Porcine Cerebroside Sulfate Activator: Further Structural Characterization and Disulfide Identification†

Richard L. Stevens,‡ Kym F. Faull,§ Kenneth A. Conklin,§ Brian N. Green,∥ and Arvan L. Fluharty*,§,⊥

Department of Biological Chemistry, the Neuropsychiatric Institute, and the Department of Psychiatry and Biobehavioral Sciences, UCLA School of Medicine, Los Angeles, California 90024, VG Biotech, Tudor Road, Altrincham, WA14 5RZ England, and the Mental Retardation Research Group, Lanterman Developmental Center, Pomona, California 91768

Received September 30, 1992; Revised Manuscript Received January 19, 1993

ABSTRACT: Cerebroside sulfate activator (CS-Act) is a small compact protein which binds and solubilizes certain glycosphingolipids. Following the recent publication of the purification and preliminary sequence of pig kidney CS-Act [Fluharty, A. L., Katona, Z., Meek, W. E., Frei, K., & Fowler, A. V. (1992) *Biochem. Med. Metab. Biol. 47*, 66–85], we now report the primary sequence of the C-terminal portion of this protein and the assignment of the three disulfide bonds. Cyanogen bromide (CNBr) treatment of native CS-Act produced three major and several minor peptide fragments. Analysis of one HPLC-purified fragment revealed the C-terminus 14 amino acid sequence. This established the length of the native protein at 79 residues. In conjunction with the sequence data for one other major HPLC-purified CNBr fragment, it could be concluded that the three intrachain disulfide bonds were located at half-cystine residues 4 and 77, 7 and 71, and 36 and 47. Mass spectrometry (fast atom bombardment and electrospray ionization) showed the molecular weight of the major component of the CS-Act preparation to be 9720.5 Da, which was in close agreement with the calculated mass of the 79 amino acid peptide with five covalently attached sugar residues and three internal disulfide bonds. The mass spectrometric molecular weight measurements also showed that the CS-Act preparation possessed microheterogeneity in its carbohydrate moiety, as less intense signals corresponded to species containing (in decreasing order of abundance) two, one, four, and three sugar residues. The same type of measurements on native and chemically reduced CS-Act revealed a mean difference of 5.9 ± 0.2 Da (mean ± SEM, $n = 3$) for the three most abundant species in the preparation, confirming that in the native state the molecules were oxidized, with each containing three disulfide bonds. With account made for the contribution from the carbohydrate components, the molecular weight measurements of the CNBr fragments agreed closely with the predicted values, concordant with the disulfide bridge assignments. The secondary structure prediction analysis shows a high preference for α-helical content in this protein. The disulfide placement data, the prediction analysis information, and the property of high stability has led us to revise our previously proposed CS-Act structure to one which contains a bundle of four amphipathic α-helices forming an internal hydrophobic core.

Cerebroside sulfate activator (CS-Act), a small, heat-stable, water-soluble glycoprotein which is essential for the in vivo hydrolysis of cerebroside sulfate by arylsulfatase A, was first described over 25 years ago (Mehl & Jatzkewitz, 1964). In 1979 renewed interest was generated in CS-Act when its deficiency (Shapiro et al., 1979; Stevens et al., 1981) was discovered to be the cause of a rare form of metachromatic leukodystrophy, a disease ordinarily due to the deficiency of the lysosomal enzyme arylsulfatase A. Subsequently Cs-Act was shown to be identical to activators of certain other sphingolipid hydrolases, most notably the G_{M1} ganglioside and globotriaosylceramide activators (Li et al., 1985). Thus, besides being called CS-Act (Sandhoff, 1984), it is known by several other names, such as nonspecific activator (Li et al., 1988), sphingolipid activator protein 1, or SAP-1 (Innui & Wenger, 1984), and saposin B (O'Brien & Kishimoto, 1991).

The designation saposin B was coined when it became apparent that this activator protein was derived from a larger precursor protein, prosaposin. The precursor, glycosylated at one or more of five potential glycosylation sites, is approximately 60–70 kDa in mass and encodes four regions of similar structure, designated saposins A, B, C, and D, each of which leads to a separate sphingolipid activator protein (O'Brien et al., 1988; O'Brien & Kishimoto, 1991). A recent report (Holschmidt et al., 1991) on the detailed organization of the human prosaposin gene, which is on chromosome 10 (Inui et al., 1985), showed it to be in excess of 17 kb in size and to have 15 exons, with 3 or 4 exons comprising the saposin B segment. The processing of the human precursor protein would be expected to lead to mature activators of 80–84 amino acids with at least one N-linked carbohydrate chain (Fujibayashi & Wenger, 1986; Furst et al., 1988). The amino acid sequence of human Cs-Act protein (saposin B) was initially deduced from the cDNA sequence (Dewji et al., 1986) as well as the cDNA sequence of prosaposin (Dewji et al., 1987; Furst et al., 1988). Human CS-Act protein has also been directly sequenced and was found to contain 80 amino acids, with an N-linked carbohydrate chain presumably located at the missing Asn-21 position (Nakano et al., 1989). The sequence was colinear with that from the cDNAs (O'Brien et al., 1988). Inspection of the Cs-Act sequences shows the molecule to have six cysteine residues; these are presumed to be in the

† This research was supported in part by an NIH grant to A.L.F. (NS-11665) and by a grant from the United Leukodystrophy Foundation. K.F.F. and K.A.C. were supported in part by a NIDA grant (DA05010) and a grant from the W. M. Keck Foundation.

* Address correspondence to this author at UCLA-MRRC, 760 Westwood Plaza, Los Angeles, CA 90024-1759.

‡ Department of Biological Chemistry, UCLA School of Medicine.

§ Neuropsychiatric Institute and Department of Psychiatry and Biobehavioral Sciences, UCLA School of Medicine.

∥ VG Biotech.

⊥ Mental Retardation Research Group.

0006-2960/93/0432-4051$04.00/0

oxidized form, but the specific disulfide bridging patterns were not known.

Recently we have reported (Fluharty et al., 1992a,b) the purification and properties of CS-Act protein from pig kidney, a source of higher content and greater availability than the human sources. The final product, of greater than 90% purity by gel electrophoretic and HPLC criteria, was sequenced to the 64th residue from the N-terminus. The sequence differed at about 10% of the residues from the human CS-Act protein. From the behavior on sizing columns the active molecule appeared to be approximately 20 kDa in mass, and a guanidine hydrochloride treated preparation appeared half this size, consistent with an approximately 80 amino acid residue monomer. Sugar analyses of several preparations showed the presence of glucosamine and fucose with smaller amounts of mannose. A stable 1:1 complex of cerebroside sulfate with the dimer form of CS-Act was demonstrated. The CS-Act protein was extremely refractory to denaturation by high temperatures and was highly resistant to most proteases.

In the present study a purified preparation of pig kidney CS-Act protein was used to complete the determination of the amino acid sequence of the protein, establish the various glycoforms, and identify the placement of the three disulfide bridges. These parameters were evaluated by sequence and mass spectrometric analysis of the intact protein and HPLC-purified CNBr fragments of the native activator.

