Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 6345–6348, September 1987
Biochemistry

Extensive amino acid sequence homologies between animal lectins

P. PAROUTAUD*, G. LEVI†, V. I. TEICHBERG†, AND A. D. STROSBERG*

*Department of Biotechnology, Pasteur Institute, Centre National de la Recherche Scientifique, University Paris VII, 25, rue du Docteur Roux, 75724 Paris Cedex 15, France; and †Department of Neurobiology, The Weizmann Institute for Science, Rehovot 76100, Israel

Communicated by Elvin A. Kabat, May 7, 1987

ABSTRACT **We have established the amino acid sequence of the β-D-galactoside binding lectin from the electric eel and the sequences of several peptides from a similar lectin isolated from human placenta. These sequences were compared with the published sequences of peptides derived from the β-D-galactoside binding lectin from human lung and with sequences deduced from cDNAs assigned to the β-D-galactoside binding lectins from chicken embryo skin and human hepatomas. Significant homologies were observed. One of the highly conserved regions that contains a tryptophan residue and two glutamic acid residues is probably part of the β-D-galactoside binding site, which, on the basis of spectroscopic studies of the electric eel lectin, is expected to contain such residues. The similarity of the hydropathy profiles and the predicted secondary structure of the lectins from chicken skin and electric eel, in spite of differences in their amino acid sequences, strongly suggests that these proteins have maintained structural homologies during evolution and together with the other β-D-galactoside binding lectins were derived from a common ancestor gene.**

Embryonic and differentiating tissues of various vertebrates from teleosts (1, 2) to amphibians (3), birds (4, 5), and mammals (6–9) contain a β-D-galactoside binding protein (lectin) that agglutinates trypsinized rabbit erythrocytes and is specifically inhibited by β-D-galactopyranosyl β-D-thiogalactopyranoside and lactose. Several of these lectins have been isolated and characterized and found to share common properties such as subunit molecular weight, saccharide specificity, and the requirement for the presence of reducing agents during the purification procedure (1, 10). In some cases, they possess similar amino acid compositions (4, 10). The fact that several β-D-galactoside binding lectins from phylogenetically distinct species cross react immunologically further indicates that these proteins possess common structural determinants that have been maintained during evolution (4, 12–14).

To determine the extent of the structural homology between various β-D-galactoside binding lectins, we have undertaken to establish the amino acid sequences of β-D-galactoside binding lectins from the electric eel (15) and human placenta (9). We compare here the established amino acid sequences of these proteins to the sequences deduced from the cDNAs encoding the β-D-galactoside binding lectins from embryonic chicken skin (16) and from human hepatomas (17) and to the partial amino acid sequence of the human lung lectin (17).

The homologies observed suggest that these proteins are derived from a common ancestor gene and have maintained through evolution part of the structure of the β-D-galactoside binding site putatively assigned to residues 70–76.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Tissues. Live electric eels, *Electrophorus electricus*, were obtained from Worldwide Paramount Aquarium (Ardsley, NY). They were decapitated, and the main electric organ was cut in small cubes and frozen at −20°C. Fresh human placentas obtained during night deliveries were kept on ice until processing the next morning.

Purification of β-D-Galactoside Binding Lectins. The tissue homogenization and fractionation were carried out as described by Levi and Teichberg (15). The lectins were isolated by affinity chromatography using a lactosyl-Sepharose matrix (15). The eluting buffer consisted of 0.01 M phosphate-buffered saline (0.15 M NaCl, pH 7.2) supplemented with 100 mM lactose and 14 mM 2-mercaptoethanol. The purity of the eluted lectin was verified by $NaDodSO_4$/PAGE, and its activity after dialysis against phosphate-buffered saline was monitored on trypsinized rabbit erythrocytes in a quantitative hemagglutination assay performed on microtiter plates as reported (15).

