Sequence of a Full-Length cDNA for Rat Lung β-Galactoside-Binding Protein: Primary and Secondary Structure of the Lectin†

Linda B. Clerch,*,‡ Philip Whitney,‡ Michael Hass,‡ Keith Brew,§ Todd Miller,§ Rudolf Werner,§ and Donald Massaro‡,∥

Pulmonary Research Laboratories, Calvin and Flavia Oak Asthma Research and Treatment Facility, Department of Medicine, and Department of Biochemistry, University of Miami School of Medicine, Miami, Florida 33136, and Veterans Administration Medical Center, Miami, Florida 33125

Received August 17, 1987; Revised Manuscript Received October 6, 1987

ABSTRACT: A full-length cDNA for rat lung β-galactoside lectin (subunit M_r ~14000, lectin 14K) was cloned and the nucleotide sequence determined. The deduced amino acid sequence agrees with the amino acid composition and direct amino acid sequence analysis of purified rat lung lectin peptides. We found that the amino-terminal alanine is blocked with an acetyl group. Comparison of the amino acid sequence with other proteins shows a high degree of homology only with other vertebrate lectin sequences, supporting the suggestion that these lectins may constitute a unique class of vertebrate proteins. The amino acid composition and sequence of lectin peptides, the sequence of lectin cDNA, and isoelectric focusing of purified lectin indicate that rat lung lectin 14K is composed predominantly of a single protein. In addition, rat uterus lectin 14K was found to be the same protein as that present in lung. We characterized the secondary and tertiary structure of rat lung lectin 14K by circular dichroism, by analytical ultracentrifugation, and by computer analysis of its primary structure. Results of these experiments suggest that lectin 14K is primarily a hydrophilic protein with an asymmetric, elongated structure consisting of approximately equal amounts of α helix, β sheet, β turn, and random coil. We found that Cys-2 and Cys-130 react most rapidly with iodoacetamide; one or both of these residues may be primarily responsible for the thiol requirement of lectin activity.

The lectin that is the subject of this report is an endogenous, soluble, dimeric β-galactoside-binding protein with subunit molecular weight ~14000 (lectin 14K); it is found in rat lung, uterus, and heart (Powell, 1980). Lectin 14K is a member of the class of soluble vertebrate lectins that requires exogenous thiols but not metal ions for hemagglutinating activity (Barondes, 1984). At present, their physiological function is unknown. It is hoped that more complete definition of the structures and properties of these lectins will provide some insights for further studies on lectin function.

The primary structures of some lectins have been reported. The amino acid sequence of chick skin lectin 14K was deduced from the cDNA nucleotide sequence (Ohyama et al., 1986) and from peptide sequences (Hirabayashi et al., 1987a). Southan et al. (1987) reported a nearly complete sequence of bovine heart lectin 14K by direct amino acid sequence analysis. Gitt and Barondes (1986) determined a partial amino acid sequence for human lung lectin 14K and deduced partial sequences based on two cDNA clones from a human hepatoma. The studies by Gitt and Barondes showed that there are three human lectin genes. All of the reported vertebrate lectin 14K amino acid sequences show striking homologies with each other, but computer searches of protein sequence data bases have not detected significant homologies with other proteins. This suggests these vertebrate lectins are members of a unique class of proteins. We report here the first full-length sequence of a mammalian lectin of this class and confirm that it is only homologous with the soluble vertebrate lectins.

Our studies also provide a more detailed explanation of the requirement for exogenous thiols to maintain lectin activity. We previously showed that reaction of iodoacetamide with cysteinyl residues in rat lectin 14K enhances lectin activity and protects against loss of activity that would occur as the lectin thiol groups form disulfide bonds under oxidizing conditions (Whitney et al., 1986). In the current study, we used radioactive iodacetamide to identify which cysteinyl residues are reactive in rat lectin in its native, reduced state.

In addition, we have used information from the primary structure, circular dichroism, and hydrodynamic properties of rat lectin 14K to characterize the secondary and tertiary structures.

Materials and Methods

Materials. DNA modifying and restriction enzymes were obtained from New England Biolabs, Beverly, MA, or Bethesda Research Laboratories, Gaithersburg, MD. Radioactive chemicals were obtained from New England Nuclear, Boston, MA.

Animals. Adult Sprague–Dawley albino rats were obtained from Charles River Laboratories, Portage, ME. They were allowed rodent laboratory chow 5001 (Ralston Purina, St. Louis, MO) and water ad libitum. Lighting was provided from 7 a.m. to 7 p.m. Rats were killed by exsanguination after anesthesia with sodium pentobarbital (~60 mg/kg, intraperitoneal).

Lectin Purification. Lectin from rat lung or rat uterus was purified and assayed by hemagglutination as previously described (Whitney et al., 1985). The dimeric lectin (M_r ~

† This research was supported by NIH Grants HL32109, HL20633, HL07283, and GM21363 and by V.A. Research Funds. D.M. is a Medical Investigator of the Veterans Administration. This paper is based on work done toward a Ph.D. degree (L.B.C.) at the Department of Biochemistry, University of Miami.

* Address correspondence to this author.

‡ Pulmonary Research R120, University of Miami School of Medicine, P.O. Box 016960, Miami, FL 33101.

§ Department of Biochemistry, University of Miami School of Medicine.

∥ V.A. Medical Center.

0006-2960/88/0427-0692$01.50/0

14 000 subunits) elutes from lactosyl-Sepharose before the monomeric lectins [M_r ~18 000 and 29 000; see Cerra et al. (1985)]. The purified lectin 14K migrated as a single band on polyacrylamide gel electrophoresis (Laemmli, 1970). To achieve the best results from isoelectric focusing, lectin must be purified quickly.

Differential Labeling of Lectin Cysteinyl Residues. Rat lung lectin [3 mg in 3.6 mL of 20 mM sodium phosphate buffer (pH 7.2)–140 mM NaCl–1 mM ethylenediaminetetraacetic acid (EDTA)] in its native, reduced state was alkylated by adding 0.4 mL of 1.8 mM iodo[^{14}C]acetamide (ICN Biochemicals, 5 Ci/mol) in 1.0 M tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), pH 7.8. After 14 h at 4 °C, the reaction was terminated by adding 6.2 mg of dithiothreitol (10 mM). Fifteen minutes later, the protein was denatured by adding sufficient solid guanidine hydrochloride to make a 6 M solution. Six hours later, the protein was reacted with 30 mM iodoacetamide for 1 h and then dialyzed against 20 mM sodium phosphate buffer (pH 7.2)–140 mM NaCl–1 mM EDTA. After this two-stage procedure to selectively label the surface cysteinyl residues, the alkylated protein had 24 000 dpm/nmol of subunit.

