

A Structural Comparison of the A and B Subunits of *Griffonia simplicifolia* I Isolectins¹

JAMIE E. LAMB² AND IRWIN J. GOLDSTEIN³

Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109

Received July 19, 1983, and in revised form October 5, 1983

A structural comparison between the A and B subunits of the five tetrameric *Griffonia simplicifolia* I isolectins (A₄, A₃B, A₂B₂, AB₃, B₄) was undertaken to determine the extent of homology between the subunits. The first 25 N-terminal amino acids of both A and B subunits were determined following the enzymatic removal of N-terminal pyroglutamate blocking groups with pyroglutamate aminopeptidase. Although 21 amino acids were common to both subunits, there were four unique amino acids in the N-terminal sequence of A and B. Residues 8, 9, 17, and 19 were asparagine, leucine, lysine, and asparagine in subunit A and threonine, phenylalanine, glutamic acid, and serine in subunit B. The last six C-terminal amino acids, released by digestion with carboxypeptidase Y, were the same for both subunits: Arg-(Phe, Val)-Leu-Thr-Ser-COOH. Subunit B, which contains one methionyl residue, was cleaved by cyanogen bromide into two fragments, a large ($M_r = 31,000$) and a small ($M_r = 2700$) polypeptide. Failure of the small fragment to undergo manual Edman degradation indicated an N-terminal blocking group, presumably pyroglutamate. Both subunits were digested with trypsin and the tryptic peptides were analyzed using reverse-phase HPLC. Tryptic glycopeptides were identified by labeling the carbohydrate moiety of the A and B subunit using sodium [³H] borohydride. Cysteine-containing tryptic peptides were similarly identified by using [1-¹⁴C]iodoacetamide. Approximately 30% of the tryptic peptides were common to both subunits. Thus, although the N- and C-terminal regions of A and B are similar, the subunits each possess unique sequences.

One of the four lectins (1) present in *Griffonia simplicifolia* (GS)⁴ seeds exhibits

anti-B hemagglutinating activity (2). This lectin, designated GS-I, is composed of a family of five tetrameric α -D-galactosyl binding isolectins (3) which contain two distinct subunits, A and B. The isolectins are designated A₄, A₃B, A₂B₂, AB₃, and B₄. Although both subunits bind terminal nonreducing α -D-galactosyl groups with approximately the same affinity (4, 5), the A subunit binds 2-acetamido-2-deoxy- α -D-galactosyl end groups with a 1000-fold greater affinity than the B subunit (6). Hence the B subunit (and therefore the B₄ isolectin) is highly specific for α -D-galactosyl end groups (4-6).

Murphy and Goldstein (4) performed an extensive physicochemical characterization of the A and B subunits. They found that the subunits shared many common antigenic determinants. The amino acid

¹ Dedicated to the memory of Lee A. Murphy. This study was supported by Public Health Service Grant GM 29470 from the National Cancer Institute. This study was submitted by J.E.L. to the University of Michigan in partial fulfillment of the requirements for the Ph.D. degree in Biological Chemistry.

² Present address: Department of Pathology, University of Utah Medical Center, Salt Lake City, Utah 84132.

³ To whom correspondence should be addressed.

⁴ Abbreviations used: Con A, concanavalin A; GS, *Griffonia simplicifolia*; SDS, sodium dodecyl sulfate; TPCK, L-(tosylamido-2-phenyl)ethylchloromethyl ketone; PTH, phenylthiohydantoin; TCA, trichloroacetic acid; CPY, carboxypeptidase Y; ATEE, N-acetyl-L-tyrosine ethyl ester; TRH, thyrotropin releasing hormone; PAP, pyroglutamate aminopeptidase.

compositions were very similar, the most notable difference being the presence of one methionyl residue in B and none in A. Both subunits contained one cysteine but these were not equally reactive with the thiol reagent 5,5'-dithiobis(2-nitrobenzoic) acid. The minimum molecular weight, calculated from the amino acid and carbohydrate analysis, was 30,800 for A and 30,400 for B. However, the apparent molecular weights based on SDS-polyacrylamide gel electrophoresis were 32,000 for A and 33,000 for B (3). The carbohydrate composition of both subunits was also similar, with A containing 6.9% and B 7.2%, of the same neutral sugars.

Recently, Lamb *et al.* (7) postulated, on the basis of statistical data, that the B subunit might be a precursor of the A subunit via a post-translational proteolysis which included removal of the sole methionyl residue from B. In order to test the validity of this hypothesis, we have undertaken a structural comparison between the A and B subunits.

The structural studies described herein include N- and C-terminal sequencing, cyanogen bromide cleavage of B subunits followed by separation of the fragments on Sephadex LH-60, and tryptic peptide mapping of the subunits using reverse-phase HPLC.

MATERIALS AND METHODS

G. simplicifolia seeds and *N*-acetyl-L-tyrosine ethyl ester were purchased from Calbiochem (La Jolla, Calif.). Carboxypeptidase Y and pyroglutamate aminopeptidase were purchased from Boehringer-Mannheim Biochemicals (Indianapolis, Ind.). TPCK-trypsin was obtained from Worthington (Freehold, N. J.). Type VI horse heart cytochrome *c*, bovine erythrocyte carbonic anhydrase, and type III horse heart myoglobin were purchased from Sigma Chemical Company (St. Louis, Mo.). Thyrotropin-releasing hormone was obtained from Vega Biochemicals (Tucson, Ariz.). HPLC solvents were obtained from Burdick and Jackson Labs, Inc. (Muskegon, Mich.). Sodium [³H] borohydride, (142.0 mCi mmol⁻¹) and [¹⁴C]iodoacetamide (19.1 mCi mmol⁻¹, 100 μCi ml⁻¹) were purchased from New England Nuclear (Boston, Mass.).

Purification of G. simplicifolia I isolectins. The method of Delmotte and Goldstein (8) was employed for the purification of the five *G. simplicifolia* I isolectins. Nondenaturing pH 8.9 polyacrylamide gel

electrophoresis (9) was employed to determine the purity of the GS-I isolectins. The purity of the A and B subunits was assessed using SDS-gel electrophoresis in the presence of 2-mercaptoethanol as described by Laemmli (10). Only isolectins showing greater than 95% homogeneity were employed in the structural analyses.