Digestions. *Tryptic digestion.* The digestion was performed in 1% NH_4HCO_3 by adding 10 μg of tosylphenylalanyl chloromethyl ketone-treated trypsin dissolved in 0.01 mM HCl to 1 mg of lectin. After 6 hr at 37°C, an additional 10 μg of trypsin was added, and 6 hr later the digestion was stopped by lowering the pH to 4. The digest was separated by HPLC on a RP300 column (Brownlee) using the following solvents: solvent A = H_2O/0.1% trifluoroacetic acid; solvent B = 80% (vol/vol) CH_3CN/20% (vol/vol) H_2O/0.1% trifluoroacetic acid.

Both solvents were degassed by a helium stream. The separation was completed using a gradient from 0% solvent B to 40% (vol/vol) solvent B in 8 hr with a flow rate of 0.5 ml/min. The variable wavelength detector was set at 206 nm, with a 0.2 sensitivity. The operation temperature was 25°C.

Staphylococcus aureus *protease digestion.* The digestion (18) was performed under conditions designed for cleavage at the Glu-Xaa peptide bonds as follows: 10 μg of enzyme (Miles Laboratories) was added to 1 mg of lectin in 50 mM NH_4HCO_3 at pH 7.8, and the digestion was allowed to proceed for 18 hr at 37°C. The resulting peptides were separated by HPLC as described above.

Cyanogen bromide cleavage. The cleavage was performed in 70% (vol/vol) formic acid, at 25°C for 24 hr in the dark, using 10 mg of cyanogen bromide for 1 mg of electrolectin.

Resulting fragments were separated by gel filtration on Sephadex G-50 superfine, in the presence of 5 M guanidine hydrochloride. Fragments were desalted on Sephadex G-25 in 1% NH_4HCO_3 and lyophilized.

Separation and Purification of Peptides. Tryptic and *S. aureus* protease peptides were separated by reverse-phase chromatography on a Waters HPLC instrument and monitored at 206 nm with a model 440 spectrophotometer (Waters). Most separations were done on a RP300 column (supplied by Brownlee) in 0.1% trifluoroacetic acid buffers using gradients described above.

Sequence Analysis. The various peptides were analyzed by automated Edman degradation in a gas-phase microsequencer from Applied Biosystems. Chemubrene (Chemunex, Paris) was added to prevent wash out of peptides and for