EXPERIMENTAL PROCEDURES

Preparation of CS-Act. The protein was prepared from the kidneys of freshly butchered pigs essentially as previously described (Fluharty et al., 1992a). In brief, frozen ground kidneys were homogenized in distilled water and then stirred at 90 °C for 5 min. The supernatant was filtered and placed on a Con A Sepharose column. The breakthrough fractions were adsorbed onto octyl Sepharose 4B, and CS-Act was eluted with octyl β-D-glucopyranoside. The dialyzed solution was then sequentially chromatographed on DEAE-Sephadex and Sephacryl S-300 HF. Greater than 90% purity was estimated by gel electrophoresis and reversed-phase HPLC criteria. Activity measurements with arylsulfatase A and [^{35}S]cerebroside sulfate showed greater than 2500 units/mg protein. A unit of activity is defined as hydrolysis of 1 nmol of cerebroside sulfate/4 h under the prescribed assay conditions (Mehl & Jatzkewitz, 1964).

Chemical Reduction of CS-Act. To 2.5 nmol of CS-Act in 40 μL of 3 mM Tris-HCl buffer (pH 7.5) was added 1 μL of 100 mg/mL dithiothreitol (600 nmol, 40-fold molar excess of dithiothreitol to cysteine sulfhydryl groups). The solution was incubated for 3 h at 37 °C and then taken to dryness in a vacuum concentrator.

Cyanogen Bromide Cleavage. Purified pig kidney CS-Act (162 μg in 30 μL of 25 mM Tris-HCl buffer, pH 7.5) was treated with 70 μL of formic acid containing 4.5 mg of cyanogen bromide (CNBr:methionine molar ratio 330:1) for 24 h at room temperature. The mixture was diluted 3-fold with water and taken to dryness in a vacuum concentrator. The residue was redissolved in water (300 μL), and the drying procedure was repeated. The final residue was dissolved in 100 μL of 0.1% trifluoroacetic acid (TFA) in preparation for HPLC.

HPLC Separation of the CNBr Fragments. After removal of 1 μL for amino acid analysis, the remaining CNBr mixture was separated by C-4 HPLC (column, 4.6 mm × 25 cm Vydac C-4 reverse phase; buffer, 0.1% TFA in water with a linear gradient of acetonitrile containing 0.1% TFA from 0 to 50% over a time period of 5–65 min at 1 mL/min; 1-mL fractions; detection by absorbance at 214 nm). The collected fractions were frozen, and subsequently aliquots were subjected to amino acid and mass spectrometric analysis. Fractions corresponding to peaks 2, 3, and 4 (Figure 1) were concentrated and then refractionated by C-18 HPLC (column, 2 mm × 30 cm Waters Bondapak C-18 reverse phase; buffer, 0.1% TFA in water with a linear gradient of acetonitrile from 0 to 100% over 120 min at 0.2 mL/min; 0.2-mL fractions). Amino acid analysis, Edman sequencing, and mass spectrometry were performed on the major peaks.

Amino Acid Analysis. All amino acid compositions were determined on aliquots that were performic acid oxidized prior to 6 N HCl vapor hydrolysis at 110 °C for 18 h. The neutralized hydrolysates were derivatized to the phenylthiocarbamoyl amino acids and were separated and quantitated by using the Waters Picotag system (Cohen & Strydom, 1988).

Peptide Sequencing. Samples (100–800 pmol) were adsorbed onto a Polybrene-treated glass filter and sequenced on an ABI 470A Sequenator equipped with an ABI 900A computer and a 120A on-line HPLC phenylthiohydantoin (PTH) analyzer.

Mass Analysis. Fast atom bombardment (FAB) spectra were obtained with a VG ZAB-SE instrument (Fisons Instruments, VG Analytical, Manchester, U.K.) using an 11/250 data system, an 8-kV accelerating potential, zenon bombarding gas at 8 kV and 1 mA, and a mass resolution of 500 (10% valley, M/ΔM). Aliquots (1–2 μL) of native CS-Act (5.4 mg/mL in water) and HPLC-purified samples (30–100 pmol in water/acetonitrile/TFA, 50/50/1) were applied to the static FAB probe tip onto which had already been placed 1–2 μL of liquid matrix. Glycerol, thioglycerol, *m*-nitrobenzyl alcohol (mNBA), and a mixture of mNBA/thioglycerol/trifluoroacetic acid (100/100/1, NOD matrix) were tried as matrices; the most intense signals were obtained with NOD matrix. Data were collected by scanning from *m/z* 10 000 to 8000 for CS-Act and appropriate lower ranges for the fragments, and about 10 scans were collected from each sample into a multichannel analyzer. The data were smoothed and centroided, and the mass was measured using cesium iodide ion clusters for calibration. Under ideal conditions the accuracy of the mass measurements carried out in this manner is typically about ±1 Da.

Electrospray ionization (ESI) mass spectra were obtained by dissolving the samples in water/acetonitrile/formic acid (50/50/1) and injecting 10-μL aliquots of the resulting solutions into the infusion line connected to an electrospray ion source attached to a quadrupole mass spectrometer (VG Bio-Q, Fisons Instruments, VG Biotech, Altrincham, U.K.). Data were collected by scanning from *m/z* 400 to 2000, and about 10 scans were summed from each sample into a multichannel analyzer. The raw data were smoothed and centroided, and the mass was measured using the multiple charged ion series from a separate introduction of horse heart myoglobin (mol wt 16951.5) for calibration. In this form the data are represented on the abscissa as the mass/charge (*m/z*) ratio. Because ESI spectra typically show for each component several ions of differing charge state, it aids interpretation to transform the data into a true molecular mass spectrum in which the abscissa is in units of mass. This was achieved by the software package supplied with the spectrometer. Under ideal conditions the accuracy of the mass measurement carried out in this manner is typically ±0.01% of the molecular weight. Selected samples were also analyzed on an API III (Sciex, Thornhill, Canada) and a TSQ 700

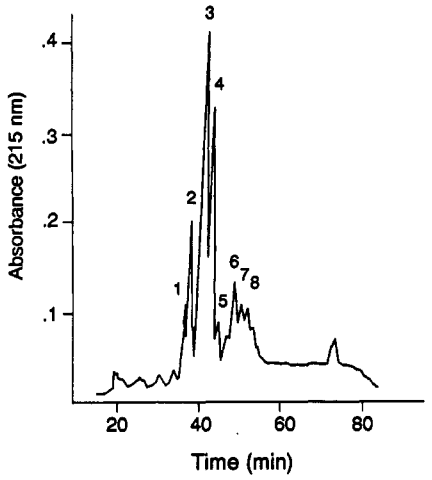

FIGURE 1: HPLC profile from a C-4 reversed-phase column following injection of the CNBr digest of purified porcine CS-Act (for details, see the text). Fractions most closely corresponding to peaks 1–8 were used for further processing and analysis.