Peptide Production and Purification. The amino terminus of lectin was found to be blocked to Edman degradation, so the protein had to be cleaved to generate fragments suitable for sequence analysis. Peptides produced chemically or proteolytically were purified by high-performance liquid chromatography (HPLC) using a reverse-phase C-4 column (Vydac 214TP54) with acetonitrile gradients in 0.1% trifluoroacetic acid. Some peptides were subjected to sequence analysis using an Applied Biosystems Model 470A gas-phase sequencer. Specific peptides are designated by residue numbers as given in Figure 4.

Following cleavage with cyanogen bromide (Findlay & Brew, 1972), the reaction mixture was dried and the residue dissolved in 8 M urea for purification of the peptides. Peptides 115–120 and 121–134 were sequenced: peptide 1–114 was found to have a blocked N-terminus.

CNBr peptide 1–114 (700 μg) was digested by 25 μg of endopeptidase Arg C (Boehringer Mannheim Biochemicals) for 10 h at 37 °C in 1 mL of 50 mM NH_4HCO_3. Peptides 1–20, 1–48, 21–48, 49–73, and 74–114 were purified. Peptides 1–20, 1–48, and 74–115 were subjected to sequence analysis; the first two were found to have a blocked amino terminus.

Arg C peptide 1–48 (40 μg) was digested by 1 μg of trypsin [tosylphenylalanine chloromethyl ketone (TPCK) treated; Worthington Biochemical Corp.] for 5 h in 0.4 mL of 50 mM Tris-HCl buffer (pH 8.2)–1 mM $CaCl_2$. Peptides 1–18, 19–28, 29–36, and 37–48 were purified; peptide 29–36 was sequenced.

Arg C peptide 1–20 was digested by 10 μg of endoproteinase Glu C (from *Staphylococcus aureus* V8, Boehringer Mannheim Biochemicals) in 0.3 mL of 0.1 M Tris-HCl buffer (pH 7.8) overnight at 37 °C.

Fast Atom Bombardment Mass Spectrometry. Tryptic peptide 1–18 (200 ng) was analyzed by Dr. Roger Laine (Department of Biochemistry, Louisiana State University) using a Kratos MS 80 with 8-keV zenon fast atom bombardment and scanning at 300 s/decade through the range up to 2400-dalton ionic mass.

Protein Analytical Procedures. Samples for amino acid analysis were hydrolyzed in 6 N HCl for 22 h at 110 °C in a vacuum. Most samples were analyzed with a JEOL 5AH using a 0.8 cm × 46 cm column (sulfonated polystyrene; 12 ± 2 μm beads) and ninhydrin detection. The trypsin peptides were analyzed by reverse-phase HPLC following derivatization with phenyl isothiocyanate (Scholze, 1985).

Analytical ultracentrifugation was done by Dr. J. F. Woessner and Caroline Taplin (Department of Biochemistry, University of Miami) using a Beckman Model E instrument with interference optics. A sample of lectin [0.9 mg/mL in 10 mM Tris-HCl buffer (pH 7.2)–150 mM NaCl–0.1 M EDTA] from adult rat lung was centrifuged at 47 900 rpm at 20 °C in a double-sector cell; $s_{20,w}$ was calculated from the sedimentation rate of the midpoint of the protein boundary with corrections for solvent density and viscosity.

Isoelectric focusing was done by using a Hoefer Mighty Small electrophoresis unit with 0.5-mm-thick slabs of 7.5% (w/v) polyacrylamide at 15 °C in 8 M urea–2% (v/v) Triton X-100–10% (v/v) glycerol–1% (v/v) BioLyte 3/10–2% (v/v) BioLyte 4/6–2% (v/v) Biolyte 5/7. Samples were dissolved in 8 M urea–2.5% (v/v) Triton X-100–15% (v/v) glycerol–1% (v/v) Biolyte 3/10–5% (v/v) BioLyte 4/6–100 mM dithiothreitol (DTT)–50 μg/mL bromophenol blue. The ampholytes in the gel were prefocused for 1 h with the power supply initially set at a high enough current to give 500 V; with time, the voltage increased to a maximum of 1200 V. After samples were loaded, the current was set at 2 mA and the voltage again allowed to go to the maximum of 1200 V; sample focusing time was 3.5 h.

Polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate was also done with the Hoefer Mighty Small unit at 15 °C using 13% (w/v) polyacrylamide in the separating gel (Laemmli, 1970). After electrofocusing or electrophoresis, the proteins were made visible by silver staining (Merril et al., 1981).

Circular dichroism were measured in an Aviv-modified Cary 60 dichrograph. Lectin samples (0.2 mg/mL) were dialyzed against 2 mM Tris-HCl buffer (pH 7.2)–100 mM NaCl under N_2 and placed in a cuvette (2-mm path length), and the solutions were scanned from 250 to 200 nm. Each recorded spectrum was the average of five repetitive scans.

Lectin concentration was determined by amino acid analysis.

Construction of a Rat Uterus cDNA Library. Poly(A)-containing RNA was isolated from estrogen-primed uteri of 4–6-month-old Sprague–Dawley rats by the method of Chirgwin et al. (1979). Double-stranded cDNA was synthesized from this RNA by the procedure of Huynh et al. (1984) and ligated to phage λgt10 DNA. The recombinant phage DNA was then packed in vitro (Packagene, Biotech, Madison, WI). The library had a total of 2.4 × 10^5 phage.

Construction of a Lung cDNA Library. Poly(A)-containing RNA was isolated from adult rat lungs by the method of Chirgwin et al. (1979). DNA complementary to lung RNA was synthesized with a kit (Amersham, Arlington Heights, IL) based on the method of Gubler and Hoffman (1983). The cDNA was ligated to λgt11 DNA and packaged (Packagene). The cDNA library had a recombinant titer of 2.2 × 10^5 phage.

