

Isolation and sequencing of a cDNA clone encoding 107 kDa sialoglycoprotein in rat liver lysosomal membranes

Masaru Himeno, Youichiro Noguchi, Hiroyuki Sasaki⁺, Yoshitaka Tanaka, Koji Furuno, Akira Kono*, Yoshiyuki Sakaki⁺ and Keitaro Kato

Division of Physiological Chemistry, Faculty of Pharmaceutical Sciences, Kyushu University, Higashi-ku, Fukuoka,

*⁺Research Laboratory for Genetic Information, Kyushu University, Higashi-ku, Fukuoka and *National Kyushu Cancer Center, Minami-ku, Fukuoka, Japan*

Received 19 November 1988; revised version received 3 January 1989

A cDNA for 107 kDa sialoglycoprotein (LGP 107), the major protein component of rat liver lysosomal membranes, was isolated and sequenced. The 1.8 kbp cDNA contained an open reading frame encoding a polypeptide consisting of 386 amino acid residues (M_r 41 914). The deduced NH₂-terminal 10-residue sequence is identical with that determined for purified LGP 107. The primary structure deduced for LGP 107 contains 20 potential *N*-glycosylation sites and exhibits 82.5, 43 and 60% sequence similarities to mouse LAMP-1, chicken LEP 100, and a 120-kDa human lysosomal glycoprotein, respectively. Among these lysosomal glycoproteins, the amino acid sequence of the putative transmembrane segment is highly conserved. Northern blot hybridization analysis identified a single species of LGP 107 mRNA (2.1 kbp in length) in rat liver, kidney, brain, lung, spleen, heart and pancreas, although its level in pancreas was very low.

Sialoglycoprotein, 107-kDa; Lysosomal membrane; cDNA cloning; Amino acid sequence

1. INTRODUCTION

The lysosomal membrane proteins play a pivotal role in the functions of lysosomes by separating the lysosomal matrix from cytosolic materials. Lysosomal membrane proteins are thought to be involved in such important roles as lysosomal fusion with autophagic vacuoles and endosomes. Recently, lysosomal glycoproteins have been purified from mouse [1,2], rat [3–6], chicken [7] and human [8] cells. They are all highly sialylated and have low isoelectric points, suggesting that one of their functions is to maintain an acidic intralysosomal environment. These proteins have

been shown to occur not only in lysosomes but also in endosomes and plasma membranes [4–6]. This fact is suggestive of their involvement in facilitating the vesicular transport that is responsible for transport of materials into lysosomes for digestion. Despite such speculations, their precise functions are still unknown.

Here we report the isolation and sequencing of a cDNA clone encoding a 107-kDa sialoglycoprotein (LGP 107) in the rat liver lysosomal membranes. The primary structure predicted for rat liver LGP 107 exhibits obvious similarities to those of mouse, human, and chicken lysosomal membrane sialoglycoproteins [8–10].

2. MATERIALS AND METHODS

2.1. Purification of LGP 107 from rat liver lysosomal membranes

Major sialoglycoprotein (420 kDa) was purified from rat liver lysosomal membranes as described previously [3,11]. When analyzed by SDS-PAGE, the purified protein was

Correspondence address: K. Kato, Division of Physiological Chemistry, Faculty of Pharmaceutical Sciences, Kyushu University, Higashi-Ku, Fukuoka 812, Japan

Abbreviations: LGP 107, 107-kDa sialoglycoprotein; PAGE, polyacrylamide gel electrophoresis