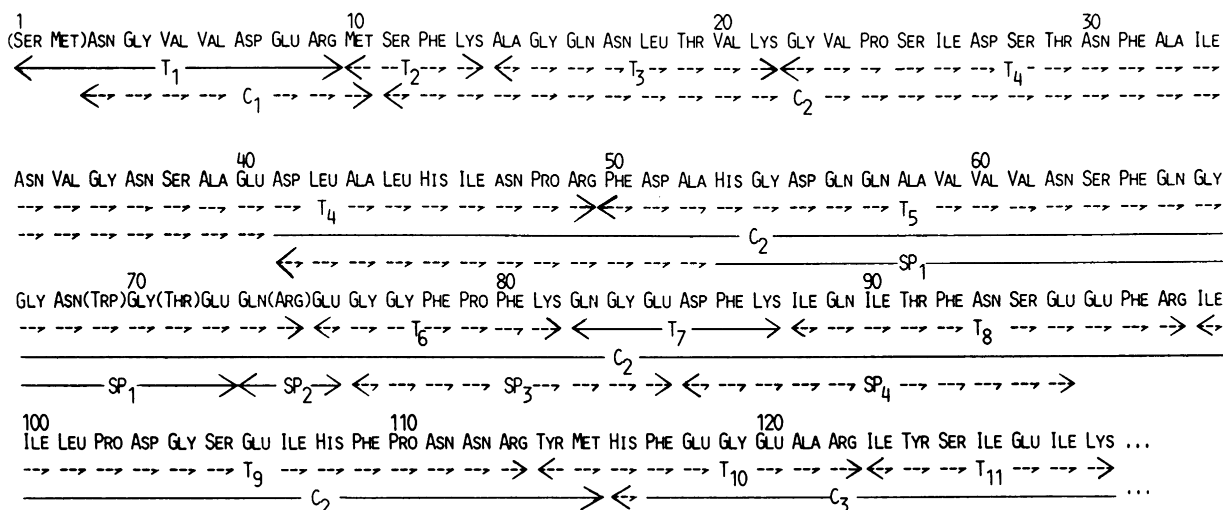

FIG. 1. Alignment of the peptides and fragments from eel lectin, based on the sequence derived from a cDNA corresponding to chicken lectin. T, tryptic peptide; C, cyanogen bromide peptide; SP, *S. aureus* peptide. Solid line, residue positioned by composition. Broken line, residues effectively sequenced.

improved initial yields. The phenylthiohydantoin amino acid derivatives were identified by HPLC on a Waters instrument equipped with a RP18 5-μm column (supplied by Brownlee), using a 0–40% sodium acetate/acetonitrile gradient (as above).

RESULTS

Amino Terminal Sequencing. Several attempts were made to sequence directly the amino terminus of electrolectin, the β-D-galactoside binding protein from *E. electricus*, and of the human placental lectin. Both manual and automated procedures failed, even after pretreatment of the protein with pyrrolidone decarboxylase, generally useful for removing cyclized glutamic acid or glutamine residues. From these negative results, it was assumed that the amino terminus was blocked and that the blocking group was not a cyclized glutamic acid or glutamine residue.

Cyanogen Bromide Fragments from Electrolectin. The amino acid composition of electrolectin indicates the presence of three methionine residues (15). To take advantage of this favorable situation, the protein was hydrolyzed using cyanogen bromide in 70% (vol/vol) formic acid. Only a single large fragment and one small fragment were isolated, purified, and sequenced partially. The expected third (blocked) amino-terminal peptide was not recovered. The results are summarized in Fig. 1 in which the fragments are aligned based on the chicken lectin cDNA sequence (16). When 70%