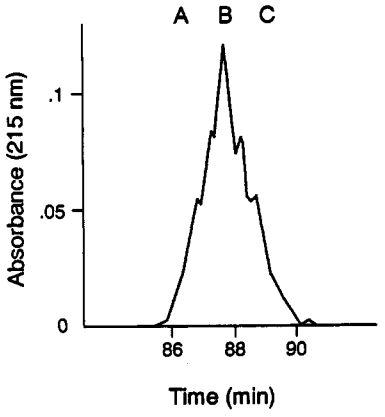

FIGURE 2: HPLC profile from a C-18 reversed-phase column following injection of the peak-4 fraction from Figure 1 (for details, see the text). Letters correspond to fractions from the leading (A), the central (B), and the trailing portion (C) of the peak. These were utilized for various analyses.

(Finnegan MAT, San Jose, CA) instrument, with essentially identical results.

RESULTS

Characterization of the CNBr Cleavage Products. Single-step reversed-phase HPLC of CNBr-treated CS-Act revealed eight discernible UV absorbing peaks (henceforth referred to as fragments 1–8, Figure 1). The untreated CS-Act preparation, chromatographed separately, eluted as a single peak toward the end of the gradient, 6 min later than the last-eluting fragment, at a retention time of 57 min. There were no discernible peaks in this portion of the chromatogram of the CNBr digest. Amino acid analysis revealed the presence of homoserine (produced from methionine by the CNBr reaction) and cysteine (determined as cysteic acid after performic acid oxidation) in all eight fragments, suggesting the presence of at least one disulfide bridge in each. The three largest peaks (fragments 2, 3, and 4), accounting for approximately 66% of the total UV absorbance, were re-chromatographed on a C-18 column. In each case the profile of UV absorbance suggested microheterogeneity (the profile of fragment 4 is shown in Figure 2; the profiles from fragments 2 and 3 were similar but are not shown), and selected fractions, labeled a, b, and c, collected across each peak were taken for sequence, amino acid, and mass analysis.

Table I: Sequence Analysis of CNBr Fragment 2[a]

cycle no.	peptide 1–10		peptide 66–79	
	amino acid	pmol	amino acid	pmol
1	Gly	314	Gln	364
2	Asp	45[b]	Pro	157
3	Val	480	Lys	412
4	Cys		Asp	215
5	Gln	236	Ile	448
6	Asp	176	Cys	
7	Cys	22[c]	Gly	365
8	Ile	337	Leu	368
9	Gln	242	Val	461
10	HSer	85[d]	Gly	285
11			Phe	303
12			Cys	7[c]
13			Glu	158
14			Glu	108
15				

[a] This sample was taken from fraction A of the C-18 chromatography of fragment 2. [b] A low value because of an incomplete drying step. [c] Detected as PTH-cystine. [d] PTH-homoserine has the same retention time as PTH-Thr.

Table II: Amino Acid Composition[a] of CNBr Fragments 2 and 4

amino acid	fragment 2[b] res 1–10 + 66–79		fragment 4[b] res 11–43 + 47–61	
	obsd	theoret	obsd	theoret
Cys[c]	3.0	4	1.8	2
Asp	2.9	3	5.9	6
Glu	4.7	5	7.1	7
Ser	0.7		2.8	3
Gly	3.4	3	3.0	2
His	0.1		1.0	1
Arg	0.2		2.1	2
Thr	0.2		2.4	3
Ala	0.3		4.2	4
Pro	1.0	1	1.1	1
Tyr	0.0		0.0	(2)[a]
Val	1.8	2	3.6	4
Met	0.1		0.2	
Ile	1.8	2	2.9	3
Leu	1.2	1	2.9	3
Phe	0.9	1	0.9	1
Lys	0.9	1	1.9	2
HSer[d]	0.9	1	2.2	2

[a] The samples were performic acid oxidized prior to HCl hydrolysis. Tyrosine is destroyed with oxidation. [b] Both samples were taken from fraction A of the corresponding C-18 chromatography separations. [c] Cysteine was determined as cysteic acid. [d] Homoserine is derived from Met after CNBr cleavage.

Fragment 2. The sequence data from fragment 2a (Table I) was best interpreted as resulting from the simultaneous sequencing of equimolar quantities of Cs-Act 1–10 joined through two disulfide bonds to a 14 amino acid peptide. The sequence of the 14 amino acid peptide, inferred by subtraction, did not match any part of the known sequence of pig kidney CS-Act. This new peptide was designated CS-Act 66–79 and accounted for the C-terminal region of the protein. The abrupt falloff in sequencing response following cycles 10 and 14 suggested that complete sequences were obtained for both peptides. This was apparently due to minimal washout from the glass filter attributable to the hydrophilic nature of the C-terminal homoserine and glutamic acid residues. Most importantly, neither cycle 4 nor 6 yielded any significant quantity of a second PTH-amino acid, and cycles 7 and 12 yielded 22 and 7 pmol of PTH-cystine, respectively.

The amino acid composition of fragment 2b (Table II) showed reasonable agreement with the theoretical composition of CS-Act 1–10 + 66–79 (determined from the sequence data).

Table III: Mass Spectrometric Analysis of CNBr Fragments 2 and 4

mol wt obsd	mol wt calcd	identity[d] (peptide res)	rel signal int
		Fragment 2[a]	
2596.8[b] 2596[c]	2596.9	1–10 + 66–79 2 disulfides, 1 lactone	100
2617.6[b]	2614.9	1–10 + 66–79 2 disulfides, 0 lactone	20
2579.9[b] 2579[c]	2578.9	1–10 + 66–79 2 disulfides, 1 lactone, 1 additional water loss	7
		Fragment 4[e]	
6186.5[b]	6185.7	11–43 + 47–61 1 disulfide, 2 lactones + 5 CHO	100
6204.8[b]	6203.7	11–43 + 47–61 1 disulfide, 1 lactone + 5 CHO	40
6223.8[b]	6221.7	11–43 + 47–61 1 disulfide, 0 lactone + 5 CHO	10

[a] This sample was taken from fraction B of the C-18 chromatography of fragment 2. [b] Data obtained with electrospray ionization mass spectrometry. [c] Data obtained with FAB mass spectrometry. [d] The lactone group(s) are postulated to arise from water loss from the C-terminal homoserine residues. The 5 sugar residues (CHO) are postulated as *N*-acetylglucosamine (2), mannose (2), and fucose (1). See the text for the justification of these assignments. [e] This sample was taken from fraction A of the C-18 chromatography of fragment 4.

The observed homoserine residue could be from Met-10. This analysis by performic acid oxidation gave 3.0 residues of cysteic acid, which was less than the 4 residues of cysteine expected for 2 disulfide bonds.