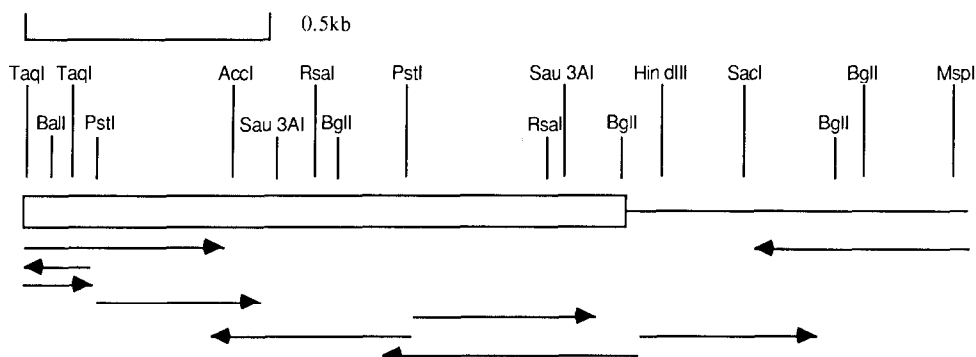


Fig.1. Restriction map of LGP 107 cDNA and the strategy adapted for nucleotide sequence determination. Arrows indicate the direction and extent of sequencing runs. The open box indicates the amino acid coding region and the line indicates the 3'-noncoding region.

	GC	GCC	CCA	GCA	CTG	TTC	GAG	GTG	AAA	GAC	AAC	AAC	GGC	ACA	GCG	TGT
		Ala	Pro	Ala	Leu	Phe	Glu	Val	Lys	Asp	Asn	Asn	Gly	Thr	Ala	Cys
												*				#
48	ATA	ATG	GCC	AGC	TTC	TCT	GCC	TCC	TTT	CTG	ACC	ACC	TAT	GAT	GCT	GGA
16	Ile	Met	Ala	Ser	Phe	Ser	Ala	Ser	Phe	Leu	Thr	Thr	Tyr	Asp	Ala	Gly
96	CAT	GTT	TCT	AAG	GTC	TCG	AAT	ATG	ACC	CTG	CCA	GCC	TCT	GCA	GAA	GTC
32	His	Val	Ser	Lys	Val	Ser	Asn	Met	Thr	Leu	Pro	Ala	Ser	Ala	Glu	Val
							*									
144	CTG	AAG	AAT	AGC	AGC	TCT	TGT	GGT	GAA	AAG	AAT	GCT	TCT	GAG	CCC	ACC
48	Leu	Lys	Asn	Ser	Ser	Ser	Cys	Gly	Glu	Lys	Asn	Ala	Ser	Glu	Pro	Thr
			*				#				*					
192	CTC	GCA	ATC	ACC	TTT	GGA	GAA	GGA	TAT	TTA	CTG	AAA	CTC	ACC	TTC	ACA
64	Leu	Ala	Ile	Thr	Phe	Gly	Glu	Gly	Tyr	Leu	Leu	Lys	Leu	Thr	Phe	Thr
240	AAA	AAC	ACA	ACA	CGT	TAC	AGT	GTC	CAG	CAC	ATG	TAT	TTC	ACA	TAT	AAC
80	Lys	Asn	Thr	Thr	Arg	Tyr	Ser	Val	Gln	His	Met	Tyr	Phe	Thr	Tyr	Asn
		*														*
288	CTG	TCA	GAC	ACA	CAA	TTC	TTT	CCC	AAT	GCC	AGC	TCC	AAA	GGG	CCC	GAC
96	Leu	Ser	Asp	Thr	Gln	Phe	Phe	Pro	Asn	Ala	Ser	Ser	Lys	Gly	Pro	Asp
									*							
336	ACT	GTG	GAT	TCC	ACA	ACT	GAC	ATC	AAG	GCA	GAC	ATC	AAC	AAA	ACA	TAC
112	Thr	Val	Asp	Ser	Thr	Thr	Asp	Ile	Lys	Ala	Asp	Ile	Asn	Lys	Thr	Tyr
													*			
384	CGA	TGT	GTC	AGC	GAC	ATC	AGG	GTC	TAC	ATG	AAG	AAT	GTG	ACC	ATT	GTG
128	Arg	Cys	Val	Ser	Asp	Ile	Arg	Val	Tyr	Met	Lys	Asn	Val	Thr	Ile	Val
		#										*				
432	CTC	TGG	GAC	GCT	ACT	ATC	CAG	GCC	TAC	CTG	CCG	AGT	AGC	AAC	TTC	AGC
144	Leu	Trp	Asp	Ala	Thr	Ile	Gln	Ala	Tyr	Leu	Pro	Ser	Ser	Asn	Phe	Ser
														*		
480	AAG	GAA	GAG	ACA	CGC	TGC	CCA	CAG	GAT	CAA	CCT	TCC	CCA	ACT	ACT	GGG
160	Lys	Glu	Glu	Thr	Arg	Cys	Pro	Gln	Asp	Gln	Pro	Ser	Pro	Thr	Thr	Gly
						#										
528	CCA	CCC	AGC	CCC	TCA	CCA	CCA	CTT	GTG	CCC	ACA	AAC	CCC	AGT	GTG	TCC
176	Pro	Pro	Ser	Pro	Ser	Pro	Pro	Leu	Val	Pro	Thr	Asn	Pro	Ser	Val	Ser
												*				
576	AAG	TAC	AAT	GTG	ACT	GGT	GAC	AAT	GGA	ACC	TGC	CTG	CTG	GCC	TCT	ATG
192	Lys	Tyr	Asn	Val	Thr	Gly	Asp	Asn	Gly	Thr	Cys	Leu	Leu	Ala	Ser	Met
		*						*			#					
624	GCA	CTG	CAA	CTC	AAC	ATC	ACC	TAC	ATG	AAG	AAG	GAC	AAC	ACG	ACT	GTG
208	Ala	Leu	Gln	Leu	Asn	Ile	Thr	Tyr	Met	Lys	Lys	Asp	Asn	Thr	Thr	Val
					*								*			
672	ACC	AGA	GCA	TTC	AAC	ATC	AAC	CCA	AGT	GAC	AAA	TAT	AGT	GGG	ACT	TGC
224	Thr	Arg	Ala	Phe	Asn	Ile	Asn	Pro	Ser	Asp	Lys	Tyr	Ser	Gly	Thr	Cys
							*									#