Reduction and alkylation of A and B subunits. The A₄, B₄, and A₂B₂ isolectins were reduced and alkylated in 0.1 M NH₄HCO₃ buffer, pH 8.2, containing 8 M urea and 1 mM Na₂EDTA (11). A 1.5-fold molar excess of 1.0 mM dithiothreitol solution was added and reduction was allowed to proceed for 3 h at room temperature under N₂. A 1.5-fold molar excess of iodoacetamide (25 mM), purified according to the method of Stark and Stein (12), was added under N₂ in the dark. After 4 h at room temperature the solutions were dialyzed against double-distilled H₂O and lyophilized.

The alkylation reaction was also performed using [1-¹⁴C]iodoacetamide diluted with unlabeled iodoacetamide in a 1:6 molar ratio, respectively. The ¹⁴C radioactivity was measured in aqueous counting scintillant (ACS, Amersham). The A₄ and B₄ isolectins had specific activities of 55 × 10⁶ and 62 × 10⁶ cpm μmol⁻¹, respectively.

Carboxypeptidase Y digestion and amino acid analysis. The carboxypeptidase Y (CPY) digestion was performed following the method of Black *et al.* (13) with the modifications described below. Reduced and alkylated A₄ or B₄ isolectins were used in each reaction. The isolectins were dissolved in 50 mM potassium phosphate, pH 6.0, containing 1% SDS. Following heating at 80–90°C for 15 min, the SDS was diluted to 0.5% with 50 mM potassium phosphate, pH 6.0 (CPY is active at concentrations of SDS ≤ 0.5%). The denatured lectins (1 mg ml⁻¹) were preincubated at 25°C for 15 min before the addition of CPY (molar ratio lectin: CPY ~ 74:1). Samples (10–15 nmol of A or B subunits) were withdrawn at timed intervals and the CPY activity was quenched immediately with 10% TCA (0°C). A thick slurry of precycled Dowex 50×-12 (0.8 ml) was added to each sample and incubated at 0°C for 60 min to ensure maximum binding of the released amino acids. The Dowex slurry was packed into a 3-ml plastic syringe, plugged with glass wool, and washed with double-distilled H₂O (20 ml). Bound amino acids were eluted with 5 M NH₄OH (3 × 0.5 ml). The eluates were combined and the NH₃ was removed by evaporation under vacuum. The amino acids contained in each sample were analyzed on a Beckman 120 C amino acid analyzer using a 6-mm column packed with Beckmann W-2 resin. As a control, a standard sample of amino acids (Pierce Standard H) was treated with CPY and analyzed in parallel with the lectin samples. The kinetics, and the identity, of each amino acid released were calculated after correcting for the losses incurred by the corresponding amino acid in the external standard.

Because serine and threonine coelute with glutamine and asparagine during the amino acid analyses, we hydrolyzed one set of samples (6 N HCl, 110°C, 24 h, under N₂) and converted glutamine and asparagine to glutamic and aspartic acids which do not coelute with serine and threonine. Positive identification of serine and threonine, in the C-terminal sequence, was thereby made possible.

CPY was assayed using the substrate *N*-acetyl-L-tyrosine ethyl ester (ATEE). A 1 mM solution of ATEE in 0.05 M potassium phosphate, pH 7.5, was incubated (25°C) with CPY and the decrease in absorbance at 237 nm (1 OD = 10 μmol) was recorded. One unit of enzyme activity will hydrolyze 1 μmol of ATEE per minute under these conditions. The CPY had an activity of 70–80 units mg⁻¹.

Pyroglutamate aminopeptidase digestion and N-terminal sequence determination. Thyrotropin releasing hormone (TRH) and reduced and alkylated A₄, A₂B₂, and B₄ were digested with pyroglutamate aminopeptidase (PAP). TRH has the sequence L-pyroglutamyl-L-histidyl-L-prolineamide and served as a control substrate. The substrates (A₄, A₂B₂, B₄, or TRH) were weighed into screw-cap culture tubes and dissolved (1 mg ml⁻¹) in 0.1 M potassium phosphate buffer, pH 7.7, containing 10 mM Na₂EDTA, 5 mM dithiothreitol, and 0.05% glycerol (unblocking buffer). Sodium cholate (0.5%) was added and the solution was heated to 60°C for 10 min. Lyophilized PAP was added (lectin:PAP, molar ratio 80:1), the tube flushed with N₂, and the incubation continued at 4°C for 18–24 h. A second, equal aliquot of PAP was added and the tube was flushed with N₂. After 16–24 h at 4°C the proteins were precipitated with 50% TCA and removed by centrifugation in a Beckmann Microfuge B (2 min). The pellet, containing unblocked lectin subunits, was saved for N-terminal sequence analysis and the supernatant was mixed with a thick slurry of Dowex 50×-12 (0.8 ml). The Dowex was transferred to a 3-ml disposable syringe, plugged with glass wool. After washing with double-distilled H₂O (10 ml) the combined eluates were evaporated under vacuum. Acid hydrolysis was performed as described above. The hydrolyzate was passed over a second Dowex 50×-12 column (0.8 ml). Amino acids (notably glutamic acid) which bound to the second Dowex column were eluted with 5 M NH₄OH, the eluate was evaporated, and the amino acids were analyzed as described above. A control sample containing only unblocking buffer and PAP was run in parallel and the samples were corrected accordingly. The recovery of pyroglutamate from TRH was approximately 22%. A 30% recovery was usually obtained from the lectins.

The TCA-precipitated pellets were washed extensively with 10% acetic acid and subjected to N-terminal sequence analysis. An automated degradation procedure (14) and a Beckmann Model 890 C spinning-cup automated sequencer using a modified Beckmann

program (No. 122972) were employed (15). PTH amino acids were identified by reverse-phase HPLC (16). Between 65 and 75 nmol of unblocked A and B subunits were subjected to Edman degradation. Initial and repetitive yields were approximately 90%. Data were compiled from three N-terminal sequence analyses of each subunit. Manual Edman sequencing was performed according to the method of Tarr (17, 18). Approximately 2–10 nmol of peptide were employed. Norleucine was added as the internal standard and the PTH amino acids were identified by reverse-phase HPLC (16).