Construction and Labeling of Oligonucleotide Probes. The amino acid sequences of residues 73–88 and 114–121 (determined by peptide sequencing) were used as the basis for the synthesis of oligonucleotide probes to screen lectin-specific cDNA clones from the rat uterus cDNA library constructed in λgt10. Probe 114–121 was a mixture of 16 sequences 23 nucleotides long with the sequence 5′GCCATWTAWTTIATUGCXTCCAT3′ (New England Biolabs, Beverly, MA) (where X = C or T, U = I or C, and W = A or G). Probe 73–88 was a single sequence 48 bases in length with the sequence 5′ACAGACCTCTGTTATGGATCCCGGCTGGAAAGGGAAGGCCGTCTCCCT3′

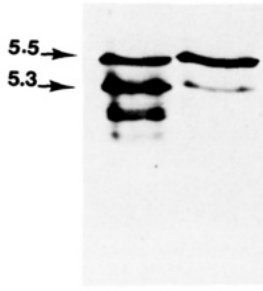

FIGURE 1: Isoelectric focusing of rat lung lectin. (Lane A) Lectin was purified from adult rat lungs by the usual procedure (Whitney et al., 1985) and stored at 4 °C in 20 mM sodium phosphate buffer (pH 7.2)–140 mM NaCl–5 mM dithiothreitol–10 mM lactose for 4 months followed by an 18-h incubation with 20 mM DTT; 250 ng of lectin was electrofocused. (Lane B) Lectin was quickly purified from adult rat lungs by using a single lactosyl-Sepharose column; 150 ng was electrofocused as described under Materials and Methods.

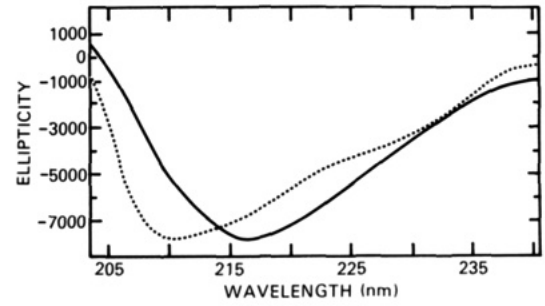

FIGURE 2: Native lectin, maintained in its reduced state, was prepared from lungs of neonatal rats (age 8 days). The circular dichroism (solid line) is expressed as mean residue weight ellipticity in degrees centimeter squared per decimole. The dotted line is a computer-generated curve for 27% α helix, 22% β sheet, 28% β turn, and 23% random coil.

(synthesized by Dr. Richard Voellmy, Department of Biochemistry, University of Miami).

The oligonucleotide probes were end-labeled with ^{32}P using T4 polynucleotide kinase and [γ-^{32}P]ATP by standard methods (Maniatis et al., 1982) and were isolated from the labeling reaction mixture by column chromatography on Sephadex G-25.

Screening of the Uterus cDNA Library. A total of 6×10^5 plaques of the λgt10 uterus cDNA library were screened by standard methods (Davis et al., 1980). Nitrocellulose filter replicas were hybridized with probe 114–121. (Probe 73–88 was not useful for screening because we found that it hybridized nonspecifically to λDNA.) Twenty-three clones that hybridized with the oligonucleotide probe were identified and plaque purified. The cDNA inserts of six phage clones that hybridized to both probe 114–121 and probe 73–80 on a Southern blot (Southern, 1975) were isolated and inserted into the plasmid vector pGEM-blue (Promega Biotech, Madison, WI) for further characterization.

Isolation and Labeling of the Uterus cDNA Probe. The 147 base pair insert from the rat lectin uterus cDNA clone, designated U7A, was isolated by *Eco*RI digestion, electrophoresis through a 0.8% agarose gel, and electroelution. The insert was labeled with [α-^{32}P]dCTP by nick translation (Maniatis et al., 1982) and isolated from the reaction mixture by column chromatography on Sephadex G-50.

Screening of the Lung cDNA Library. A total of 6×10^5 plaques of the λgt11 lung cDNA library were screened by the method of Davis et al. (1980). Replicas on nitrocellulose filters were hybridized with the nick-translated cDNA insert from uterus clone U7A that corresponds to the coding region for the amino acid sequence of residues 57–105. Seventy clones were identified. The cDNA inserts of four phage clones were isolated and inserted into pGEM-blue plasmid vectors for further characterization.

DNA Sequence Analysis. All DNA sequences were determined by the Sanger dideoxynucleotide termination method (Sanger et al., 1977) using the GEM Seq K/RT sequencing system (Promega Biotech., Madison, WI). Both the T7 and SP6 promoter primers were used to sequence from both strands of each insert. The nucleotide sequences were analyzed by using the sequence analysis software package of the University of Wisconsin Genetics Computer Group.

RESULTS

Molecular Characterization. Isoelectric focusing of rapidly purified lectin from adult rat gave a major component at pH 5.5 and a minor one at pH 5.3 (Figure 1). Purified lectin from neonatal rat (age 8 days) gave the same result (data not shown). If the samples were stored for more than a few days at 4 °C, the band at pH 5.3 increased, and one or two bands at lower pH were observed (Figure 1). The reason for the shift in isoelectric points is not known, but it is not reversed by addition of dithiothreitol. Substitution of thiodigalactoside for lactose during purification and storage did not alter the course of the change in isoionic points.

An indication of secondary structural features was obtained by measuring circular dichroism from 250 to 200 nm (Figure 2). The spectrum shown in Figure 2 is of lung lectin from age 8-day rats; lectin from adult rats gave the same spectrum (data not shown). The spectrum was analyzed with a computer program (Pro Sec version 2.1, Aviv Associates, Lakewood, NJ) based on the work of Chang et al. (1978). The best fit corresponded to 27% α helix, 22% β sheet, 28% β turn, and 23% random coil. A theoretical curve using these parameters does not match the experimental curve very well (Figure 2), so these values must be considered to be approximations. Addition of lactose to the lectin solution did not produce a significant change in the spectrum; this indicates that specific saccharide binding is not accompanied by large changes in lectin structure.

A measure of the overall conformation of the dimeric lectin can be obtained from its hydrodynamic properties. We measured its sedimentation rate in the ultracentrifuge and found an $s_{20,w} = 2.48$ S. Although the sedimentation coefficient was not determined at several concentrations and extrapolated to zero protein concentration, the sample was dilute enough so that this should not have been a significant source of error. We had previously determined that lectin remains as a dimer in this concentration range (Whitney et al., 1986). Using the sedimentation coefficient and an estimate of 0.73 for the partial specific volume [calculated from the amino acid composition according to Cohn and Edsall (1943)], we calculated a frictional coefficient of 5.34×10^{-8} compared to 3.86×10^{-8} for a nonhydrated sphere of molecular mass 29 800 daltons. The ratio of frictional coefficients is 1.38. If a hydration factor of 0.2 g of H_2O/g of protein is assumed, this frictional ratio decreases to 1.27 and would correspond to an axial ratio of about 5 for an elongated ellipsoid or 6 for a flattened ellipsoid (Cohn & Edsall, 1943), indicating that the lectin has an asymmetric structure in solution.