Fig.2. The nucleotide sequence and deduced amino acid sequence of LGP 107 cDNA. The amino acid sequence obtained from protein sequencing is indicated by the box. Putative polyadenylation signals are enclosed with the box. The sequences corresponding to the (serine-proline)-rich region (dashed line) and the putative transmembrane portion (solid line) are indicated. Cysteine residues, potential *N*-glycosylation sites and the termination codon are indicated by #, * and ---, respectively.

separated into two polypeptide bands having molecular masses of 107 and 96 kDa. The two polypeptides, separated by preparative SDS-PAGE, were extracted from the gel and used as purified preparations. Antibodies were raised in rabbits against LGP 107 and affinity purified on a Sepharose 6B column conjugated with purified LGP 107.

2.2. Protein sequence analysis

The NH₂-terminal sequence of purified LGP 107 was determined in an Applied Biosystem 470A protein sequencer/Spectra Physics SP8100 HPLC system.

2.3. Screening of the cDNA library

A rat liver cDNA expression library [12] in λ gt11 was screened with specific rabbit antibodies raised against LGP 107 using horseradish peroxidase-conjugated anti-rabbit IgG as the second antibodies, as described previously [13]. Positive plaques were isolated and bacteriophage DNAs were digested with *Eco*RI. The *Eco*RI-excised cDNA inserts were subcloned into plasmid vector pUC118 [14], and characterized by restriction-endonuclease mapping.

2.4. DNA sequencing

Nucleotide sequences were analyzed by the dideoxynucleotide chain termination method [15], in which denatured plasmid DNA was used as a template as described [16].

2.5. Northern-blot analysis

Twenty μ g each of total RNA prepared from various rat tissues by the method of Chirgwin et al. [17] was denatured with 1 M glyoxal and subjected to electrophoresis on 1.1% (w/v) agarose gel as described previously [18]. The RNAs thus separated were transferred to a nylon membrane (Zeta-probe, Bio-Rad, USA) and hybridized with ³²P-labeled LGP 107 cDNA, which had been prepared with Multiprime DNA labeling system (Amersham International, England) as described [19].