Cyanogen bromide cleavage. The cyanogen bromide (CNBr) cleavage reaction was performed in 70% formic acid (19). CNBr (100 g CNBr in 50 ml acetonitrile) was added in a lectin:CNBr molar ratio of 1:100. The reaction mixture was incubated at room temperature for 24 h in the dark. Twenty volumes of double-distilled H₂O was added and the solution was thoroughly degassed to remove HCN. The solution was lyophilized and the CNBr fragments were stored at 4°C. Cytochrome *c* and myoglobin were also cleaved using the same procedure. CNBr fragments of B subunits were separated by hydrophobic chromatography using Sephadex LH-60 (114 × 1.5 cm) in formic acid and ethanol (1:4 v/v) (20), or, more simply, by dialysis (acetonitrile:0.1% H₃PO₄, 1:1 v/v) since the small fragment was not retained by the dialysis membrane. The molecular weights (31,000 and 2700) and purity of these fragments were assessed using SDS-gel electrophoresis (21) and SDS-urea gel electrophoresis (Bethesda Research Laboratories, Focus, 2 (6) 6, 1980). The CNBr fragments of cytochrome *c* (7760, 2780, and 1810 Da) and myoglobin (8270, 6420, and 2550 Da) (21) served as molecular weight standards for the SDS-urea polyacrylamide gel.

An amino acid analysis of the small fragment was performed and the most probable composition was calculated using a computer program described by Hoy *et al.* (22). A molecular weight of 3300 was calculated.

Periodate oxidation and sodium [³H] borohydride reduction of A and B subunits. A modified version of the method described by Lotan *et al.* (23) was employed to label the carbohydrate moiety of both A and B subunits.

Reduced and alkylated A₄ or B₄ isolectins (1.0 mg) were dissolved in 250 μl of 60 mM sodium acetate buffer, pH 4.5, containing 0.05% SDS. Sodium metaperiodate (250 μl of 60 mM) was added and the solution was stirred for 4 h in the dark (0–4°C). The oxidation reaction was terminated with 50 μl of 3 M ethylene glycol (1 h, 0–4°C). The solution was dialyzed against 10 mM NH₄HCO₃, pH 7.9. A 0.5 M solution of sodium [³H] borohydride was prepared in 0.1 N NaOH; 30 μl of this solution was added and the reaction was continued for 40 min at 50°C. The reduction reaction was terminated by adding 0.75 M H₂SO₄ until the pH

dropped to 5.0. The solution was dialyzed against 0.1 M NH_4HCO_3 , pH 7.9. The labeled isolectins were diluted fivefold with unlabeled A_4 or B_4 which had been oxidized with periodate and reduced with unlabeled sodium borohydride. The specific activities of A_4 and B_4 were 120×10^6 and 146×10^6 cpm mol^{-1} , respectively.

Trypsin digestion. Reduced and alkylated A_4 , A_2B_2 , and B_4 isolectins were dissolved separately at a concentration of 1 mg ml^{-1} in 0.1 M NH_4HCO_3 buffer, pH 7.9, and digested with TPCK-treated trypsin (1 mg ml^{-1} in 1 mM HCl) added in four equal aliquots (lectin:trypsin ratio 1:40 w/w) over 10–20 h. The solution was stirred throughout the reaction. One hour after the last addition of trypsin, the reaction was quenched (adjusted to pH 3.0 with acetic acid) and quickly lyophilized, and the tryptic peptides were stored at 4°C. Those isolectins which had been labeled in the carbohydrate moiety with sodium [^3H] borohydride and those labeled with [^{14}C]iodoacetamide were also digested with trypsin as described above. The tryptic peptides were separated using reverse-phase HPLC.

Separation of tryptic peptides by reverse-phase HPLC. A Waters Associates (Milford, Mass.) HPLC system was employed. The system was comprised of Model 6000A pumps, a Model 600 solvent programmer, a Model 450 variable-wavelength detector, and a Model U6K injector. All separations were performed on a Waters C-18 $\mu\text{Bondapak}$ column (4 mm \times 30 cm, 10 μm). Tryptic peptides were separated according to the method of Fulmer and Wasserman (24). A flow rate of 1.5 ml min^{-1} was maintained. Tryptic peptides were collected either manually or using a fraction collector.

Protein and neutral sugar assay. Protein was determined by the method of Lowry *et al.* (25) using bovine serum albumin as the standard. The lectin concentrations were also calculated from the absorbance at 280 nm ($E_{1\text{cm}}^{0.1\%} = 1.41$). Covalently bound carbohydrate was detected by SDS-gel electrophoresis according to the method of Racusen (26).

RESULTS

Isolectin purification. *G. simplicifolia* isolectins (A_4 , A_3B , A_2B_2 , AB_3 , B_4) were purified according to the method of Delmotte and Goldstein (8). The purity of the individual isolectins was assessed using pH 8.9 nondenaturing polyacrylamide gels and SDS-gel electrophoresis in the presence of 2-mercaptoethanol. The relative proportions of A and B subunits in each isolectin and the purity of the preparations are illustrated in Fig. 1. The molecular weights of the A and B subunits were calculated to be 36,000 and 38,500, respectively. These molecular weights are somewhat higher

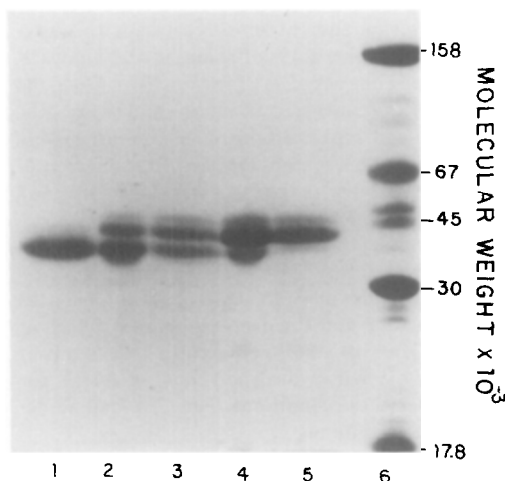


FIG. 1. Demonstration of the separation of GS-I isolectins and relative purity of A and B subunits. The five isolectins were purified by affinity chromatography (8) and subjected to SDS-gel electrophoresis on a 12.5% polyacrylamide slab gel in the presence of 2-mercaptoethanol. Lanes: 1, A_4 ; 2, A_3B ; 3, A_2B_2 ; 4, AB_3 ; 5, B_4 ; 6, molecular weight markers aldolase, BSA, ovalbumin, carbonic anhydrase, and myoglobin.

than those previously reported (4). Only those preparations of A_4 and B_4 which were at least 95% homogeneous were employed in the structural studies of the A and B subunits.