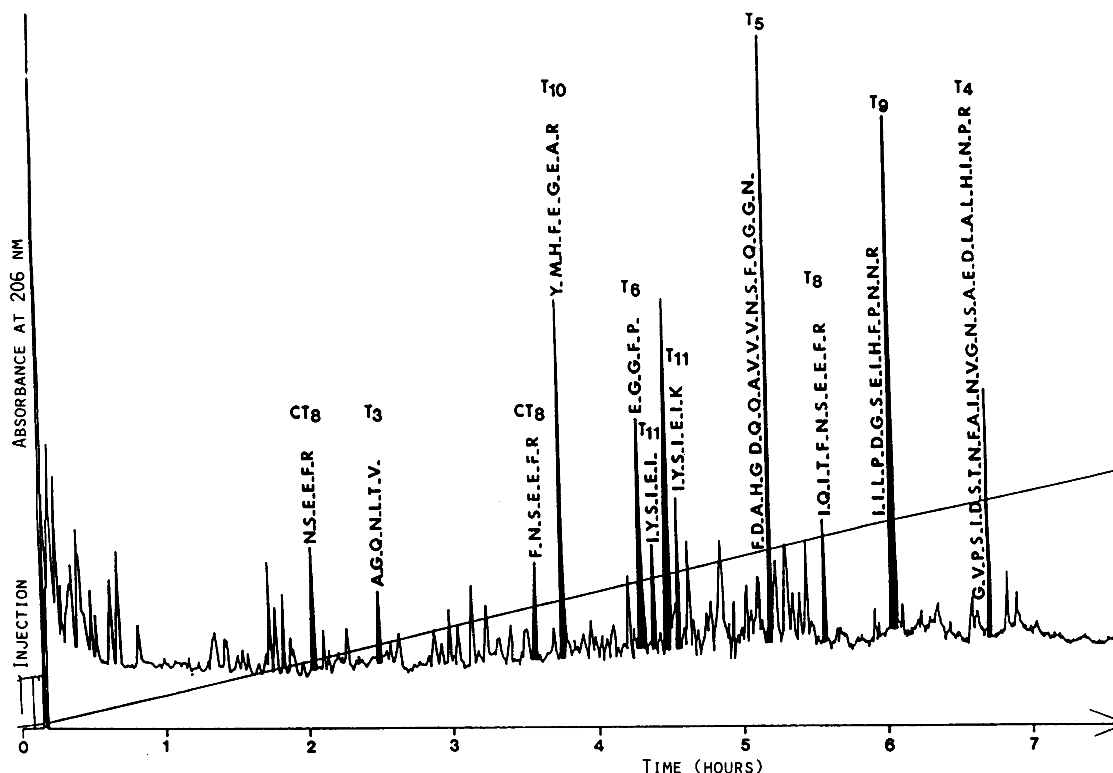

FIG. 2. A typical HPLC profile of a tryptic digest of electrolectin. Each fraction (in black) was analyzed in a gas-phase sequenator, and the corresponding sequence is indicated above the peak in the single-letter amino acid code.

(vol/vol) formic acid was replaced by a 50% (vol/vol) heptafluorobutyric acid/50% (vol/vol) formic acid solution, cleavage was also observed at an additional site, likely to correspond to a Trp-Xaa peptide bond. This new peptide was sequenced partially to yield the sequence Gly-(Thr)-Glu-Gln.

Tryptic Peptides. Most of the sequence of electrolectin and all the results on the human protein were obtained by analysis of the tryptic peptides that had been separated on a RP300 column.

We present in Fig. 2 a typical separation and indicate along the peaks the sequences of the corresponding peptides. The peaks were numbered according to the presumed position of the peptides in the whole sequence presented in Fig. 1. Two additional peptides (CT8) were obtained, probably by chymotryptic activity in the trypsin.

***S. aureus* Protease.** Additional electrolectin sequences and overlapping sequences were obtained from the analysis of peptides resulting from the digestion of the eel lectin by the protease from *S. aureus*, which in the conditions used in this work cleaves the proteins at the Glu-Xaa peptide bonds. Peptides were separated by HPLC prior to automated sequence analysis.

Alignment of Peptides. The sequence analyses of peptides obtained from the chemical and proteolytic cleavages have permitted the assignment of the amino acid sequence of the galactoside-binding lectin from electric eel and of >60% of that of the human placenta lectin. The alignment of the peptides was based on the extensive homologies of these lectins with the galactoside-binding 14-kDa binding protein from chicken embryo (16) (Fig. 3).

Secondary Structures of Eel and Chicken Lectins. Characteristic parameters of the secondary structures of the chicken and eel lectins, as predicted from their sequences, were calculated using a software package of the University of Wisconsin Genetics Computer group with a Microvax II computer. Fig. 4 shows the results obtained for these lectins in terms of their respective propensity to form β-sheets and their hydrophobic character (19). Similar calculations can be made for the propensity to form α-helices but since the results predicted that none of the two lectin sequences form α-helices, the actual data are not shown.

In Fig. 4 the regions of the curve above the threshold line with a value of 1.00 indicate the amino acid residues probably involved in β-sheet structures. Accordingly, both the chicken and eel lectins would appear to be composed of at least 10 β-sheets all formed with amino acid residues at equivalent positions in the two proteins.

By using the hydropathy scale of Kyte and Doolittle (20), one observes a striking similarity in the hydropathy profiles of the two proteins.

DISCUSSION

Since their discovery (1), the vertebrate β-D-galactoside binding lectins have been the subject of a large number of investigations and of several reviews (11, 21, 22). These studies show that the β-D-galactoside binding lectins from species as distinct as the electric eel, the frog, the chicken, and humans share a number of common properties. They all have subunits in the range of 12–16 kDa, have a higher affinity for β-D-galactopyranosyl β-D-thiogalactopyranoside than for lactose and galactose, require reducing agents for maintaining their agglutinating activity, and often cross react immunologically. These results suggest that these vertebrate lectins have maintained structural homologies during evolution.