Confirmatory evidence that the two peptide segments were covalently attached was supplied by the mass spectrometric measurements (Table III), in which the major signals from ESIMS transformed into a mass value of 2596.8 Da, very close to the calculated mass value of 2596.9 Da for CS-Act 1–10 + 66–79 in the lactone form (presumably derived from dehydration of the C-terminal homoserine residue) with two disulfide bridges. Less intense signals in the mass spectral data agreed after transformation with calculated mass values of the same structure with no lactone (2617.6 Da) and with the same structure with one lactone and an additional loss of water (2579.9 Da). There were no detectable signals in the mass range of the individual peptide segments.

Fragment 4. The amino acid composition of fragment 4A (Table II) was found to be reasonably consistent with that of CS-Act 11–43 + 47–61. The analysis showed 1.8 residues of cysteine, representing Cys-36 and Cys-47, and 2.2 residues of homoserine, corresponding to Met-43 and Met-61 of the CS-Act protein. When fragment 4B was subjected to 23 cycles of Edman degradation, the results (Table IV) were best explained as primarily resulting from simultaneous sequencing of equimolar quantities of CS-Act 11–43 and 47–61 covalently attached through a single disulfide bond. A minor sequence evident in this analysis appeared to be composed of Cs-Act 44–61, which was detected at about 20% of the level of the major peptides. Segment 47–61 was apparantly completely sequenced. Segment 44–61 was sequenced to the penultimate residue; the response for the last predicted residue (homoserine) was too weak for confident assignment. The existence of this minor peptide (44–61) suggested incomplete CNBr cleavage at Met-46. Within the partial sequence of segment 11–43 was the absence of any detectable quantity of a PTH-amino acid at cycle 11 (carbohydrate-linked Asn-21).

Confirmatory evidence that the two peptide segments CS-Act 11–43 and 47–61 were covalently attached was provided by mass spectrometry, which showed the predominant signal

Table IV: Sequence Analysis of Fragment 4[a]

cycle no.	peptide 11–43 amino acid	pmol	peptide 47–61 amino acid	pmol	peptide 44–61 amino acid	pmol
1	Val	323	Cys		Ala	60
2	Thr	201	Lys	177	Asp	58
3	Asp	102	Asn	54	Met	32
4	Leu	215	Tyr	125	Cys	
5	Gln	171	Ile	168	Lys	30
6	Asn	197	Ser	92	Asn	197
7	Ala	203	Gln	94	Tyr	34 (+21)
8	Val	168	Tyr	101	Ile	38 (+21)
9	Arg	100	Ser	77	Ser	77
10	Thr	143	Glu	53	Gln	20 (+7)
11	Asn[b]	0	Ile	43	Tyr	18 (+7)
12	Ser	76	Ala	44	Ser	76
13	Thr	91	Ile	39	Glu	14 (+14)
14	Phe	61	Gln	18	Ile	21 (−18)
15	Val	55	HSer[c]	20	Ala	15 (+5)
16	Glu	45			Ile	10 (+1)
17	Ala	56			Gln	5 (+1)
18	Leu	44			HSer[c]	0
19	Val	41				
20	Asn	43				
21	His	19				
22	Ala	38				
23	Lys	16				

[a] This sample was taken from fraction B of the C-18 chromatography of fragment 4. Values in parentheses represent the net change of residue from the previous cycle. [b] The absence of PTH-Asn indicates N-linked carbohydrate on residue 21. [c] PTH-homoserine has the same retention time as PTH-Thr.

Table V: Mass Spectrometric Analysis of CNBr Fragments of CS-Act[a]

CNBr fragment peak	mol wt measd	mol wt calcd	possible identity[b]
1	2877	2865.3	1–10 + 64–79, 1 lactone, 2 disulfides
4	6490.6	6485.1	11–43 + 44–61, 5 CHO, 2 lactones, 1 additional water loss, 1 disulfide
4	6508.6	6503.1	11–43 + 44–61, 5 CHO, 2 lactones, 1 disulfide
5	5654	5658.2	11–43 + 47–61, 2 CHO, 2 lactones, 1 disulfide
6	6024	6023.6	11–43 + 47–61, 4 CHO, 2 lactones, 1 disulfide
6	5879	5879.4	11–43 + 47–61, 3 CHO, 1 lactone, 1 disulfide
		5877.4	11–43 + 47–61, 3 CHO, 2 lactones, 1 disulfide

[a] Data is presented for signals additional to those reported in Table III. These samples were purified by a single reversed-phase HPLC (C4) step. Fractions 1, 5, and 6 were analyzed by FAB. Fraction 4 was measured by electrospray ionization. [b] The sugar residues (CHO) used in the calculations are as explained in Table VI except that 3 sugar residues represents 2 *N*-acetylglucosamines + 1 fucose (calculated molecular weight 5879.4) or alternatively 2 *N*-acetylglucosamines + 1 mannose (calculated molecular weight 5877.4).

from ESIMS transformed into a mass value of 6186.5 Da, very close to the calculated mass of 6185.7 Da for CS-Act 11–43 + 47–61 with one disulfide bond, two lactones (presumably derived from dehydration of the C-terminal homoserine residues), and five sugar residues (Table III; an explanation for the assignment of the sugar residues is given later). Relatively minor signals in the mass spectral data corresponded to the same peptide and carbohydrate components but with different amounts of lactone formation. This interpretation of the variable amounts of homoserine lactone formed upon cleavage of the methionine residues offers an explanation for at least some of the microheterogeneity evident in rechromatography of fragments 2, 3, and 4. There were no detectable signals in the mass range of the individual peptides.

Mass Spectrometric Analysis of CNBr Fragments 1, 3, and 5–8. This analysis was then pursued to try to determine the identity of minor fragments and to investigate the possibility