3. RESULTS AND DISCUSSION

A rat liver cDNA library constructed with λ gt11 as vector [12] was screened with anti-LGP 107 an-

```

720  GGT GCC CAG TTG GTG ACC CTG AAG GTG GGG AAC AAG AGC AGA GTC CTG
240  Gly Ala Gln Leu Val Thr Leu Lys Val Gly Asn Lys Ser Arg Val Leu
      *
768  GAG CTG CAG TTT GGG ATG AAT GCC ACT TCT AGC CTG TTT TTC CTG CAA
256  Glu Leu Gln Phe Gly Met Asn Ala Thr Ser Ser Leu Phe Phe Leu Gln
      *
816  GGA GTT CAG TTG AAC ATG ACT CTT CCT GAT GCC ATA GAG CCC ACG TTC
272  Gly Val Gln Leu Asn Met Thr Leu Pro Asp Ala Ile Glu Pro Thr Phe
      *
864  AGC ACC TCC AAC TAT TCC CTG AAA GCT CTT CAG GCC AGT GTC GGC AAC
288  Ser Thr Ser Asn Tyr Ser Leu Lys Ala Leu Gln Ala Ser Val Gly Asn
      *
912  TCA TAC AAG TGC AAC AGT GAG GAG CAC ATC TTT GTC AGC AAG GCG CTC
304  Ser Tyr Lys Cys Asn Ser Glu Glu His Ile Phe Val Ser Lys Ala Leu
      *
960  GCC CTC AAT GTC TTC AGC GTG CAA GTC CAG GCT TTC AGG GTA GAA AGT
320  Ala Leu Asn Val Phe Ser Val Gln Val Gln Ala Phe Arg Val Glu Ser

1008 GAC AGG TTT GGG TCT GTG GAA GAG TGT GTA CAG GAC GGT AAC AAC ATG
336  Asp Arg Phe Gly Ser Val Glu Glu Cys Val Gln Asp Gly Asn Asn Met
      *
1056 CTG ATC CCC ATT GCT GTG GGC GGG GCC CTG GCA GGG CTG GTC CTC ATC
352  Leu Ile Pro Ile Ala Val Gly Gly Ala Leu Ala Gly Leu Val Leu Ile

1104 GTC CTC ATC GCC TAC CTC ATC GGC AGG AAG AGG AGT CAC GCG GGC TAT
368  Val Leu Ile Ala Tyr Leu Ile Gly Arg Lys Arg Ser His Ala Gly Tyr

1152 CAG ACC ATC TAG CCTGGTGGGCAGGTGCGCCACAGACGCACGGGCCTGTTCTCACAT
384  Gln Thr Ile ---

1211 CCCCAGCTTAGATAGGTGTGGAAGGGAGGGAGGCACACTTGTGGCAAAGTGTTCAAATCT
1274 GCTTTATCAAAATGTGAAGCTCATCTTGCACATTTACTATGCACAAAGGAATACTATTGAAA
1337 TGACGGTGTTAATTTTGCTAACTGGGTAAATATTTTGCTAACTGGTTAAATGTTAATATGTT
1400 ACCAAAGTAGAGCTCTAAAGAGGACAAAGAGGCTCCACGCATTTGACTTTTAAGACTTGGTGT
1463 TTGGTCTTCTATCTTTTACTCAGATTTACGCCTTACAAAGGGAATCTCTGGTCCAGACACTT
1526 TGGCTGGCAAGGGTGGCTGATGGTGGTTAGGCTGCACACTTGAGAAGCAAAACAGGAGCAGGG
1589 ACGTCTGCCACACAGGCACGCACAGGGTCAGCCTCTGGACACTTGGCTTGGGCTGCCTGGCC
1652 TGGGGGGCTAGACTCTGGCATCTGGCTGGGCACACCCAGTGTTCGTGCTCTGCCGCT
1715 GTGAGCTACCACTTTCCTAAATAGAAAATGGCAATTATGGGGTTGGGGTTTGGCTCAGTGSTA
1778 GAGCCCTTGCCTAGGAAGCGCAAGGCCCTGGGTTTCGGTCCCCAGCTCCGGAAAAA
1841 AAAAAAAAAAAAAA

```

tibodies, resulting in the isolation of three positive clones from approximately 2.1×10^5 phages. The inserts of these clones, termed λ a, λ b, and λ c (insert lengths, 1.8, 1.5 and 1.8 kbp, respectively), were subcloned into the plasmid vector pUC118 and analyzed by restriction mapping. Fig.1 shows the restriction map of the insert of λ a, which was

the longest among the three clones, as well as the strategy adopted for its sequencing. The nucleotide sequence determined for the λ a insert and the primary structure deduced therefrom are shown in fig.2.