The amino acid sequence data, the propensity to form β-sheets, and hydropathy profiles presented here illustrate the extent of this homology. The chicken and eel lectins have 51 identical residues over 130 positions that were compared, i.e., a homology of >39%. The identities are distributed over the whole length of the proteins, with the exception of the last third of the chains that display much less similarity than the rest of the polypeptides.

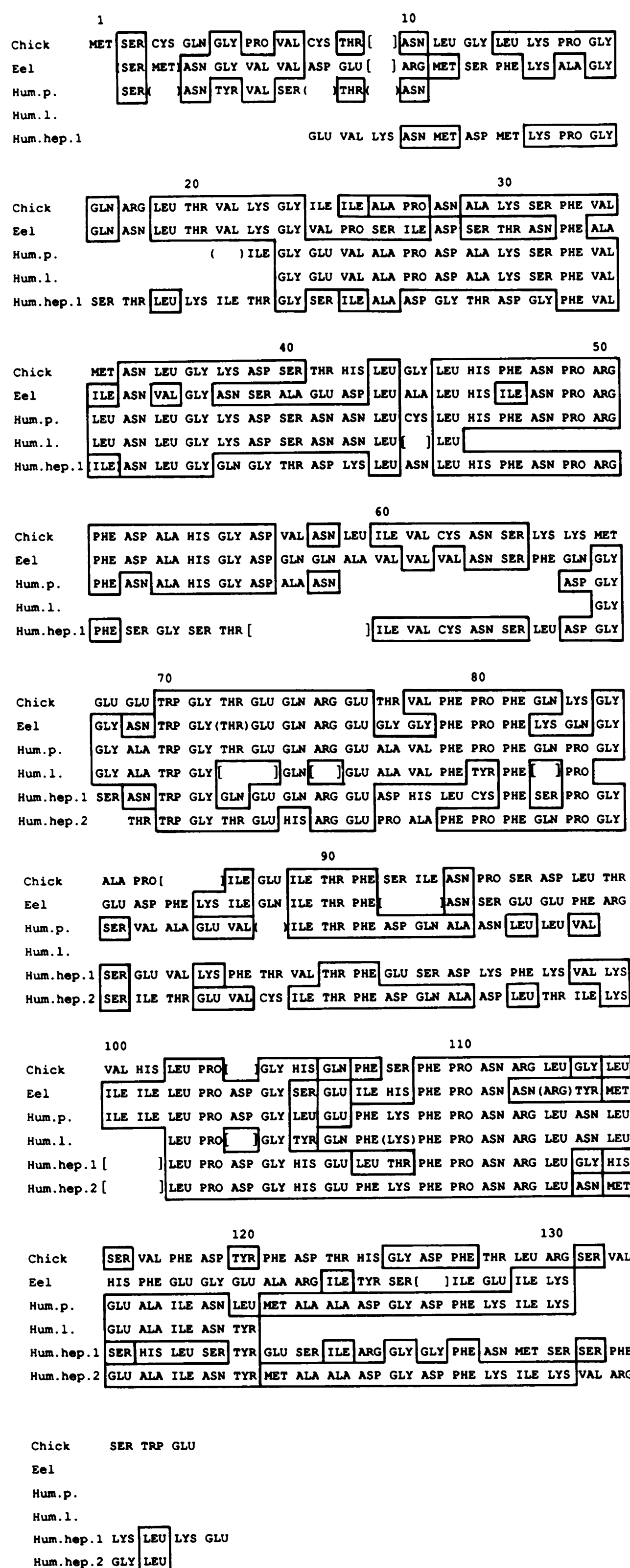

FIG. 3. Comparison of the sequences obtained for the β-galactoside binding lectins from eel and human placenta (Hum.p) and lung (Hum.l) (17) with those derived from cDNAs of chicken embryo skin (16), and human hepatoma Li-7 cells. (Hum.hep.1 and Hum.hep.2) (17).