Pyroglutamate aminopeptidase digestion and N-terminal sequence analysis. When TRH and reduced and alkylated A_2B_2 were digested with PAP, glutamic acid was released in both cases. There was no ambiguity in the identity of the amino acid released, since after subtraction of the control digest (PAP and unblocking buffer) glutamic acid was the only residue which remained. When A_4 and B_4 (reduced and alkylated) were treated with PAP, glutamic acid was again released from both isolectins. These data support the view that both A and B subunits are blocked with N-terminal pyroglutamate which prevents N-terminal amino acid sequencing. A 30% yield of N-terminal pyroglutamate was obtained routinely from the A and B subunits. The TRH yielded 22% pyroglutamate; hence the low yields are probably attributable to experimental manipulations and do not reflect a partial blocking of the subunits. Confirmation of the success of PAP

digestion was obtained by automated Edman degradation. The N-terminal sequence of both A and B subunits are presented in Table I. Although 21 of the amino acids are similar there are four unique amino acids. Residues 8, 9, 17, and 19 are asparagine, leucine, lysine, and asparagine in subunit A and threonine, phenylalanine, glutamic acid, and serine in subunit B.

N-terminal sequence homology. Table II illustrates a comparison among the N-terminal sequences of the A and B subunits and those of other D-galactosyl and N-acetyl- α -D-galactosaminyl binding lectins. Also included is the sequence of con A beginning at residue 123. All lectin subunits examined so far possess the invariant phenylalanines at positions 6 and 11. The sequence, beginning at residue 22, is conserved though not invariant. With minor

shifts in the position of a few amino acids (possibly due to insertions, deletions, or substitutions in an ancestral gene) the lectins show considerable homology. However, the N-terminal blocking group pyroglutamate has not been demonstrated to be present in any other lectin subunit which has been examined thus far.

Carboxypeptidase Y digestion—C-terminal sequence. The results of the CPY digestion of reduced and alkylated A₄ and B₄ isolectins are shown in Figs. 2 and 3, respectively. The kinetics of release of each C-terminal amino acid are the same for both A and B subunits. A concentration of either A₄ or B₄ that would yield 10–15 nmol of A or B subunits, in the samples withdrawn at timed intervals, was employed in the digestion. A typical digestion yielded 2–10 nmol of C-terminal amino acids (depending on the location of the amino acid in the sequence). The yield of each amino acid was corrected for loss of the corresponding amino acid when a standard mixture of amino acids was digested with CPY. The C-terminal sequence was Arg-(Val, Phe)-Leu-Thr-Ser-COOH for both A and B subunits.

Cyanogen bromide cleavage reaction. Previous studies (4) have shown that B subunits contain one methionyl residue whereas A subunits contain none. When both B₄ and A₄ were reacted separately with cyanogen bromide only B subunits were cleaved. The B subunits were cleaved into a large-molecular weight-fragment (CNBr B₄-I) and a small-molecular-weight fragment (CNBr B₄-II). The size of these fragments was determined using two types of polyacrylamide electrophoresis. The large fragment (CNBr B₄-I, Fig. 4, lane 3) was analyzed by SDS-gel electrophoresis in the presence of 2-mercaptoethanol using bovine serum albumin, ovalbumin, carbonic anhydrase, and myoglobin as molecular weight standards and was calculated to have a molecular weight of approximately 31,000. A sample of CNBr-treated A₄ is illustrated in lane 5. No cleavage reaction occurred which was consistent with the view that A subunits contain no methionine (compare lanes 4, untreated and 5, treated). Figure 5 illustrates the re-

TABLE I
COMPARISONS BETWEEN AMINO TERMINAL SEQUENCE
OF GS I -A₄ AND -B₄^a

| Residue No. | A ₄ | B ₄ |
|-------------|----------------|----------------|
| 1 | PyroGlu | PyroGlu |
| 2 | Ser | Ser |
| 3 | Asp | Asp |
| 4 | Ser | Ser |
| 5 | Val | Val |
| 6 | Ser | Ser |
| 7 | Phe | Phe |
| 8 | <i>Asn</i> | <i>Thr</i> |
| 9 | <i>Leu</i> | <i>Phe</i> |
| 10 | Pro | Pro |
| 11 | Asn | Asn |
| 12 | Phe | Phe |
| 13 | Trp | Trp |
| 14 | Ser | Ser |
| 15 | Asp | Asp |
| 16 | Val | Val |
| 17 | <i>Lys</i> | <i>Glu</i> |
| 18 | Asp | Asp |
| 19 | <i>Asn</i> | <i>Ser</i> |
| 20 | Ile | Ile |
| 21 | Ile | Ile |
| 22 | Phe | Phe |
| 23 | Gln | Gln |
| 24 | Gly | Gly |
| 25 | Asp | Asp |
| 26 | Ala | Ala |

^a Residues which differ are italicized.

TABLE II

COMPARISON AMONG N-TERMINAL SEQUENCES OF Gal*- AND GalNAc-BINDING LECTINS

| | | 1 | 5 | 10 | 15 | 20 | 25 |
|--|----------------------|-----------------------|---------------------|---------------------|----------------|----|----|
| <i>G. simplicifolia</i> I-A | PyroGln ⁶ | S D S | V S F N L P | N F W S D V K D N | I I F Q G D A | | |
| <i>G. simplicifolia</i> I-B | PyroGln | S D S | V S F T F | P N F W S D V E D S | I I F Q G D A* | | |
| <i>Arachis hypogaea</i> ¹ | | A E T V S F N F | N S F S E G N P A | I N F O G D V* | | | |
| <i>Dolichos biflorus</i> ² | | A N T Q S F S F | Q N F N S P S F | I L O G D A | | | |
| <i>Glycine max</i> ¹ | | A E T V S F S W N K F | V P K E P D M I L E | G D A | | | |
| <i>Phaseolus vulgaris</i> ³ E | | A S E T S F S F | E R F N E T | N L I L O R D A | | | |
| L | | S N D I Y F N F | E R F N E T | N L I L O R D A | | | |
| <i>Vicia cracca</i> GalNAc ⁴ | | T E S T S F S F | T E F N Q N Q E | N L I L E G D A | | | |
| con A ⁵ | 123 | T D A L H F M F | N Q F S K D Q K D | L I L Q G D A | | | |

¹ Foriers, A., Wuilmart, C., Sharon, N., and Strossberg, A. D. (1977) *Biochem. Biophys. Res. Commun.* **75**, 980-986.

² Etzler, M. E., Talbot, C. F., and Ziaya, P. R. (1977) *FEBS Lett.* **82**, 39-41.