As can be seen, this cDNA lacked the 5'-untranslated region, initiation codon and a cer-

	1	10	20	30	40	50
LGP 107	APALFEVKD	NNGTACIM	ASFSAFL	TTYDAGH	VSKVSNM	TLPASAEVLKN
	:	:	:	:	:	:
mLAMP-1	---LFEVKNN	-GTCIMAS	FSASFLL	TYETANG	SQIVNISL	PASAEVLKN
	:	:	:	:	:	:
	60	70	80	90	100	
LGP 107	SSSCGEKN	ASEPTLAI	TFGEGYLL	KLTFKN	TTRYSVQH	MYFTYNLSDTQ
	:	:	:	:	:	:
mLAMP-1	GSSCGKEN	VSDPSLT	ITFGRGYLL	TLNFTKN	TTRYSVQH	MYFTYNLSDE
	:	:	:	:	:	:
	110	120	130	140	150	
LGP 107	FFPNASSK	GPDVTVD	STTDIKAD	INKTYRC	VSDIRVY	MKNVTIVLWDATIQ
	:	:	:	:	:	:
mLAMP-1	HFPNAISKE	IYTMDST	TDIKADINK	AYRCVSD	IRVYMK	NVTIVLVRDATIQ
	:	:	:	:	:	:
	160	170	180	190	200	
LGP 107	AYLPSSNF	SKEETRC	PQDQPS	PTTGPP	SPSPPL	VPTNPSVSKYNVTGDNG
	:	:	:	:	:	:
mLAMP-1	AYLSSGNF	SKEETHC	TQDGP	STTGPP	SPSPPL	VPTNPTVSKYNVTGNG
	:	:	:	:	:	:
	210	220	230	240	250	
LGP 107	TCLLASMA	LQLNITY	MKKDNT	TVTRAF	NINPSDK	YSGTCGAQLVTLKVG
	:	:	:	:	:	:
mLAMP-1	TCLLASMA	LQLNITY	LKKDNK	TVTRAF	NISPND	TSSGSGCINLVT
	:	:	:	:	:	:
	260	270	280	290	300	
LGP 107	KSRVLELQ	FGMNAT	SSLFFL	QGVQLN	MTLPDA	IEPTFSTSNYSLKALQAS
	:	:	:	:	:	:
mLAMP-1	KNRALELQ	FGMNASS	SLFFLQ	GVRLNM	TLPDAL	VPTFSISNHSLKALQAT
	:	:	:	:	:	:
	310	320	330	340	350	
LGP 107	VGNSYKCN	SEEHIFV	SKALAL	NVFSVQ	VQAFR	VESDRFGSVEECVQDGNN
	:	:	:	:	:	:
mLAMP-1	VGNSYKCN	TEEHIFV	SKMLSL	NVFSVQ	VQAFK	VDSDRFGSVEECVQDGNN
	:	:	:	:	:	:
	360	370	380			
LGP 107	MLIPIAVG	GALAGLV	LIVLIA	YLI	GRKRSH	HAGYQTI
	:	:	:	:	:	:
mLAMP-1	MLIPIAVG	GALAGLV	LIVLIA	YLI	GRKRSH	HAGYQTI

Fig.3. Amino acid sequence similarity between LGP 107 and mouse LAMP-1 [9]. The LGP 107 (top) and mouse LAMP-1 sequences are represented in the standard one-letter code. Identical residues are indicated by (:), gaps by (-).