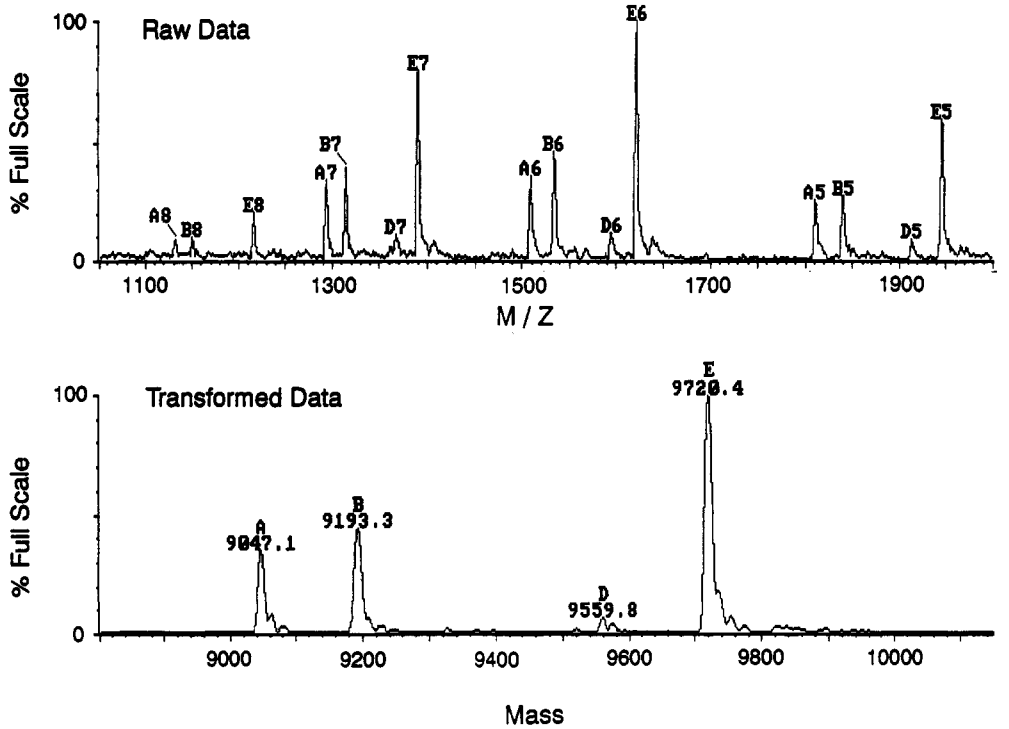

FIGURE 3: Top: Electrospray mass spectrum of purified porcine CS-Act (abscissa represented as mass/charge ratio) in which the peaks are identified according to the series to which they belong (A, B, D, and E) and their charge state (+8, +7, +6, and +5). Bottom: The same data following computerized transformation so that the abscissa is represented in units of mass, with the four principal components identified as A, B, D, and E with their corresponding mass values.

Table VI: Mass Spectrometric Analysis of Native Porcine CS-Act

measd mol wt							
ESI[a]				mean measd	calcd	identity[b]	rel
expt 1	expt 2	expt 3	FAB	ESI mol wt	mol wt	(peptide res)	signal int
9720.4	9720.7	9720.4	9724	9720.5 ± 0.2	9721.0	1–79 + 5 CHO	100
9193.3	9193.1	9192.4	9194	9192.9 ± 0.3	9193.5	1–79 + 2 CHO	42
9047.1	9048.0	9047.7	9043	9047.6 ± 0.3	9047.4	1–79 + 1 CHO	37
9559.8	9558.1	9559.2	9553	9559.0 ± 0.5	9558.9	1–79 + 4 CHO	8

[a] Three separate measurements on the same sample were made by electrospray ionization (ESI). [b] The sugar residues (CHO) are 5 CHO, *N*-acetylglucosamine (2), mannose (2), and fucose (1); 4 CHO, *N*-acetylglucosamine (2), mannose (1), and fucose (1); 2 CHO, *N*-acetylglucosamine and fucose; and 1 CHO, *N*-acetylglucosamine. See the text for the justification of these assignments.

of locating any different disulfide bond arrangements. Table V lists the mass values for some of the additional signals observed with their possible identity. Signals corresponding to more than one molecule were evident in all the spectra, and some fragments had as many as four clear signals. On the basis of variable lactone formation, variable sugar length, and/or variable peptide length due to incomplete CNBr peptide cleavage, a probable identity could be assigned to all of the observable mass values. An example of incomplete CNBr peptide cleavage was the observation in fragment 3 of a mass value corresponding to the minor sequence CS-Act 44–61 + 11–43 with a single disulfide bond. Mass analyses of fragments 5–8 all indicated the presence of molecules that had shortened oligosaccharide chains.

Mass Spectrometric Analysis of Native CS-Act. With native CS-Act both FAB and ESI gave relatively intense signals which were in close agreement with one another and verified the presence of several different molecules in the preparation. For example, the ESI mass spectrum for native CS-Act gave signals for several molecules, each having multiple charged states, with the strongest signal corresponding to a molecule of mass 9720.4 Da (Figure 3). Data from several determinations (Table VI) showed the predominant species in the preparation had a mean mass value as determined by ESIMS of 9720.5 ± 0.2 Da (±SEM, three separate determinations). This measurement was in excellent agreement with the calculated mass of CS-Act 1–79 with three internal disulfide bonds and five covalently attached sugar residues (two *N*-acetylhexosamines, two hexoses, and one deoxyhexose) of 9721.0 Da. As explained in the Discussion, it is most likely that these sugars are two *N*-acetylglucosamines, two mannoses, and one fucose and for simplicity will be referred to as such. The other less intense signals corresponded to the same molecule with two (*N*-acetylglucosamine and fucose), one (*N*-acetylglucosamine), and four carbohydrate residues (two *N*-acetylglucosamines, one mannose, and one fucose), respectively.

In addition, ESIMS revealed a weak signal at 9822.9 Da, which agrees reasonably well with the calculated mass for CS-Act 1–80 with three internal disulfide bonds and five sugar residues (two *N*-acetylglucosamines, two mannoses, and one fucose; 9821.0 Da) with an additonal valine at position 80. The assignment of valine to position 80 is made on the basis of the sequence of human CS-Act deduced from the human cDNA and because about 20% of isolated human liver CS-Act protein contains valine at position 80 (Furst et al., 1990). In the pig kidney protein the importance of this component

Table VII: Mass Spectrometric Analysis[a] of Reduced[b] and Native Porcine CS-Act

identity[c] (peptide residues)	measd mol wt		reduced minus native	rel signal int[d]
	reduced	native		
1–79 + 5 CHO	9725.9	9720.4	5.5	100
1–79 + 2 CHO	9199.6	9193.3	6.3	35
1–79 + 1 CHO	9052.9	9047.1	5.8	25
1–79 + 4 CHO	9564.0	9559.8	4.2	5

[a] Electrospray ionization. The measured values for the native form are the same data as that presented in experiment 1, Table VI. [b] CS activator was reduced by dithiothreitol as described in Experimental Procedures. [c] The assignments of the sugar residues used in the calculations are explained in Table VI. [d] Mean of the relative intensities of the signals in the spectra of the reduced and native forms.

appears to be minimal because on the basis of the mass spectrometric data it is present at less than the 1% level, and it could not be detected in any of the sequence analyses.

The CS-Act protein has six cysteine residues located from the N-terminus at positions 4, 7, 36, 47, 71, and 77. The molecular weight measurements on the various CNBr fragments were all in agreement with the native molecule being in the fully oxidized state. Additional proof of this was provided by mass measuring reduced CS-Act. The ESIMS spectrum of the dithiothreitol-reduced protein produced four signals in the same relative intensities as the native protein shown in Figure 3. Mass values of the three predominant species in the reduced preparation increased an average of 5.9 ± 0.2 Da (mean ± SEM, n = 3) mass units (Table VII), which is in good agreement with six cysteine residues forming from their respective disulfide bonds.