In addition to the 51 identical residues, there are 27 positions that are occupied by homologous residues, coded

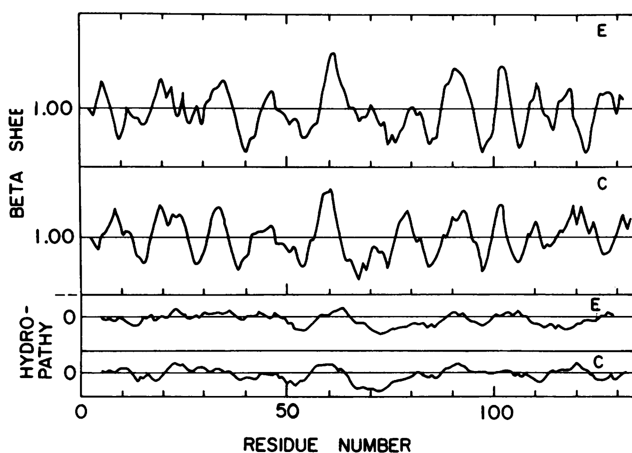

FIG. 4. Predicted β-sheet structures and hydropathicity profiles of chicken and eel β-galactoside binding lectins. The two upper curves were calculated using the method of Chou and Fasman (19). The lower two curves were obtained using the hydropathy scale of Kyte and Doolittle (20).

by codons that only differ by a single nucleotide. Although the blocked amino-terminal peptide has not been formally isolated and sequenced, it is likely to be composed of an *N*-acetylated Ser-Met sequence as was also predicted for the other lectins. The tryptophan was assigned to position 70 by the analysis of the peptides resulting from cyanogen bromide cleavage in a heptafluorobutyric acid/formic acid mixture. The assignment of the tryptophan residue to position 70 is also made likely in view of the fact that residues 70–76 are highly conserved in all the sequences presented in Fig. 3. The similarity of the β-sheet forming regions and the hydropathy profiles of the chicken and eel lectin sequences strengthens the validity of the alignment of the peptides and emphasizes the strong structural homology of the lectins. The comparison between the chicken and human placenta lectins yields 59 identical residues over 111 positions, which are comparable once the insertions of three gaps in the chicken protein are introduced. Several peptide sequences are in fact identical in the two proteins: residues 29–33, 35–40, 45–51, 53–56, 70–76, 78–82, 89–91, 109–113, and 125–127. The comparison between the eel lectin and the human placenta lectin yields 47 identical residues over 111 positions, with insertions of three gaps in the eel lectin. To these one should add the likely presence in the placenta lectin of a number of conserved lysine and arginine residues immediately preceding the start of tryptic peptides. This is the case for position 65 (lysine in the chicken lectin) and positions 50 and 99 (arginine in the chicken and the eel lectin, respectively).

The comparison between the partial sequences of the four lectins of human origin shows significant homologies but also systematic differences. Thus 50 residues out of 54 comparable positions (93% homology) are identical in the lung and placenta lectins whereas only 41% homology is found (42 residues identical out of the 103 comparable residues) between the placenta lectin and the putative product of the hepatoma cDNA clone 1. These results indicate that at least four different genes encoding β-D-galactoside binding lectins may be present in the human genome. However, the homologies between the human lectins and those from the eel and chicken suggest that the lectin-encoding genes have been derived from a common ancestor gene.