tain length of the 5'-coding segment. To obtain information concerning the missing 5'-coding segment (and therefore the NH₂-terminal amino acid sequence), we determined the NH₂-terminal sequence of purified LGP 107. The sequence thus determined, Ala-Pro-Ala-Leu-Phe-Glu-Val-Lys-Asp-Asn, was found to be identical with the NH₂-terminal sequence deduced from the nucleotide sequence (see fig.2). It was thus evident that the cDNA insert of λ a codes for the entire length of mature LGP 107 but lacks the segment encoding a NH₂-terminal, cleavable signal peptide. The molecular mass predicted for mature LGP 107, 41.9 kDa, is lower by about 3 kDa than the value (45 kDa) estimated for the immunoprecipitate from the products of cell-free translation directed by rat liver poly(A)⁺ RNA (not shown). This is consistent with the fact that LGP 107 synthesized in vitro still retains the signal peptide. The termination codon TAG (nucleotides 1161–1163) was followed by the 694-nucleotide 3'-noncoding region and a poly(A) tail. As is the case for lysosomal membrane glycoproteins from other sources, the primary structure of rat LGP 107 contains a potential hinge region (residues 169–191) that is rich in proline and serine and thought to separate the polypeptide into two domains. The primary structure of LGP 107 also contains 20 potential asparagine-linked glycosylation sites and 8 cysteine residues at the same spacings as those of mouse LAMP-1 [9] (see fig.2). It is clear that the much higher molecular mass estimated for purified LGP 107 by SDS-PAGE (107 kDa) than that predicted here (41.9 kDa) is due to the presence of sugar chains in the former. Like mouse LAMP-1, LGP 107 possesses a 24-residue strongly hydrophobic stretch (residues 351–386) near the COOH-terminus (see fig.2). It is highly likely that this stretch serves as a transmembrane segment and, if so, it is expected that the remaining short COOH-terminal portion (residue 375 to the COOH-terminus) is extruded to the cytoplasm. The secondary structure pattern of LGP 107, predicted by the algorithms of Garnier et al. [20] and Kyte and Doolittle [21], is almost the same as that obtained for mouse LAMP-1 [9] (not shown).

Although no typical polyadenylation signal (AATAAA) was found in the 3'-noncoding region we sequenced, it is likely that either one of the two DNA segments enclosed with the box (ATTAT or

ATTTA in fig.2) may serve as the polyadenylation signal instead of the sequence of AATAAA.

The deduced amino acid sequence of LGP 107 is 82.5% similar to that of mouse LAMP-1 [9]. LGP 107 also exhibits 60 and 40% sequence similarities with a 120-kDa human lysosomal membrane glycoprotein [8] and chicken LEP 100 [7], respectively. If the mismatches between chemically similar amino acid residues can be assumed to be 'identical', then the sequence similarity between LGP 107 and mouse LAMP-1 reaches more than 90% (see fig.3). Despite such a high sequence similarity, the LGP 107 polypeptide is longer than that of mouse LAMP-1 by four amino acid residues. Three of them are located at the NH₂-terminus of LGP 107 and the remaining one between residues 7 and 8 of the mouse LAMP-1 sequence (see fig.3). In any case, it appears that rat LGP 107 is functionally identical with mouse

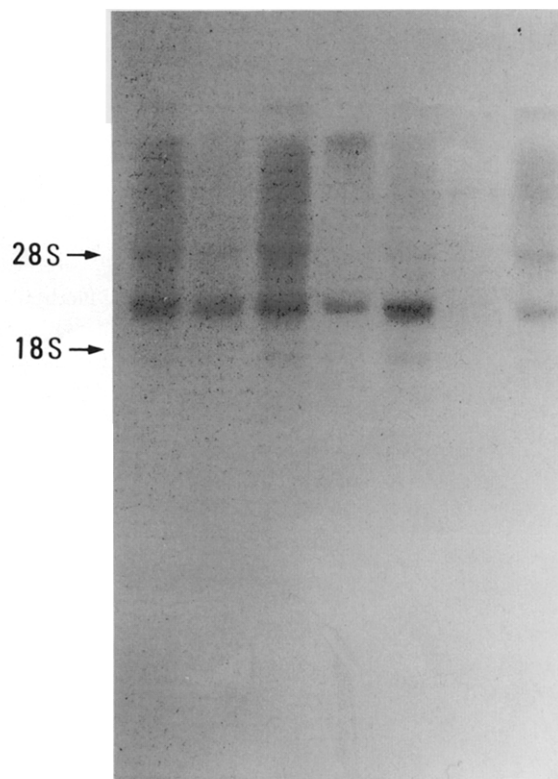


Fig.4. RNA blot hybridization analysis. RNA size markers are 28 S (5.5 kbp) and 18 S (2.1 kbp) rRNAs. Lanes from left to right represent samples from brain, heart, lung, liver, kidney, pancreas, and spleen, respectively.