Peptide Purification and Sequencing. Since the amino acid composition of lung lectin indicated the presence of two methionyl residues, it was expected that CNBr cleavage would generate three peptide fragments. HPLC separation of this CNBr digest gave three well-resolved peaks, suggesting that

Table I: Amino Acid Composition of Lectin Peptides[a]

peptide:	CNBr 3	CNBr 1	CNBr 2	Arg 2	Arg 3	Arg 1	Arg 4
residue no.:	1–114	115–120	121–134	1–20	21–48	49–73	74–114
Asp + Asn	19.1 (19)	1.0 (1)	2.0 (2)	2.0 (2)	7.3 (6)	5.9 (6)	5.0 (5)
Thr	6.5 (7)	0	0	0	1.3	2.8 (3)	3.5 (4)
Ser	4.5 (5)	0	0	0.9 (1)	1.4 (2)	0.9 (1)	1.2 (1)
Glu + Gln	9.0 (9)	1.0 (1)	1.1 (1)	1.0 (1)	1.4 (1)	2.0 (2)	5.1 (5)
Pro	6.7 (7)	0	0	0.9 (1)	1.9 (2)	0	4.0 (4)
Gly	8.6 (9)	0	1.0 (1)	2.0 (2)	2.3 (2)	2.9 (3)	2.2 (2)
Ala	7.8 (8)	1.0 (1)	2.9 (3)	2.0 (2)	2.0 (2)	2.0 (2)	2.3 (2)
Cys (Cam-Cys)	5.3 (5)	0	1.1 (1)	1.9 (2)	1.0 (1)	0.9 (1)	1.0 (1)
Val	4.2 (5)	0	0.8 (1)	1.9 (2)	1.1 (1)	0.9 (1)	1.1 (1)
Met (Hse)	0.6 (1)	0.7 (1)	0	0	0	0	0.6 (1)
Ile	3.9 (4)	1.0 (1)	1.0 (1)	0	0.3	0.9 (1)	3.0 (3)
Leu	11.8 (12)	0	0	4.0 (4)	3.6 (5)	0	3.4 (3)
Tyr	0	1.0 (1)	0	0	0	0	0
Phe	8.5 (8)	0	2.0 (2)	0	2.0 (2)	0.9 (1)	5.1 (5)
His	3.0 (3)	0	0	0	1.4 (1)	1.0 (1)	1.1 (1)
Lys	6.9 (7)	0	2.0 (2)	2.0 (2)	1.9 (2)	1.0 (1)	2.2 (2)
Arg	4.2 (4)	0	0	1.0 (1)	1.0 (1)	1.1 (1)	1.2 (1)
Trp	NT[b] (1)					NT (1)	

[a] Peptides from cleavage with CNBr or endoproteinase Arg C were purified by HPLC and analyzed for amino acid composition. The peptides are identified by the mode of cleavage and the order of elution from the reverse-phase column. Residue number is taken from numbering as shown in Figure 4. The expected compositions based on the amino acid sequence are given in parentheses. Methionine was present as homoserine (Hse) and cysteine as the carboxamidomethyl (Cam) derivative. Results were not corrected for losses during hydrolysis. [b] NT = not tested.

complete cleavage had occurred. The sequences of two peptides, CNBr 1 and CNBr 2, identified retrospectively as being derived from residues 115–120 and 121–134, respectively, were readily determined [underlined residues below bases 346–405 (Figure 3)] and were in good agreement with the amino acid compositions (Table I). CNBr 3 (residues 1–114) had a blocked N-terminus; its composition is in good agreement with the deduced sequence for this region (Figure 4).

Fragments generated from CNBr 3 by cleavage at arginyl residues were separated by HPLC (Arg peptides 1–4, Table I). Arg peptide 4 (residues 74–114) was subjected to sequencing for 15 cycles to give the sequence underlined below bases 223–267 in Figure 3. Sequence analysis of Arg peptide 2 indicated the presence of a blocked amino terminus. Arg peptides 1, 2, and 3 were identified from their amino acid compositions (Table I).

Because of incomplete cleavage of CNBr 3 at Arg-20, another fragment (corresponding to residues 1–48) with a blocked amino terminus was isolated from the above digest. Tryptic hydrolysis of this large fragment gave four fragments. One was sequenced and found to correspond to residues 29–36 (underlined residues below bases 88–111 in Figure 3); the compositions of the other fragments were consistent with the deduced sequences for residues 1–18, 19–28, and 37–48.

Amino-Terminal Blocking Group. Tryptic fragment 1–18 was analyzed by fast atom bombardment mass spectrometry and gave an ion at 1986.6 ± 1 daltons. The calculated molecular mass of the peptide ion without a blocking group is 1943.6 daltons; the difference is exactly what would be predicted for an acetyl group. We conclude that the N-terminal alanine is blocked with an acetyl group.

Cysteine Reactivity. One aim of these studies was to identify which of the cysteinyl residues are on the surface of the native protein molecule. This can be assessed from the specific radioactivity of each cysteinyl residue since radioactive iodoacetamide was used only during reaction with lectin in its native, reduced form. CNBr and arginine cleavages separated the cysteines into separate peptides except that Cys-2 and Cys-16 were both in Arg peptide 1–20. This peptide was cleaved at Glu-15, and two radioactive peptides were found. Most of the radioactivity (92%) was found in peptide 1–15. Setting the specific radioactivity of Cys-2 at 1.00, the relative reactivity of the other cysteinyl residues was the following:

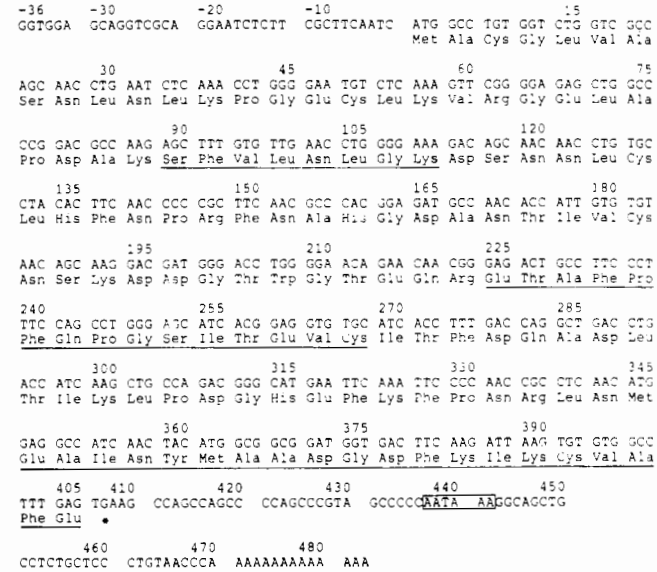

FIGURE 3: Nucleotide sequence and deduced amino acid sequence of lung β-galactoside-binding lectin cDNA. The amino acid sequences determined by direct sequencing of lung lectin peptides are underlined. The presumptive poly(A) addition signal is boxed. The superscript numbers refer to the numbers of the corresponding nucleotides of the cDNA sequence. The asterisk indicates the stop codon.

Cys-16, 0.09; Cys-42, 0.000; Cys-60, 0.003; Cys-88, 0.12; Cys-130, 0.72.

Characterization of Uterus Lectin. Uterus was previously reported to have lectin 14K (Powell, 1980; Gabius et al., 1986). However, further characterization was necessary to show that uterus and lung lectins are the same. The hemagglutinating activity of the purified uterus lectin, its subunit molecular weight (determined by sodium dodecyl sulfate–polyacrylamide gel electrophoresis), its isoelectric focusing pattern, and its amino acid composition were not significantly different from corresponding values for purified rat lung lectin (data not shown). Therefore, uterus and lung lectins appeared to be the same protein, and it was appropriate to screen the uterus cDNA library for lectin cDNA.

Isolation of a Partial Lectin cDNA from the Rat Uterus cDNA Library. As described under Materials and Methods, we isolated phage clones based on insert hybridization to ^{32}P

FIGURE 4: Comparison of amino acid sequences for seven vertebrate β-galactoside-binding lectins. RLL, cDNA of rat lung lectin determined in this work; CSL, cDNA of chick skin lectin (Ohyama et al., 1986); BHL, amino acid sequence of bovine heart lectin (Southan et al., 1987); HLL, peptides from human lung lectin (Gitt & Barondes, 1986); HHC1, cDNA from human hepatoma clone 1 (Gitt & Barondes, 1986); HHC2, cDNA from human hepatoma clone 2 (Gitt & Barondes, 1986); HPL, peptides of human placenta lectin (Hirabayashi et al., 1987). Dashes indicate gaps introduced for optimal alignment; complete gaps indicate undetermined sequences. The superscript numbers refer to the number of the corresponding amino acid in the sequence. Sequence identities between rat lectin and one or more chains are enclosed by boxes. Percent identity of the rat lung lectin sequence to the other vertebrate lectins is shown at the bottom right.

end-labeled oligonucleotide probes. We chose to pursue the clone designated U7. *Eco*RI digestion of U7 yielded two fragments, U7A and U7B, that on a Southern blot hybridized with probe 73–88 and probe 114–121, respectively. An *Eco*RI restriction site in the coding region of a lectin clone would account for the presence of two fragments. [An *Eco*RI restriction site at a location corresponding to a point between the two probes was in fact found in the lectin cDNA of human hepatoma clone 2 (Gitt & Barondes, 1986).] U7A and U7B were subcloned into pGEM-blue plasmid and sequenced. The deduced amino acid sequences of U7A and U7B match exactly the amino acid sequences from lung peptides (residues 74–88 and 115–134). U7A is a 147 base pair fragment that corresponds to amino acid residues 57–105. U7B is a 143 base pair fragment that corresponds to positions 106–134 in the amino sequence together with a noncoding section of 53 base pairs on the 3′ side of the stop codon. The nucleotide sequences of U7A and U7B are identical with the corresponding sequence of lung lectin cDNA. U7A is the same as bases 172–318, and U7B is the same as bases 319–461 in Figure 3. Clone U7B includes the poly(A) addition signal. The insert U7A was used as a specific probe to screen the lung cDNA library.

Isolation of a Full-Length Lectin cDNA from the Rat Lung cDNA Library. Of the four clones chosen for further characterization because they hybridized with ^{32}P nick-translated cDNA probe U7A, we pursued the clone designated L25. *Eco*RI digestion of L25 gave two fragments, L25A and L25B. On a Southern blot, they hybridized to cDNA probe U7A and oligonucleotide probe 114–121, respectively. L25A and L25B were subcloned into the plasmid vector pGEM-blue and sequenced. The results of DNA sequencing are shown in Figure 3.

L25A is a 351 base pair fragment that extends from base −36 to base 315 in Figure 3; L25B extends from base 316 to base 483. The full-length cDNA has 519 nucleotides. A single open-reading frame of 405 nucleotides codes for a polypeptide of 134 amino acids with a molecular weight of 14840. There are 36 base pairs 5′ and 78 base paris 3′ of the coding region. The 3′ region consists of a TGA stop codon at position 406 followed by a polyadenylation recognition site 28 base pairs downstream. Poly(A) starts 61 base pairs 3′ of the stop codon and is 14 A's long. The deduced amino acid sequence from bases 87–111, 223–267, and 396–405 was confirmed by direct amino acid sequence analysis of purified lung lectin peptides (underlined amino acids in Figure 3). In addition, the remainder of the deduced amino acid sequence agreed with amino acid composition analysis of lung peptides shown in Table I. A comparison of the deduced amino acid sequence of rat lung lectin with other vertebrate lectins is shown in Figure 4. The percentages of identical amino acids indicate a significant homology between these lectins. A computer search of the National Biomedical Research Foundation (NBRF) protein sequence data base did not reveal homologies with any other protein. This result supports the previous suggestion that these vertebrate lectins are a novel class of proteins (Southan et al., 1987). Computer analysis using the program ALIGN of the protein identification resource of the NBRF did not reveal any significant internal homologies in lung lectin sequence when comparing residues 1–31, 31–63, 63–103, and 103–134.