DISCUSSION

We have previously identified methionine residues in pig kidney CS-Act at positions 10, 43, 46, 61, and 63 (Fluharty et al., 1992a). Assuming homology to the human activator, methionine residues were also expected at positions 62 and 65. Thus, in the fully reduced state complete CNBr cleavage of the pig kidney material was predicted to result in the formation of six peptides, four of which would be greater than four amino acid residues in length (res 1–10, 11–43, 47–61, and 66–end), while in the fully oxidized state the number of peptide products resulting from complete CNBr cleavage would be fewer, depending on the number and positions of the intrachain disulfide bonds. The unexpected complexity of the HPLC profile of the CNBr digestion mixture can be attributed to a variety of factors, including variable lactone formation, variable sugar length, variable CNBr fragment length due to incomplete cleavage, and a quantitatively minor 80 amino acid residue component in the preparation. Despite this complexity, determination of the various components of the digestion was possible, using a combination of sequence, amino acid, and mass spectrometric data. In general, excellent agreement was obtained between all data sets, allowing the virtually unequivocal determination of the structure of the preparation and assignment of the disulfide linkages as shown in Figure 4.

The sequence analysis of peptide segment 66–79 established for the first time the sequence of the amino acids in the C-terminal portion of pig kidney CS-Act. In this 14 amino acid region there were three conservative changes from the human CS-Act molecule at residues 69 (Glu to Asp), 72 (Ala to Gly), and 78 (Asp to Glu). Also, the sequence and mass spectrometric data for peptide segment 47–61 established for the first time the identity of positions 55 (Ser) and 60 (Gln)

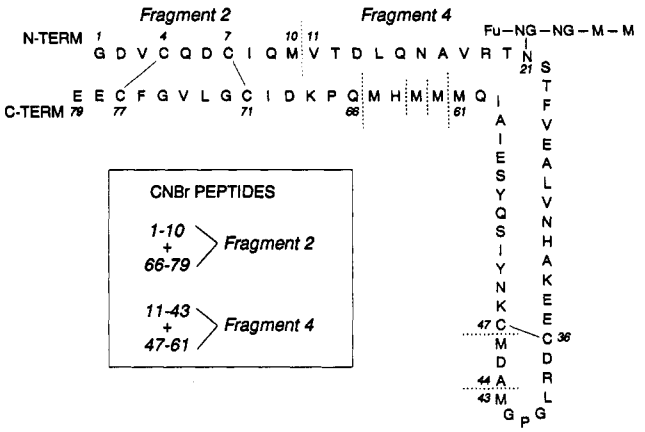

FIGURE 4: Linear representation of the amino acid sequence of porcine CS-Act showing the disulfide bridge assignments and the location and probable structure of the carbohydrate moiety (Fu = fucose, NG = *N*-acetylglucosamine, and M = mannose). The dotted lines denote CNBr cleavage positions.

and confirmed the previously uncertain assignment of Met at residues 62 and 65.

The complete pig kidney CS-Act protein sequence differs from the human kidney CS-Act protein sequence in nine positions. Six of these changes had been noted earlier (Fluharty et al., 1992a), and nearly all are conservative replacements. These changes have little influence on the highly amphipathic nature of the postulated α-helices in the human protein molecule (O'Brien et al., 1988). Human, rat, and pig kidney CS-Act contain the unique methionine-rich sequence QMMMHMQ, spanning residues 60–66 in the pig sequence. This sequence is not found in other saposin sequences or in any other known proteins and may represent a critical "signature" for this protein. The similarity between the rat and pig sequences and that of the human CS-Act at positions other than the methionine-rich sequence is 77% and 87%, respectively.

All known saposin sphingolipid activator proteins have at least one potential glycosylation site and presumably contain an N-glycosylated moiety (O'Brien et al., 1991). The absence of sequence data for peptide 11–43 at cycle 11 confirms that residue 21 is a likely site of glycosylation in the pig kidney molecule. The mass measurement data on CNBr fragments 4 and 6 and on the native protein indicate that isoforms containing one, two, three, four, and five sugar residues were present in the preparation and that the five-sugar isoform appeared to be the most abundant form. The justification for assigning *N*-acetylglucosamine, mannose, and fucose as the sugar residues in our structural model is based on multiple criteria. Firstly, the differences in measured molecular weights between the four most intense signals from native CS-Act (161.5, 527.6, and 673.1 Da, for the difference between the five- and the four-, two-, and one-carbohydrate-unit-containing molecules, respectively, Table VI) was recognized as most likely being due to one hexose (residue mass 162.1 Da), one acetylhexosamine + two hexoses (combined residue mass 527.5 Da), and two acetylhexosamines + two hexoses + one deoxyhexose (combined residue mass 673.6 Da), respectively (B. N. Green, unpublished observations). Secondly, the presence of glucosamine, mannose, and fucose had been previously observed in the chemical analysis of CS-Act (Fluharty et al., 1992a). Thirdly, although the exact linkage of these sugar residues is not established, the obtained mass values are consistent with the idea that this carbohydrate chain is of the consensus complex type Asn-GlcNAc(Fuc)-GlcNAc-Man-Man (Kornfeld & Kornfeld, 1985).

It is worthwhile to compare our proposed pattern of glycosylation with that determined by Yamashita et al. (1990) for the normal human liver protein. Roughly half the molecules were fucosylated. Mono- and di-GluNac substituents made up roughly 20% of the total, but there were no fucosylated analogues of these substituents. In contrast only about 20% of the porcine kidney material was not fucosylated, and all of this was the mono-GluNac derivative. There was no evidence of di-GluNac or di-GluNac(Fuc) substituents in the porcine preparation examined here. The human material contained substantial amounts of mono-, di- and trimannosyl oligosaccharides, with fucosylated forms being about 2-fold more frequent in each case. Permethylation data suggested that the dimannosyl components were linked α1–6 rather than α1–3. The most frequent structure (35–40%) was Manα1–6Manβ1–4GluNacβ1–4GluNac(Fucα1–6), which also appears to be the most common glycosyl residue in the porcine kidney cerebroside sulfate activator. However, only a small amount of monomannosyl and no trimannosyl substituents were present. All of the mannosyl substituents were fucosylated.

The importance of glycosylation for the biochemical activity of CS-Act is obscure. Although enzymatically deglycosylated human CS-Act is fully active in vitro (A. L. Fluharty, unpublished observations), a mutation that changes threonine-23 to isoleucine, altering the recognition site for glycosylation at asparagine-21, is responsible for a deficiency of active CS-Act in a variant form of metachromatic leukodystrophy (Kretz et al., 1990). In our experience only a small fraction of the activity in crude pig kidney homogenate is bound by concanavalin A (Fluharty et al., 1992a), and the purified CS-Act protein used in the experiments reported herein is not retained by this lectin despite the 2 mol of mannose/mol of activator. The significance of the isoforms with variable carbohydrate content found in this study is unclear, as is the greater question of the importance of glycosylation for in vivo activity.