³ Miller, J. B., Noyes, C., Heinrikson, R., Kingdon, H. S., and Yachnin, S. (1973) *J. Exp. Med.* **138**, 939-951.

⁴ Bauman, C., Rüdiger, H., and Strossberg, A. D. (1979) *FEBS Lett.* **102**, 216-218.

⁵ Cunningham, B. A., Wang, J. L., Waxdal, M. J., and Edelman, G. M. (1975) *J. Biol. Chem.* **250**, 1503-1512.

⁶ Pyroglutamate.

sults of SDS-urea gel electrophoresis of the small fragment. Cyanogen bromide cleaved sperm whale myoglobin (lane 1)

and horse heart cytochrome *c* (lane 2) were used as molecular weight standards. The CNBr B₄-II was calculated to have a molecular weight of 2700. Similar gels were treated with the thymol-sulfuric acid reagent (26) to visualize the carbohydrate covalently bound to the cyanogen bromide fragments. Only CNBr B₄-I gave a positive

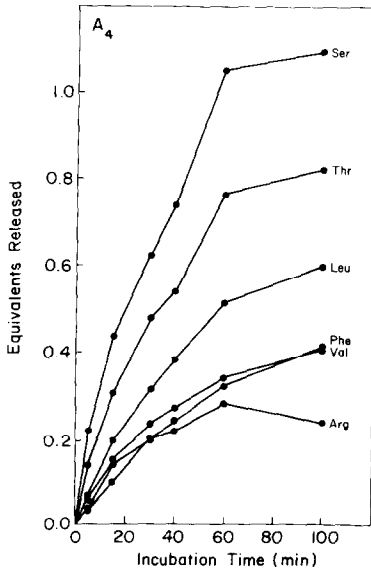


FIG. 2. Kinetics of release of C-terminal amino acids from reduced-alkylated GS-I A₄. Graph shows the number of equivalents (nmol amino acid nmol subunit⁻¹) released versus time of incubation with carboxypeptidase Y.

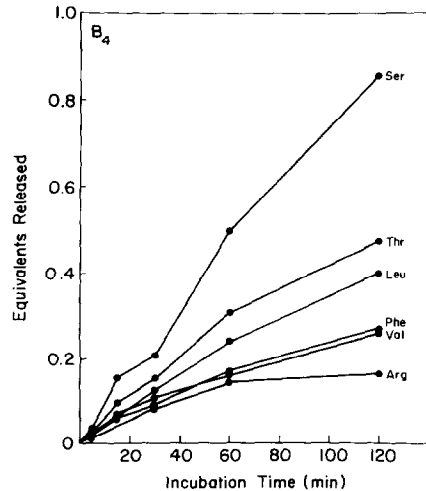


FIG. 3. Kinetics of release of C-terminal amino acids from reduced-alkylated GS-I B₄. See Fig. 2 for details.

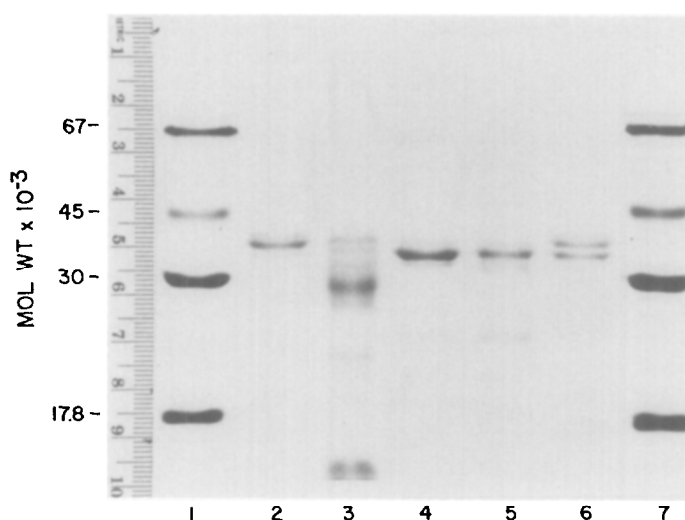


FIG. 4. Analysis of the large CNBr fragment (CNBr B₄-I) from B subunits. Purified A₄ and B₄ isolectins were reacted with CNBr and subjected to SDS-gel electrophoresis in 12.5% polyacrylamide slab gels in the presence of 2-mercaptoethanol. Lanes: 1 and 7, molecular weight markers BSA, ovalbumin, carbonic anhydrase, and myoglobin; 2, GS-I B₄; 3, CNBr-cleaved B₄; 4, GS-I A₄; 5, CNBr-treated A₄; 6, GS-I A₂B₂.

reaction. The small fragment (CNBr B₄-II) does not appear to be a glycopeptide. The large and small CNBr fragments of the B subunit were separated on a Sephadex LH 60 (114 × 1.5 cm) column using ethanol-containing formic acid as the elution solvent (Fig. 6). The sizes of the respective peaks do not indicate the relative proportions of the two fragments since the extinction coefficients were different (CNBr B₄-I has a greater number of tyrosyl and tryptophanyl residues than CNBr B₄-II). The purity of the separated fragments was assessed using either SDS-gel electrophoresis or SDS-urea gel electrophoresis. The small fragment, which appeared to be homogenous, was hydrolyzed in 6 N HCl and the amino acid composition was determined. A computer analysis of the amino acid composition was performed (22), the results of which are shown in Table III. A molecular weight of 3300 was calculated.

The small and large CNBr fragments were subjected to manual Edman degradation. No N-terminal amino acids were released from the small fragment despite attempts to unblock the peptide by treatment with methanolic HCl (50°C, 30 min). The first three N-terminal amino acids of myo-

globin were determined in parallel as a control. The first two amino acids of the large fragment were determined (data not shown). The fragment (CNBr B₄-I) was clearly not blocked with pyroglutamate and is probably at the C-terminal end of the B subunit. These data suggest that CNBr B₄-II is located at the N-terminus.