The deduced amino acid sequence was used to generate a computer-predicted secondary structure by the method of Chou and Fasman (1978). The top three graphs in Figure 5 show the propensity of certain regions of the polypeptide backbone to form specific secondary structures. The predictions indicate that there is no predominant single type of structure. Computer analysis of the sequence data by the method of Garnier (1978) predicts that the lectin will have about 23% α helix, 20% β sheet, 33% β turn, and 24% random coil. Although this must be considered to be an approximation of the true structure, it is very close to that calculated from the circular dichroism spectrum. Peaks of predicted structures occur in highly conserved regions of amino acid sequence for α helix around residue 74, β sheet around residues 34, 55, and 60, and β turn around residues 50, 56, 66, 105, and 112. With the possible exception of cysteines-2 and -130, the cysteines are not in regions predicted for α helix or β turn; cysteines-60 and -88 are in hydrophobic regions with high probability of β sheet. The overall hydropathy score and profile [by the method of Kyte and Doolittle (1982) using a window size of

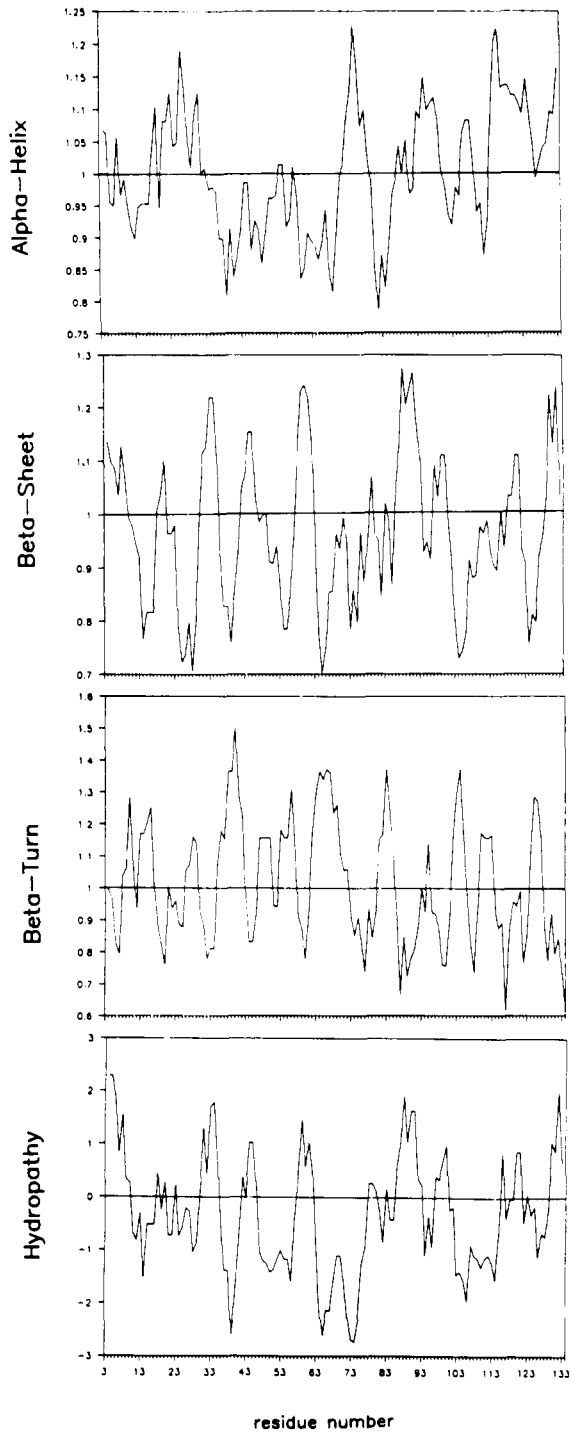

FIGURE 5: Prediction of secondary structure and hydropathy from the amino acid sequence of rat lectin 14K. α helix, β sheet, and β turn scores were computed by the method of Chou and Fasman (1978). Scores greater than 1 predict a high probability of forming a particular structure. Hydropathy scores were computed by the method of Kyte and Doolittle (1982). Values greater than 0 are indicative of hydrophobic regions; values less than 0 indicate hydrophilicity.

six amino acids] for the lectin show a preponderance of hydrophilic regions (bottom graph in Figure 5). There are no lengthy hydrophobic sequences that would indicate membrane-spanning ability.

Discussion

Gitt and Barondes (1986) found three human lectin 14K genes. In light of this report, it was necessary to isolate a single clone containing the full-length cDNA; it would not be appropriate to overlap sequences from different clones that may have arisen from homologous but distinct genes. The deduced amino acid sequence reported in this paper is taken from one clone and thus represents the product of a single gene.

Because there are multiple genes for human lectin 14K, we examined our results for evidence of multiple species of rat lectin 14K. Several lines of evidence suggest there is only one predominant lectin 14K in rat lung and uterus and that the cDNA clones we sequenced correspond to the predominant or only form expressed in these tissues: (1) The lectin from uterus was the same as the lectin from lung based on agglutinating activity, subunit molecular weight, isoelectric focusing, and the amino acid composition. (2) The sequence of the lectin partial cDNA from uterus exactly matched the corresponding sequence of the lung lectin cDNA, so the two clones coded for the same amino acid sequence. (3) The amino acid sequence of four peptides from lung lectin exactly matched the sequences deduced from the cDNA sequences. (4) The CNBr peptides were isolated in high yield, and their amino acid compositions closely matched the compositions predicted from the deduced amino acid sequences. (5) At least 90% of the lung lectin appears in one band at pH 5.5 after isoelectric focusing.

The pattern of bands obtained after isoelectric focusing changes depending on how the lectin is purified and stored. If the lectin is purified quickly, there is only the major band at pH 5.5 and a minor one at pH 5.3. Upon storage at 4 °C for a few weeks, the band at pH 5.5 becomes weaker, the band at pH 5.3 is stronger, and additional, more acidic bands appear. Treatment of the "aged" lectin with dithiothreitol did not change the pattern of bands, so disulfide bond formation is not likely to be the cause. Another possible cause for the changing isoionic points could be reaction of lysyl residues with the reducing aldehyde of lactose followed by Amadori rearrangement to form a stable product (Acharya & Manning, 1980). This does not appear to contribute significantly to the aging process because substitution of thiodigalactoside (which is not a reducing disaccharide) for lactose did not alter the course of the change in isoionic points. The "aging" process probably does not produce great changes in the lectin molecule since the lectin retains its hemagglutinating activity. We do not know if the minor component at pH 5.3 in freshly isolated lectin is the product of a different gene or if it is due to a natural posttranslational modification or a time-dependent process during purification. In any event, these results support the conclusion that one protein accounts for at least 90% of the soluble, dimeric, β-galactoside-specific lectin in rat lung. The expression of multiple lectin genes in human hepatoma cells (Gitt & Barondes, 1986) does not necessarily imply that there are multiple forms of lectin in normal human lung; perhaps the multiple bands observed after isoelectric focusing of human lung lectin (Sparrow et al., 1987) are due to the same type of aging process we find with rat lung lectin.