The sequence data on the CNBr fragments allowed us to identify the three disulfide bonds in pig kidney CS-Act protein as Cys4–Cys77, Cys7–Cys71, and Cys36–Cys47. In addition, the mass spectrometric data confirmed that the native molecule contained three internal disulfide linkages and that peptide fragments 1–10 and 66–79, and 11–43 and 47–61, were covalently attached to one another. The evidence for identification of the two disulfide bonds connecting the N-terminal portion of the molecule to the C-terminal portion was the absence of detectable quantities of PTH-cystine at Edman degradation cycles 4 and 6 and detection of PTH-cystine at Edman degradation cycles 7 and 12 of fragment 2 (Table II). Our rationale is based on the concept that cystine in one peptide segment is not released by Edman degradation until its other half is also released (Lu et al., 1986). The only combination of linkages that will explain all the data is 4–77 and 7–71. PTH-cystine is less stable than most other PTH-amino acids, thus accounting for the lower recovery of PTH-cystine at cycles 7 and 12 than were obtained for the PTH-amino acids in nearby cycles. In addition, the low recovery of PTH-cystine at cycles 7 and 12 was consistent with recoveries of PTH-cystine in similar cycles in recently reported studies of disulfide determination of human von Willebrand factor (Marti et al., 1987). The Cys36–47 linkage is the only one available for the covalent attachment of the 11–43 and 47–61 peptide segments. In analyzing the sequence and mass spectrometric data for the other CNBr fragments, with the single exception referred to below, we found no evidence for any disulfide bond

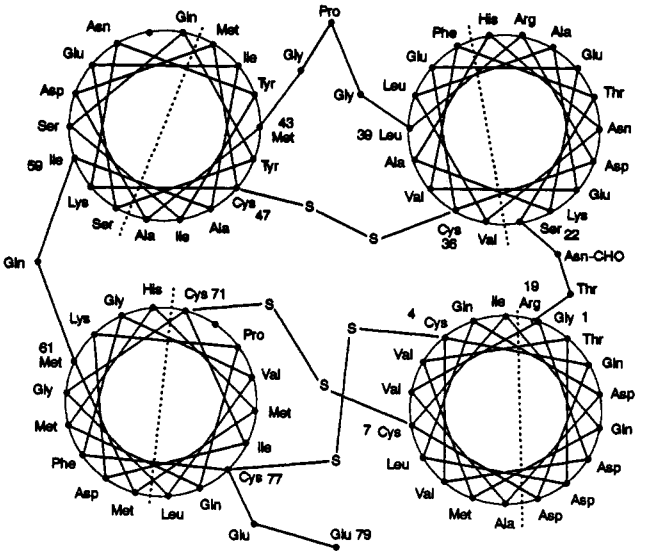

FIGURE 5: Model of four-α-helical bundle of CS-Act showing a proposed arrangement of the amphipathic helical segments incorporating the newly established disulfide bridges. The preponderance of hydrophobic amino acid residues are located in the central core, and the hydrophilic amino acid residues are on the outside.

arrangements other than those proposed for fragments 2 and 4.

One set of sequence data for fragment 3 was best explained as resulting from the simultaneous sequencing of equimolar quantities of three peptide segments. Actually this is the expected result for the trihelical peptide configuration originally proposed for this protein. However, this result could not be confirmed in a repeat experiment and was not supported by the mass measurements. Therefore, if this alternate disulfide pairing scheme is also present, it must be of relatively minor significance. The new information on the placement of the three disulfide bonds in pig kidney CS-Act has necessitated the generation of a new model (Figure 5) for this activator protein.

The agreement between the measured molecular weights and those calculated on the basis of the structural model deserves some comment. Typically, ESI measurements are generally within ±0.01%, and FAB measurements are usually within ±1 Da, of the molecular weight. With ESIMS, differences between measured and calculated masses of $<\pm$0.01% of the molecular weight were found for eight of nine measurements on the native molecule (the single exception had a difference of 0.012%), but for only two of six measurements on the CNBr fragments. With FABMS, differences between the measured and calculated masses of ±1 Da were found for one of four measurements on the native preparation and for three of seven measurements on the CNBr fragments. In general we noticed that the weaker signals were usually those that give the greatest differences between measured and expected values, indicating perhaps that the inaccuracy with which the signals were centroided may have contributed to a greater than expected error. In addition, we cannot dismiss the possibility of the presence of unresolved minor components which influenced the centroiding and mass measurement processes; again, the effect of such components would be greatest with the weaker signals.

Previously a structural model for saposin activator proteins, including CS-Act, was proposed which consisted of three antiparallel α-helices stabilized by internal disulfide cross-linking (O'Brien et al., 1988). This model was proposed after Wynn suggested that CS-Act might have a helical bundle

with a lipid binding center (Wynn, 1986). One feature of the three-α-helical model was that it provided for a convenient disulfide cross-linking pattern stabilizing a compact molecular structure. However, this particular disulfide cross-linking scheme could not be substantiated by the results reported here. We now propose a new monomer model for pig CS-Act protein which includes the new C-terminal sequence information and the identification of the three disulfide bonds (Figure 5). This model has four α-helices and, like the previous model, each proposed α-helical segment of 18 amino acids has a nearly ideal amphipathic structure. In this four-α-helical bundle the hydrophobic faces of each α-helical segment point toward each other while the highly hydrophilic faces form the exterior of the molecule. This arrangement accomodates two properties of CS-Act, namely, the water solubility of the protein and the ability of the protein to bind water-insoluble cerebroside sulfate. The current disulfide cross-linking pattern can be accomodated by this model and can be used to stabilize the interaction of the helices. In a recent de novo design of a synthetic 79 amino acid protein, a disulfide bond was engineered into the molecule for the specific purpose of bringing the α-helices into juxtaposition to more easily form a four-α-helical bundle (Hecht et al., 1990). A highly compact structure of CS-Act, such as that proposed here, is consistent with the protein exhibiting an unusually low molecular weight by gel filtration, resistance to protease digestion, and a high stability to heat inactivation (Fluharty et al., 1992a). By the secondary structure prediction method of GOR (Garnier et al., 1978) the CS-Act showed 72% α-helix formation, and circular dichroism studies indicated that the predominant secondary structure is indeed α-helical (O'Brien & Kishimoto, 1991; A. Waring, unpublished observations). This information is reasonably compelling, but it must be realized that our model is only an attempt to rationalize certain properties of CS-Act and fails to take into account other important properties such as the apparant dimeric nature of the active protein. However, it does provide a useful hypothetical model until NMR and/or X-ray crystallography can fully define the tertiary and quaternary structures.

CS-Act (saposin B) has sequence similarity to other proteins, most notably the other activator molecules, saposins A, C, and D (O'Brien & Kishimoto, 1991). The sequence similarity of CS-Act with saposins A, C, and D ranges from 47% to 51%, whereas the sequence identity ranges from 16% to 25%. There is also strong correlation of alignment of the six cysteine residues in the saposin molecules as well as the possibility of an N-linked carbohydrate in each. The saposin activators are all derived from the same precursor, prosaposin, and the cleaved products, saposins A, B, C, and D, are most likely the result of gene duplication. All of this evidence points to the possibility that the three other saposin molecules could have disulfide bridges analogous to CS-Act and that they could also have similar tertiary structures.