Reverse-phase HPLC analysis of tryptic peptides. Figure 7 illustrates the tryptic peptide maps obtained following digestion of reduced-alkylated A₄, B₄, and A₂B₂. In all three column profiles the gradient was 0-50% CH₃CN in 0.1% H₃PO₄. The tryptic peptides were monitored by the absorbance of the peptide bond at 220 nm. Approximately seven peptides were common to both subunits, eight were unique to B, and approximately 10 were unique to A. The cysteine-containing peptides (A and B each contain one cysteine) were detected using a radiolabeled ([1-¹⁴C]iodoacetamide) alkylating agent. Single peaks of radioactivity, coeluting with tryptic peptides, were detected for both A and B subunits. The retention times of the A and B cysteine-containing tryptic peptides were different, 36.5 and 50.5 min, respectively. Similarly, the tryptic glycopeptides were detected by

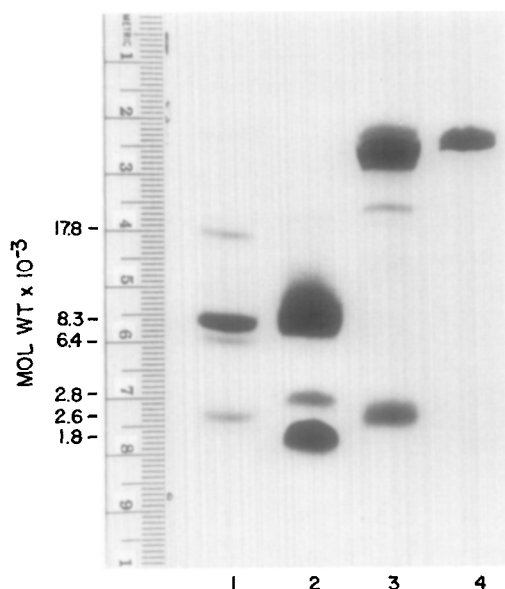


FIG. 5. Analysis of the small CNBr fragment (CNBr B₄-II) from B subunits. Purified B₄, horse heart cytochrome c, and sperm whale myoglobin were reacted with CNBr and the resulting fragments subjected to SDS-urea gel electrophoresis in 15% polyacrylamide gels in the presence of 2-mercaptoethanol. Lanes: 1, myoglobin; 2, cytochrome c; 3, GS-I B₄; 4, GS-I A₂B₂ (A and B subunits are not resolved in this gel).

treating the carbohydrate moiety of either A or B subunit with periodate followed by sodium [³H] borohydride. A single peak of radioactivity coeluted with a tryptic peptide. Again, however, the retention times of the A and B tryptic glycopeptides were different, 37.5 and 56.5 min, respectively. The tryptic glycopeptide and the cysteine-containing peptide of the A subunit eluted with virtually the same retention time (37.5 and 36.5 min) but appeared to be different peptides. These peptides were more widely separated in the B subunit tryptic peptide map (56.5 and 50.5 min, respectively) and were clearly different peptides.

These data support the notion that A and B subunits share some regions of homology but also have unique sequences within the molecule.

DISCUSSION

The goal of this study was to make a structural comparison of the A and B sub-

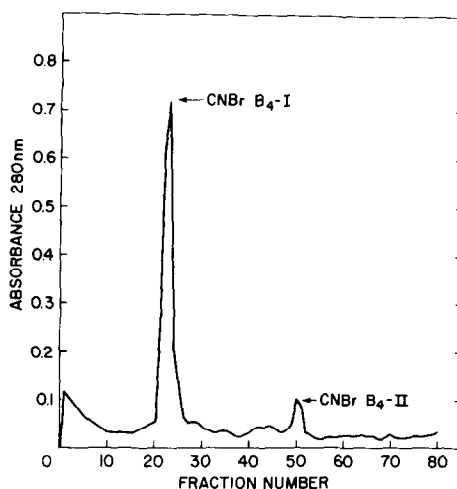


FIG. 6. Separation of the large and small CNBr fragments of B subunits. Purified B₄ was reacted with CNBr and the fragments (dissolved in formic acid) were applied to a Sephadex LH-60 column (114 × 1.5 cm) equilibrated with formic acid (88%):ethanol (100%) 1:4 v/v and eluted with the same solvent.

units which comprise the five tetrameric α -D-galactosyl binding isolectins purified from the seeds of *G. simplicifolia*. In a previous report (7) we postulated that the B subunit was converted, via some form of post-translational proteolytic modification at the C-terminus, into the A subunit. This conversion of B to A was also assumed to include the removal of a methionine, since

TABLE III

AMINO ACID COMPOSITION OF CNBr B₄-II

| Amino acid | Number |
|------------|--------|
| Asx | 4 |
| Thr | 3 |
| Ser | 4 |
| Glx | 3 |
| Pro | 1 |
| Gly | 1 |
| Ala | 2 |
| Val | 3 |
| Ile | 1 |
| Leu | 3 |
| Phe | 2 |
| Lys | 1 |
| Arg | 2 |

Note. Molecular weight = 3300.

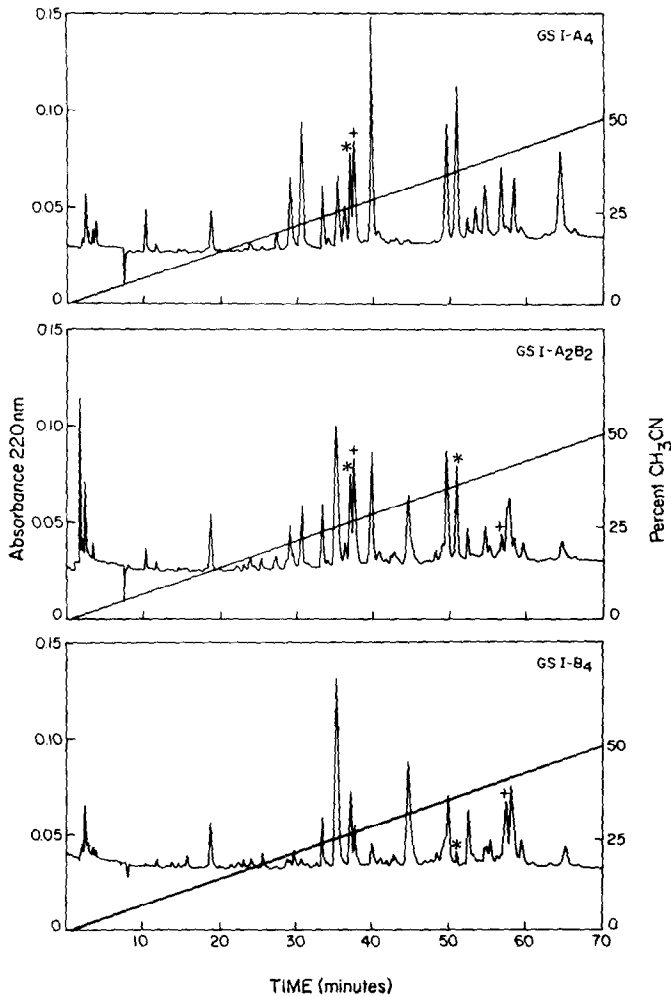


FIG. 7. Comparison among GS-I A₄, A₂B₂, and B₄ tryptic peptides using reverse-phase HPLC. Purified reduced-alkylated A₄, A₂B₂, and B₄ isolectins were digested with trypsin and the peptides separated on a C₁₈ μ Bondapak column using a linear (0-50%) gradient of CH₃CN in 0.1% H₃PO₄. The retention time of tryptic peptides labeled with [1-¹⁴C]iodoacetamide (*) and sodium [³H] borohydride (+) are indicated. The retention time, in minutes, is shown at the bottom for all three peptide maps.

