish if the GSI and *B. purpurea* lectin cysteines are homologous. The present data also show that the cysteine is not in the N-terminal region unlike that of the sainfoin lectin [18].

The amino acid compositions of *B. purpurea* and GSI lectins were compared with compositions determined in this laboratory for fourteen Papilinoideae lectins of various specificities, using the composition index of Cornish-Bowden [19]. Comparisons of the *B. purpurea* lectin with all the other lectins gave index values well outside the strong test criterion of Cornish-Bowden [20] and hence it appeared only weakly related to them. In contrast, the GSI lectin gave values indicating homology to several Papilinoideae lectins including those from *Cytisus scoparius, Cytisus sessilifolius, Caragana arborescens, Dolichos biflorus* and soybean agglutinin and concanavalin A.

The CD spectrum of *B. purpurea* lectin is a distinctive one. In the 200–240 nm region, its magnitude is below the 4000–6000 deg·cm^2·dmol^{-1} of other lectins. In the 250–290 nm region, it is of greater magnitude and has a maximum 10–20 nm lower than those of other lectins specific for *N*-acetyl-D-galactosamine or D-galactose including the GSI lectin (unpublished).

While the above analysis indicates no particular relationship between the *B. purpurea* and GSI lectins, the more definitive criterion of N-terminal amino acid sequence clearly shows a strong homology between the two (fig.2) and also that they are distinctly different from the Papilinoideae lectins. Furthermore, with 16 identities in the region sequenced, the *B. purpurea* and GSI lectins are more homologous than any other pair of lectins specific for *N*-acetyl-D-galactosamine or D-galactose. Peanut and soybean agglutinins have 12 residues in common which is higher than other pairs, and had, respectively, 9 and 8 residues homologous with *B. purpurea* lectin and 10 and 8 with GSIA. These are one or two more identities than other lectins showed.

It is therefore possible to discern the division of the legumes into these two subfamilies in the N-terminal sequence data for the lectins, but other properties give more equivocal results for the GSI lectin. Hence, for chemotaxonomic purposes the N-terminal sequence data appear to be the only suitable parameter.

REFERENCES

[1] Toms, G.C. and Western, A. (1971) in: Chemotaxonomy of the Leguminosae (Harborne, J.B., Boulter, D. and Turner, B.L. eds) pp.367–462, Academic Press, New York.

[2] Toms, G.C. (1981) in: Advances in Legume Systematics (Polhill, R.M. and Raven, P.H. eds) part 2, pp.561–577, Royal Botanic Gardens, Kew.

[3] Foriers, A., De Neve, R. and Strosberg, A.D. (1979) Physiol. Veg. 17, 597–606.

[4] Strosberg, A.D., Lauwereys, M. and Foriers, A. (1983) in: Chemical Taxonomy, Molecular Biology and Function of Plant Lectins (Goldstein, I.J. and Etzler, M.E. eds) pp.7–20, Alan R. Liss, New York.

[5] Polhill, R.M., Raven, P.H. and Stirton, C.H. (1981) in: Advances in Legume Systematics (Polhill, R.M. and Raven, P.H. eds) part 2, pp.1–26, Royal Botanic Gardens, Kew.

[6] Lamb, J.E. and Goldstein, I.J. (1984) Arch. Biochem. Biophys. 229, 15–26.

[7] Irimura, T. and Osawa, T. (1972) Arch. Biochem. Biophys. 151, 475–482.

[8] Young, N.M., Watson, D.C. and Williams, R.E. (1984) Biochem. J. 222, 41–48.

[9] Young, N.M. and Jackson, G.E.D. (1984) J. Chromatogr. (Biomed. Appl.) 336, 397–402.

[10] Edelhoch, M. (1967) Biochemistry 6, 1948–1954.

[11] Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A. and Smith, F. (1956) Analyt. Chem. 28, 350–356.

[12] Provencher, S.W. and Glockner, J. (1981) Biochemistry 20, 33–37.

[13] Vodkin, L.O., Rhodes, P.R. and Goldberg, R.B (1983) Cell 34, 1023–1031.

[14] Young, N.M. (1983) FEBS Lett. 161, 247–250.

[15] Dorland, L., Van Halbeek, H., Vliegenthart, J.F.G., Lis, H. and Sharon, N. (1981) J. Biol. Chem. 256, 7708–7711.

[16] Lönngren, J. and Goldstein, I.J. (1976) Biochim. Biophys. Acta 439, 160–166.

[17] Gould, N.R. and Scheinberg, S.L. (1970) Arch. Biochem. Biophys. 141, 607–613.

[18] Young, N.M., Williams, R.E., Roy, C. and Yaguchi, M. (1982) Can. J. Biochem. 60, 933–941.

[19] Cornish-Bowden, A. (1979) J. Theor. Biol. 65, 735–742.

[20] Cornish-Bowden, A. (1980) Anal. Biochem. 105, 233–238.