The identity/similarity between human saposins can also be extended to include human pulmonary surfactant protein (SP-B), as has been previously noted (Sano et al., 1988; O'Brien & Kishimoto, 1991; Patthy, 1991). There is a 25% identity and a 45% similarity (O'Brien & Kishimoto, 1991) between CS-Act and SP-B, which also, coincidently, has 79 amino acids. A comparison of these two sequences shows that six cysteine residues are aligned at the same positions with only minor gap changes. Recently the three intrachain disulfide bridges in porcine surfactant protein B were identified (Johansson et al., 1991). In this molecule these disulfide placements link cysteines at residues 8 and 77, 11 and 71, and 35 and 46. This disulfide arrangement is analogous to that found by us for CS-Act. That is, the N-terminal end is linked to the C-terminal end with two disulfide bridges, and the two centrally located cysteines are linked together. Thus, the sequence and disulfide similarities of CS-Act and SP-B open up the possibility that these two functionally diverse molecules may have similar overall structures. However, dissimilar characteristics of the molecules would need to be accomodated in the model. For instance, CS-Act is completely soluble in water, whereas SP-B is soluble only in an organic environment. Also, SP-B has a very high proportion of hydrophobic amino acids and, unlike CS-Act, has no linked carbohydrate. Future biophysical studies on CS-Act and SP-B will increase our understanding of the possible relationship between these two small proteins.

Lastly, CS-Act has a very high degree of sequence similarity to certain domains of sulfated glycoprotein 1 secreted by rat Sertoli cells (Collard et al., 1988). It has been suggested that this protein is also a precursor of CS-Act. This idea is supported by the very high degree of identity (77%) between human CS-Act and domain 2 of sulfated glycoprotein 1 and the presence of the unique QMMMHMQ segment in each of them. However, CS-Act is of lysosomal origin, and sulfated glycoprotein 1 is derived from secretory cells which are thought to support the process of spermatogenesis; how the two are related is not yet clear.

ACKNOWLEDGMENT

The authors wish to thank Takeo Sakuma from Sciex (Thornhill, Canada) and Ken Johnson from Finnigan MAT (San Jose, CA) for analysis of selected samples. Dr. Alan Waring contributed to the discussion of the biophysical comparison of CS-Act and SP-B.

REFERENCES

Cohen, S. A., & Strydom, D. J. (1988) *Anal. Biochem. 174*, 1–16.

Collard, M. W., Sylvester, S. R., Tsuruta, J. K., & Griswold, M. D. (1988) *Biochemistry 27*, 4557–4564.

Dewji, N., Wenger, D., Fujibayashi, S., Donoviel, M., Esch, F., Hill, F., & O'Brien, J. S. (1986) *Biochem. Biophys. Res. Commun. 134*, 989–994.

Dewji, N., Wenger, D., & O'Brien, J. S. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8652–8656.

Fluharty, A. L., Katona, Z., Meek, W. E., Frei, K., & Fowler, A. V. (1992a) *Biochem. Med. Metab. Biol. 47*, 66–85.

Fluharty, A. L., Meek, W. E., Katona, Z., & Tsay, K. K. (1992b) *Biochem. Med. Metab. Biol. 47*, 86–96.

Fujibayashi, S., & Wenger, D. A. (1986) *Biochim. Biophys. Acta 875*, 554–562.

Furst, W., Machleidt, W., & Sandhoff, K. (1988) *Biol. Chem. Hoppe-Seyler 369*, 317–328.

Furst, W., Schubert, J., Machleidt, W., Meyer, H. E., & Sandhoff, K. (1990) *Eur. J. Biochem. 192*, 709–714.

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

Hecht, M. H., Richardson, J. S., Richardson, D. C., & Ogden, R. O. (1990) *Science 249*, 884–891.

Holtschmidt, H., Sandhoff, K., Furst, W., Kwon, H. Y., Schnabel, D., & Suzuki, K. (1991) *FEBS Lett. 280*, 267–270.

Inui, K., & Wenger, D. A. (1984) *Arch. Biochem. Biophys. 233*, 556–564.

Inui, K., Kau, F.-T., Fujibayashi, S., Jones, C., Morse, H. G., Law, M. L., & Wenger, D. A. (1985) *Hum. Genet. 69*, 197–200.

Johansson, J., Curstedt, T., & Jornvall, H. (1991) *Biochemistry 30*, 6917–6921.

Kornfeld, R., & Kornfeld, S. (1985) *Annu. Rev. Biochem. 54*, 631–664.

Kretz, K. A., Carson, G. S., Morimoto, S., Kishimoto, Y., Fluharty, A. L., & O'Brien, J. S. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2541–2544.

Li, S.-C., Kihara, H., Serizawa, S., Li, Y.-T., Fluharty, A. L., Mayes, J. S., & Shapiro, L. J. (1985) *J. Biol. Chem. 260*, 1867–1871.

Li, S.-C., Sonnino, S., Tettamanti, G., & Li, Y.-T. (1988) *J. Biol. Chem. 263*, 6588–6591.

Lu, H. S., Klein, M. L., Everett, R. R., & Lai, P.-H. (1986) in *Modern Methods in Protein Chemistry* (L'Italien, J., Ed.) pp 493–501, Plenum Press, New York.

Marti, T., Rosselet, S. J., Titani, K., & Walsh, K. A. (1987) *Biochemistry 26*, 8099–8109.

Mehl, E., & Jatzkewitz, H. (1964) *Hoppe-Seyler's Z. Physiol. Chem. 339*, 260–276.

Nakano, T., Sandhoff, K., Stumper, J., Christomanou, H., & Suzuki, K. (1989) *J. Biochem. 105*, 152–154.

O'Brien, J. S., & Kishimoto, Y. (1991) *FASEB J. 5*, 301–308.

O'Brien, J. S., Kretz, K. A., Dewji, N., Wenger, D. A., Esch, F., & Fluharty, A. L. (1988) *Science 241*, 1098–1101.

Patthy, L. (1991) *J. Biol. Chem. 266*, 6035–6037.

Sandhoff, K. (1984) in *Molecular Basis of Lysosomal Storage Disorders* (Barranger, J. A., & Brady, R. O., Eds.) pp 19–49, Academic Press, New York.

Sano, A., Radin, N. S., Johnson, L. L., & Tarr, G. E. (1988) *J. Biol. Chem. 263*, 19597–19601.

Shapiro, L. J., Alec, K. A., Kaback, M. M., Itabashi, H., Desnick, R. J., Brand, N., Stevens, R. L., Fluharty, A. L., & Kihara, H. (1979) *Pediatr. Res. 13*, 1179–1181.

Stevens, R. L., Fluharty, A. L., Kihara, H., Kaback, M. M., Shapiro, L. J., Marsh, B., Sandhoff, K., & Fischer, G. (1981) *Am. J. Hum. Genet. 33*, 900–906.

Wynn, C. (1986) *Biochem. J. 240*, 921–924.

Yamashita, K., Inui, K., Totani, K., Kochibe, N., Furukawa, M., & Okada, S. (1990) *Biochemistry 29*, 3030–3039.