A subunits contain no methionine. To test this hypothesis, we have identified the first 25 N-terminal amino acids and the last six C-terminal amino acids of A and B subunits. In order to identify the position of methionine we analyzed cyanogen bromide cleavage fragments of the B subunit. The A and B subunits were also compared using tryptic peptide mapping.

An N-terminal sequence analysis of the A and B subunits was initially hampered

by the presence of an unknown blocking group. We discovered, however, that calf liver pyroglutamate aminopeptidase could remove the blocking group from both the A and B subunits. The glutamic acid which was released was identified by amino acid analysis. The unblocked subunits were sequenced by automated Edman degradation which confirmed the success of the enzymatic digestion and, more importantly, provided the N-terminal sequence of both

A and B subunits. The discovery of a pyroglutamate group, blocking the N-terminus of the subunits, is noteworthy because there is only one other example of a plant protein containing N-terminal pyroglutamate: One of the acidic chains of pumpkin seed globulin has been reported to have an N-terminal pyroglutamate blocking group (27). It is not known, however, whether the pyroglutamate of A and B subunits is formed enzymatically during the synthesis of the isolectins or whether it is formed by spontaneous cyclization of N-terminal glutamic acid.

Although 21 of the first 25 N-terminal amino acids were the same in both subunits, there were four unique amino acids in each sequence. Three of these four amino acids (8, 9, and 17) could have arisen by point mutations. The change from asparagine at residue 19 in A to serine in B would have required a two-step mutation.

Comparing amino acid sequences of lectin subunits is a useful way to search for structure-function and evolutionary relationships. Unfortunately, few lectins, notably con A (28) and favin (29), have been completely sequenced. Foriers *et al.* (30) have approached the problem by comparing the available, published N-terminal sequences. They found significant homologies among legume lectin subunits. When the N-terminal sequences of the A and B subunits are added to this list the homology among the subunits of the legume lectins is striking.

Cunningham *et al.* (31) discovered that the lectin from *Vicia faba* and con A from Jack bean were circular permutations of a single sequence. Both lectins exhibited extensive homology. Amino acids, believed to play a role in the active site and in metal binding, were conserved. The A and B subunits also show regions of homology with con A, beginning at residue 123 of this lectin (see Table II). Conceivably, there is a functional basis for the homology among the lectins. The functional relationships will, hopefully, become clearer as more lectins are sequenced.

C-terminal sequences of only a few other lectins have been reported (32, 33) and in only one other study was carboxypeptidase

Y used (32). Carboxypeptidase Y is useful because it will remove all amino acids, including proline. Noteworthy, however, is the fact that the last three C-terminal amino acids of the subunits of the lectin from red kidney beans (*Phaseolus vulgaris*) are identical (Ser-Ile-Leu-COOH). The lectin from red kidney beans also consists of five tetrameric isolectins containing two (R and L) subunits (33) and is thus structurally similar to the *G. simplicifolia* isolectins. Both subunits from the two lectins (*P. vulgaris* and *G. simplicifolia*) have identical C-termini and similar, but not identical, N-termini. The lectin from the seeds of *Dolichos biflorus* is composed of equal amounts of two structurally similar subunits I and II. These subunits have different C-terminal sequences, however (32). It will be interesting to compare the C-terminal sequences of the subunits of legume lectins in the same manner as the N-terminal sequences have been compared.

To determine the location of methionine in the B subunit, a cyanogen bromide cleavage was performed, following which the two fragments were separated using hydrophobic chromatography (20). The two fragments were calculated to be approximately 31,000 and 2700 Da by SDS-gel electrophoresis and SDS-urea gel electrophoresis, respectively. An amino acid analysis of the small fragment was performed and a molecular weight of 3300 was calculated. This discrepancy in molecular weights is probably not significant and reflects inherent differences in the two experimental methods. To investigate the location of the small fragment an N-terminal sequence analysis was attempted. The fragment (CNBr B₄-II) appeared to be blocked, presumably with pyroglutamate. These data are consistent with the small fragment being located at the N-terminus of the B subunit. Unfortunately, the insolubility of the small fragment precluded a C-terminal sequence analysis with carboxypeptidase Y or the enzymatic removal of the pyroglutamate following cyanogen bromide cleavage. Only the first few N-terminal amino acids were obtained from the large fragment. These proved difficult to identify reliably. Notwithstanding these

problems, the data suggested that the large fragment was not blocked with pyroglutamate. Future studies will be directed towards a complete sequence analysis of both cyanogen bromide fragments. Such data should also lead to an accurate measurement of the respective molecular weights. At present, however, it seems likely that the sole methionine of the B subunit resides in the N-terminal region of the molecule.

The final part of the structural analysis involved tryptic peptide mapping of A and B subunits. The use of HPLC greatly facilitated this analysis. A comparison of the tryptic peptide maps revealed that approximately 30% of the peptides were common to both subunits. Neither tryptic glycopeptides nor cysteine containing tryptic peptides were the same for the A and B subunits (see Fig. 7). Apart from the present research only one other study has reported using HPLC to separate tryptic peptides of lectin subunits (34). This technique should facilitate comparison of structural data for other lectin subunits.

In a similar manner, Miller *et al.* (33) performed tryptic peptide mapping of R and L subunits from red kidney bean isolectins. These subunits were not identical although they did share regions of homology.