LAMP-1 and that both proteins belong to the same gene family.

Fig.4 shows the results of Northern blot analysis of total RNAs from various rat tissues with ^{32}P -labeled LGP 107 cDNA as a probe. A single species of mRNA of about 2.1 kbp long was detected in all the tissues examined, i.e. liver, kidney, brain, lung, heart, spleen, and pancreas, although its level in pancreas was very low. The apparent low content of LGP 107 mRNA in pancreas was probably due to rapid degradation of the mRNA by pancreatic RNase. It therefore appears that the LGP 107 gene is expressed in many rat tissues. The size of the mRNA (2.1 kbp) is larger than that of the cDNA isolated in this study (1.8 kbp). This is consistent with the fact that the isolated cDNA lacked the 5'-noncoding region and the coding region for the signal peptide.

Acknowledgements: We wish to thank Dr Hirohiko Hojo for instruction in DNA sequencing, Dr Hirokazu Furuya for instruction in computer analysis and Dr Yukio Ikehara and Dr Yoshio Misumi for providing the rat liver cDNA library. This study was supported in part by a grant-in-aid from the Ministry of Education, Science and Culture of Japan.

REFERENCES

- [1] Chen, J.W., Pan, W., D'Souza, M.P. and August, J.T. (1985) *Arch. Biochem. Biophys.* 239, 574-586.
- [2] D'Souza, M.P. and August, J.T. (1986) *Arch. Biochem. Biophys.* 249, 522-532.
- [3] Kato, K. (1985) in: 3rd International Congress on Cell Biology (Abstracts) Academic Press, Inc., Tokyo, Japan.
- [4] Lewis, V., Green, S.A., Marsh, M., Vinko, P., Helenius, A. and Mellman, I. (1985) *J. Cell Biol.* 100, 1839-1847.
- [5] Barriocanal, J.G., Bonifacino, J.S., Yuan, L. and Sandoval, I.V. (1986) *J. Biol. Chem.* 261, 16755-16763.
- [6] Reggio, H., Bainton, D., Harms, R. and Louvard, D. (1984) *J. Cell Biol.* 99, 1511-1526.
- [7] Lippincott-Schwartz, J. and Fambrough, D.M. (1986) *J. Cell Biol.* 102, 1593-1605.
- [8] Viitala, J., Carlsson, S.R., Siebert, P.D. and Fukuda, M. (1988) *Proc. Natl. Acad. Sci. USA* 85, 3743-3747.
- [9] Chen, J.W., Ying Cha, K., Yuksel, U., Gracy, R.W. and August, J.T. (1988) *J. Biol. Chem.* 263, 8754-8758.
- [10] Fambrough, D.M., Takeyasu, K., Lippincott-Schwartz, J. and Siegel, N.R. (1988) *J. Cell Biol.* 106, 61-67.
- [11] Ohsumi, Y., Ishikawa, T. and Kato, K. (1983) *J. Biochem.* 93, 547-556.
- [12] Misumi, Y., Tashiro, K., Hattori, M., Sakaki, Y. and Ikehara, Y. (1988) *Biochem. J.* 249, 661-668.
- [13] Huynh, T., Young, R.A. and Davis, R.W. (1985) in: *DNA Cloning: A Practical Approach 1* (Glover, D.M. ed.) pp.49-78, IRL Press, Oxford.
- [14] Norrander, J., Kempe, T. and Messing, J. (1983) *Gene* 26, 101-106.
- [15] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
- [16] Hattori, M. and Sakaki, Y. (1986) *Anal. Biochem.* 152, 232-238.
- [17] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry* 18, 5294-5299.
- [18] Thomas, P.S. (1983) *Methods Enzymol.* 100, 255-259.
- [19] Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.* 132, 6-13.
- [20] Garneir, J., Osguthorpe, D.J. and Robson, B. (1978) *J. Mol. Biol.* 120, 97-120.
- [21] Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105-132.