Examination of amino acid sequences for internal homologies can give clues regarding gene evolution. Hirabayshi et al. (1987) reported that chick lectin 14K may have evolved via several gene duplications. Our results do not add further support to this hypothesis because the scores for sequence alignments of internal regions of rat lectin were not significanly higher than those for randomized sequences of the same amino acid composition. If the initial lectin gene did evolve by gene duplication, the duplicated regions in the rat protein have diverged so much that this event is no longer evident in the amino acid sequence.

A striking feature of the amino acid sequence of rat lectin 14K is its homology with lectins from other vertebrate species.

Compared with lectins that are known to be expressed in normal tissue, the strongest homology is with lectins from human lung and placenta. Taking all of the lectin sequences into consideration, only 11 of 134 residues are unique to the rat lectin, and these are all in the N-terminal half of the molecule. Of the 12 aromatic amino acids in rat lectin, 7 are present in every sequence in Figure 4, and 3 others are found in every sequence except human clone 1. Of the polypeptide regions that are especially well conserved, only region 58–63 lacks an aromatic residue. The conservation of the aromatic residues suggests that they may be important for folding and maintaining the lectin in its native, active conformation.

In view of the importance of thiol groups in maintaining lectin function, it is surprising that Cys-16 and Cys-43 are among the 11 residues unique to rat lectin and that Cys-88 and Cys-130 have only 1 match and 3 mismatches. Therefore, the cysteinyl residues are generally poorly conserved. Cys-2 and Cys-60 are more conserved and may be more directly involved in the thiol sensitivity.

We previously reported that two cysteinyl residues in rat lectin reacted very rapidly with iodoacetamide and this reaction was accompanied by a small increase in lectin activity. More prolonged reaction resulted in the modification of one to two more cysteinyl residues. The final product was no longer sensitive to oxidative inactivation (Whitney et al., 1986). Results reported here show that Cys-2 and Cys-130 are the two residues modified in lectin activation, Cys-16 and Cys-88 are modified more slowly, and Cys-42 and Cys-60 are not modified and are likely to be buried in the native lectin structure. Oxidative inactivation of rat lectin is linked to formation of intrachain disulfide bonds (Whitney et al., 1986) accompanied by dramatic changes in secondary structures as evidenced by a shift in the trough in the circular dichroism spectrum from 216 nm for native lectin to 207 nm for oxidized lectin; the shift was reversed by reduction with dithiothreitol (unpublished results). The spectral change is consistent with the disruption of regular secondary structures (α helix β sheet) to a random coil structure. Perhaps the formation of disulfide bonds locks the protein into a new, inactive conformation that cannot form the usual secondary structures and cannot bind saccharides. This suggests that the regular secondary structure is a vital part of maintaining the active lectin conformation.

The active conformation of the dimeric lectin does not appear to be spherical. The sedimentation velocity measurements indicate that the lectin dimer is asymmetric with an estimated axial ratio of about 5:1 for an elongated ellipsoid. This asymmetry is probably due in part to an elongated structure resulting from the assembly of subunits into the dimer. An alternative explanation for the sedimentation data is that the molecule could have a flexible region between globular domains as was found in immunoglobulins (Noelken et al., 1965). The existence of separated domains in the lectin molecule is unlikely in view of its small size and lack of internal homology. Perhaps the asymmetry (or flexibility) is an important feature that gives the lectin dimer a more extended range for noncovalent cross-linking of carbohydrate structures.

An important question regarding the in vivo function of the lectin is whether it acts within the cell or extracellularly. Immunocytochemical studies have detected lectin in extracellular locations (Barondes et al., 1984); this suggests that lectin may function extracellularly. In contrast, we failed to find specific lectin secretion from human lung fibroblasts grown to confluence in tissue culture. These cells contain high concentrations of intracellular lectin (Whitney et al., 1985), but only 3% of the lectin was found in the medium after 3, 6, or 24 h; 5% of the total lactic acid dehydrogenase was found in the medium (unpublished results). This suggests that extracellular lectin was there as a result of cell damage or nonspecific exocytosis of cell contents.

Characteristics of lectin structure and activity point to an intracellular localization. (1) The requirement to keep the thiol groups in their reduced form would generally limit the lectin to an intracellular location. Lability of its binding activity in the presence of oxygen could place restrictions on the latitude of extracellular functions unless sufficient reduced thiols were present. (2) There is no signal peptide sequence before the N-terminal alanine to direct the protein for secretion so that specific secretion of the lectin would require a special mechanism. (3) The N-terminal alanine is blocked with an acetyl group. In all cases reported, lectins of this type have a blocked N-terminal amino acid. The N-terminal serine in chick lectin is also blocked with an acetyl group (Hirabayashi et al., 1987a), and this is the most likely blocking group for the other lectins as well. Acetylated N-terminal amino groups are usually associated with cytoplasmic proteins (Tsunasawa & Sakiyama, 1984). Although the presence of these characteristics of rat lung lectin does not rule out an extracellular function, it would seem prudent to also consider possible intracellular functions.

ACKNOWLEDGMENTS

We acknowledge use of the facilities of the Protein Identification Resource of the National Biomedical Research Foundation. We gratefully thank Dr. Frans Huijing from the Department of Biochemistry for his useful ideas and discussion. We thank Dr. David Puett from the Department of Biochemistry for allowing us to use his Aviv-Cary dichrograph and Ondina Garcia-Pons for her assistance in the preparation of the manuscript.

REFERENCES

Acharya, A. S., & Manning, J. M. (1980) *J. Biol. Chem. 255*, 7218–7224.

Barondes, S. H. (1984) *Science (Washington, D.C.) 223*, 1259–1264.

Barondes, S. H., Cerra, R. F., Cooper, D. N. W., Haywood-Reid, P. L., & Roberson, M. M. (1984) *Biol. Cell 51*, 165–172.

Cerra, R. F., Gitt, M. A., & Barondes, S. H. (1985) *J. Biol. Chem. 260*, 10474–10477.