At the onset of this study we sought to test the hypothesis that A subunits are derived from B subunits via some form of proteolytic processing event at the C-terminus. The A and B subunits were indeed found to be similar, especially at the N- and C-termini. The tryptic peptide maps indicated, however, that there were unique sequences within the two subunits. These data did not support the view that A subunits are derived directly from B subunits. Our alternate hypothesis (7) stated that A and B subunits were derived from a common ancestral gene which had undergone duplication and subsequent mutation. A similar proposal has been advanced by several researchers to explain the similarity among lectin subunits (35-38). The data presented here are, therefore, consistent with the latter hypothesis. Conclusive proof, however, must await a complete sequence analysis of the A and B subunits.

ACKNOWLEDGMENTS

We thank Eleanor Hart and Jana Lawton for their assistance in preparing this manuscript. We are grateful to Dr. Shaun Black for his generous assistance in the structural analyses. Dr. George Tarr is gratefully acknowledged for performing the automated Edman degradation. The technical assistance of Leslie Nuysten is appreciated.

REFERENCES

1. GOLDSTEIN, I. J. (1981) *in* Recent Advances in Phytochemistry (Loewus, F. A., and Ryan, C. A., eds.) Vol. 15, pp. 25-35, Plenum, New York.
2. MÄKELÄ, O., AND MÄKELÄ, P. (1956) *Ann. Med. Exp. Biol. Fenn.* **34**, 402-404.
3. MURPHY, L. A., AND GOLDSTEIN, I. J. (1977) *J. Biol. Chem.* **262**, 4739-4742.
4. MURPHY, L. A., AND GOLDSTEIN, I. J. (1979) *Biochemistry* **18**, 4999-5005.
5. GOLDSTEIN, I. J., BLAKE, D. E., EBISU, S., WILLIAMS, T. J., AND MURPHY, L. A. (1981) *J. Biol. Chem.* **256**, 2890-3893.
6. WOOD, C., KABAT, E. A., MURPHY, L. A., AND GOLDSTEIN, I. J. (1979) *Arch. Biochem. Biophys.* **198**, 1-11.
7. LAMB, J. E., BOOKSTEIN, F. L., GOLDSTEIN, I. J., AND NEWTON, L. E. (1981) *J. Biol. Chem.* **256**, 5874-5878.
8. DELMOTTE, F. M., AND GOLDSTEIN, I. J. (1980) *Eur. J. Biochem.* **112**, 219-223.
9. BREWER, J. M., AND ASHWORTH, R. B. (1969) *J. Chem. Educ.* **46**, 41-46.
10. LAEMMLI, U. K. (1970) *Nature (London)* **227**, 680-685.
11. CRESTFIELD, A. M., MOORE, S., AND STEIN, W. (1963) *J. Biol. Chem.* **238**, 622-627.
12. STARK, G. R., AND STEIN, W. M. (1964) *J. Biol. Chem.* **239**, 3755-3761.
13. BLACK, S. D., FRENCH, J. S., WILLIAMS, C. H., JR., AND COON, M. J. (1979) *Biochem. Biophys. Res. Commun.* **91**, 1528-1535.
14. EDMAN, P., AND BEGG, G. (1967) *Eur. J. Biochem.* **1**, 80-91.
15. BRAUER, A. W., MARGOLIES, M. N., AND HABER, E. (1975) *Biochemistry* **14**, 3029-3035.
16. TARR, G. E. (1981) *Anal. Biochem.* **111**, 27-32.
17. TARR, G. E. (1975) *Anal. Biochem.* **63**, 361-370.
18. TARR, G. E., (1977) *in* Methods in Enzymology (Hirs, C. H. W., and Timasheff, S. N., eds.), Vol. 47, pp. 335-357, Academic Press, New York.
19. SCHROEDER, W. A., SHELTON, J. B., AND SHELTON, J. R. (1969) *Arch. Biochem. Biophys.* **130**, 551-556.
20. TAKAGAKI, Y., GERBER, G. E., NIHEI, K., AND KHORANA, H. G. (1980) *J. Biol. Chem.* **255**, 1536-1541.

21. SWANK, R. T., AND MUNKRES, K. D. (1971) *Anal. Biochem.* **39**, 462-477.
22. HOY, T. G., FERDINAND, T. G., AND HARRISON, P. M. (1974) *Int. J. Peptide Protein Res.* **6**, 121-140.
23. LOTAN, R., DEBRAY, H., CACAN, M., CACAN, R., AND SHARON, N. (1975) *J. Biol. Chem.* **250**, 1955-1957.
24. FULLMER, C. S., AND WASSERMAN, R. H. (1979) *J. Biol. Chem.* **254**, 7208-7212.
25. LOWRY, O., ROSEBROUGH, N. J., FARR, A. L., AND RANDALL, R. J. (1951) *J. Biol. Chem.* **193**, 265-275.
26. RACUSEN, D. (1979) *Anal. Biochem.* **99**, 474-476.
27. OHMIYA, M., HARA, I., AND MATSUBARA, H. (1980) *Plant Cell Physiol.* **21**, 157-167.
28. CUNNINGHAM, B. A., HEMPERLY, J. J., HOPP, T. P., AND EDELMAN, G. M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3218-3222.
29. EDELMAN, G. M., CUNNINGHAM, B. A., REEKE, G. N., JR., BECKER, J. W., WAXDAL, M. J., AND WANG, J. L. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 2580-2584.
30. FORIERS, A., DE NEVE, R., AND STROBERG, A. D. (1979) *Physiol. Vég.* **17**, 597-606.
31. CUNNINGHAM, B. A., HEMPERLY, J. J., HOPP, T. P., AND EDELMAN, G. M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3218-3222.
32. ROBERTS, D. M., WALKER, J., AND ETZLER, M. E. (1982) *Arch. Biochem. Biophys.* **218**, 213-219.
33. MILLER, J. B., HSU, R., HEINRIKSON, R., AND YACHNIN, S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1388-1391.
34. ROBERTS, D. D., AND GOLDSTEIN, I. J. (1982) *J. Biol. Chem.* **257**, 11274-11277.
35. BAUMAN, C., RUDIGER, H., AND STROBERG, A. D. (1979) *FEBS Lett.* **102**, 216-218.
36. HOWARD, I. K., SAGE, H. J., STEIN, M. D., YOUNG, N. M., LEON, M. S., AND DYCKES, D. F. (1971) *J. Biol. Chem.* **246**, 1590-1595.
37. LIS, H., FRIDMAN, C., SHARON, N., AND KATCHALSKI, E. (1966) *Arch. Biochem. Biophys.* **117**, 301-309.
38. VRETBLAD, P., HJORTH, R., AND LAAS, T. (1979) *Biochim Biophys. Acta* **579**, 52-61.