The observation that some peptide sequences are highly conserved in all the lectins sequenced so far may allude to the fact that these peptides correspond to essential structural determinants that have been maintained during evolution. The β-D-galactoside binding site is possibly one of such determinants. We tend to assign at least part of the β-D-galactoside binding site to residues 70–76. This peptide sequence contains a tryptophan and two glutamic acid residues in positions 70, 73, and 76, respectively. These amino acids are expected to be present in the β-D-galactoside binding site of the eel lectin since they account for the spectrofluorometric properties of this protein. Levi and Teichberg (15) have found that lactose enhances the fluorescence of the eel lectin and prevents the deleterious effects of O_2 that causes a decrease of the lectin binding activity together with a disappearance of its fluorescence. On the basis of these observations, they concluded that the single tryptophan residue of the eel lectin is present in the lactose binding site. Moreover, they inferred from the pH dependence of the fluorescence of the eel lectin measured in the presence and absence of lactose that a carboxyl residue with an abnormal pK was present in the lactose binding site. Since vicinal carboxyl acids are known to influence each other's ionization, it is plausible that glutamic acid residues 73 and 76 mutually affect their respective pK values and that the result is that one carboxyl group has an abnormal pK value.

In any event, the significant homology demonstrated here between the β-D-galactoside binding lectins from vertebrate species as phylogenetically distant as the eel, the chicken, and man is likely to reflect an important physiological role for these lectins in ontogenic events and differentiation steps.

This research was supported by grants to A.D.S. from Centre National de la Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale, and Ministère de la Recherche et de la Technologie and to V.I.T. from the Leo and Julia Forchheimer Center for Molecular Genetics at the Weizmann Institute of Science, the Minerva Foundation, and the Gesellschaft für Biotechnologische Forschung. V.I.T. holds the Louis and Florence Katz-Cohen Professorial Chair of Neuropharmacology.

1. Teichberg, V. I., Silman, I., Beitsch, D. D. & Resheff, G. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 1383–1387.
2. Teichberg, V. I. (1978) *Methods Enzymol.* **50,** 291–302.
3. Bols, N. C., Roberson, M. M., Haywood-Reid, P. L., Cerra, R. F. & Barondes, S. H. (1986) *J. Cell. Biol.* **102,** 492–499.
4. Kobiler, D., Beyer, E. C. & Barondes, S. H. (1978) *Dev. Biol.* **64,** 265–272.
5. Levi, G. & Teichberg, V. I. (1985) *Biochem. J.* **226,** 379–384.
6. de Waard, A., Hickman, S. & Kornfeld, S. (1976) *J. Biol. Chem.* **251,** 7581–7587.
7. Powell, J. T. (1980) *Biochem. J.* **187,** 123–129.
8. Childs, R. A. & Feizi, T. (1979) *Biochem. J.* **183,** 755–758.
9. Hirabayashi, J. & Kasai, K. I. (1984) *Biochem. Biophys. Res. Commun.* **122,** 938–944.
10. Cerra, R. F., Gitt, M. A. & Barondes, S. H. (1985) *J. Biol. Chem.* **257,** 10474–10477.
11. Barondes, S. H. (1984) *Science* **223,** 1254–1264.
12. Childs, L. A. & Feizi, T. (1979) *Biochem. J.* **183,** 755–758.
13. Beyer, E. C., Tokuyasu, K. T. & Barondes, S. H. (1979) *J. Cell Biol.* **82,** 565–571.
14. Levi, G. & Teichberg, V. I. (1982) *FEBS Lett.* **82,** 565–571.
15. Levi, G. & Teichberg, V. I. (1981) *J. Biol. Chem.* **256,** 5735–5740.
16. Ohyama, Y., Hirabayashi, J., Oda, Y., Ohno, S., Kawasaki, H., Suzuki, K. & Kasai, K. I. (1986) *Biochem. Biophys. Res. Commun.* **134,** 51–56.
17. Gitt, M. A. & Barondes, S. H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7603–7607.
18. Houmard, J. & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 3506–3509.
19. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13,** 211–222.
20. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
21. Simpson, D. L., Thorne, D. R. & Loh, M. H. (1978) *Life Sci.* **22,** 727–748.
22. Barondes, S. H. (1981) *Annu. Rev. Biochem.* **50,** 207–231.