Chang, C. T., Wu, C. C., & Yang, J. T. (1978) *Anal. Biochem. 91*, 13–31.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5299.

Chou, P., & Fasman, G. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–147.

Cohn, E. J., & Edsall, J. T. (1943) *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*, Hafner Publishing Co., New York.

Davis, R. W., Botstein, D., & Roth, J. R. (1980) *A Manual for Genetic Engineering, Advanced Bacterial Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Findlay, J. B. C., & Brew, K. (1970) *Eur. J. Biochem. 27*, 65–86.

Gabius, H.-J., Engelhardt, R., Graupner, G., & Cramer, F. (1986) *Lectins: Biol., Biochem., Clin. Biochem. 5*, 237–242.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.

Gitt, M. A., & Barondes, S. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7603–7607.

Gubler, U., & Hoffman, B. J. (1983) *Gene 25*, 263–269.

Hirabayashi, J., Kawasaki, H., Suzuki, K., & Kasai, K. (1987a) *J. Biochem. (Tokyo) 101*, 775–787.
Hirabayashi, J., Kawasaki, H., Suzuki, K., & Kasai, K. (1987b) *J. Biochem. (Tokyo) 101*, 987–995.
Huynh, T. V., Young, R. A., & Davis, R. W. (1984) *DNA Cloning Techniques: A Practical Approach* (Glover, D., Ed.) IRL Press, Oxford, U.K.
Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.
Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Merril, C. R., Goldman, F., Sedman, S. A., & Ebert, M. H. (1981) *Science (Washington, D.C.) 211*, 1437–1438.
Noelken, M. E., Nelson, C. A., Buckley, C. E., III, & Tanford, C. (1965) *J. Biol. Chem. 240*, 218–224.
Ohyama, Y., Hirabayashi, J., Oda, Y., Ohno, S., Kawasaki, H., Suzuki, K., & Kasai, K. (1986) *Biochem. Biophys. Res. Commun. 134*, 51–56.
Powell, J. T. (1980) *Biochem. J. 187*, 123–129.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Scholze, H. (1985) *J. Chromatogr. 350*, 453–460.
Southan, C., Aitken, A., Childs, R. A., Abbott, W. N., & Feizi, T. (1987) *FEBS Lett. 214*, 301–304.
Southern, E. (1975) *J. Mol. Biol. 98*, 503–517.
Sparrow, C. P., Leffler, H., & Barondes, S. H. (1987) *J. Biol. Chem. 262*, 7383–7390.
Tsunasawa, S., & Sakiyama, F. (1984) *Methods Enzymol. 106*, 165–170.
Whitney, P., Maxwell, S., Ryan, U., & Massaro, D. (1985) *Am. J. Physiol. 284*, C258–C264.
Whitney, P. L., Powell, J. T., & Sanford, G. L. (1986) *Biochem. J. 238*, 683–689.

Structural Characterization of Recombinant Hepatitis B Surface Antigen Protein by Mass Spectrometry†

Mark E. Hemling,‡ Steven A. Carr,*,‡ Carine Capiau,§ and Jean Petre§

Department of Physical and Structural Chemistry, Smith Kline & French Laboratories, P.O. Box 1539, King of Prussia, Pennsylvania 19406-0939, and Department of Human Vaccines Development, Smith Kline Biologicals, rue de l'Institute 89, B-1330 Rixensart, Belgium

Received June 18, 1987; Revised Manuscript Received September 18, 1987

ABSTRACT: The primary structure of recombinant hepatitis B surface antigen protein produced in yeast has been confirmed by mass spectrometric peptide mapping. These studies corroborate more than 85% of the amino acid sequence derived by sequencing of the gene and identified the presence of an acetyl moiety on approximately 70% of the NH_2-terminal methionine residues. Prior to the present work, direct structural analysis was largely prevented by the insolubility of this integral membrane protein and its primary degradation fragments in aqueous buffers and by partial blockage of the NH_2 terminus. These difficulties were overcome by preparative isolation using electroelution of the monomeric 226 amino acid protein from a polyacrylamide electrophoretic gel in the presence of sodium dodecyl sulfate. Chymotryptic digestion of the reduced and carboxymethylated monomer produced a large number of small, predominantly hydrophobic peptides ideally suited for peptide mapping by fast atom bombardment mass spectrometry. The percentage of NH_2-terminal methionine blocked by acetyl was determined by a new strategy involving cyanogen bromide cleavage, permethylation, and gas chromatography/mass spectrometry identification and quantitation of the *N*-methyl-*N*-acetylhomoserine produced.

Hepatitis B is a health problem of enormous magnitude afflicting more than 200 million people worldwide [for a recent review, see Tiollais et al. (1985)]. In the United States, about 0.1–2.0% of the population are carriers of the virus while in Far-East Asia and tropical Africa more than 10% of the population are carriers. Infection with hepatitis B produces a spectrum of clinical manifestations ranging from asymptomatic infection to lethal fulminant hepatitis (Junge & Deinhardt, 1985). Mortality most frequently results from chronic active hepatitis and liver disease.

The infective agent of hepatitis B belongs to a new family of hepadna viruses (Marion & Robinson, 1983). The blood of infected individuals contains three morphologically distinct carriers of the hepatitis B surface antigen (HBsAg).[1] The virion, usually present in only small amounts, is a spherical particle 42 nm in diameter consisting of an envelope and a

† This work was supported in part by l'Institut pour l'Encouragement de la Recherche dans l'Industrie et l'Agriculture (IRSIA), B-1050 Brussels, Belgium. This work was presented in part at the 34th Annual Conference on Mass Spectrometry and Allied Topics, June 8–13, 1986, Cincinnati, OH.

* Correspondence should be addressed to this author.
‡ Smith Kline & French Laboratories.
§ Smith Kline Biologicals.

[1] Abbreviations: HBsAg, hepatitis B surface antigen; HBcAg, HB core antigen; HBeAg, HB e antigen; HBV, HB virus; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; HPLC, high-performance liquid chromatography; Tris, tris(hydroxymethyl)-aminomethane; EDTA, ethylenediaminetetraacetic acid; RCM, reduced and carboxymethylated; FAB, fast atom bombardment; MS, mass spectrometry; TFA, trifluoroacetic acid; DMSO, dimethyl sulfoxide; GC, gas chromatography; PTH, phenylthiohydantoin; PTC, phenyl isothiocyanate; kDa, kilodalton(s).

0006-2960/88/0427-0699$01.50/0