Sequence of Human Asialoglycoprotein Receptor cDNA

AN INTERNAL SIGNAL SEQUENCE FOR MEMBRANE INSERTION*

(Received for publication, October 29, 1984)

Martin Spiess§, Alan L. Schwartz¶, and Harvey F. Lodish
From the Whitehead Institute for Biomedical Research and
Department of Biology, Massachusetts Institute of
Technology, Cambridge, Massachusetts 02142 and the
§Division of Pediatric Hematology/Oncology, Children’s
Hospital Medical Center, Dana-Farber Cancer Institute
and Department of Pediatrics, Harvard Medical School,
Boston, Massachusetts 02115

A cDNA library from the human hepatoma cell line Hep G2 was prepared in the expression vector λ gt11. Using specific antibodies, a cDNA clone containing the entire coding sequence for the human asialoglycoprotein receptor was isolated and sequenced. The deduced amino acid sequence of 291 residues is very homologous to the sequence of the major asialoglycoprotein receptor protein from rat. The comparison shows that there is no significant post-translational processing and no leader sequence, cleaved or uncleaved, at the amino terminus. An internal signal sequence, probably the membrane-spanning segment, residues 41–59, is assumed to direct insertion of the carboxyl-terminal ligand binding portion of the receptor across the endoplasmic reticulum membrane.

Partially deglycosylated plasma glycoproteins are efficiently and specifically removed from the circulation by a receptor-mediated process (1). In mammals the asialoglycoprotein receptor (ASGP-R), specific for desialylated (galactosyl-terminal) glycoproteins, is expressed exclusively in hepatic parenchymal cells. Following ligand binding to this cell surface receptor, the receptor-ligand complex is internalized and transported by a series of membrane vesicles and tubules to an acidic sorting organelle (CURL) where receptor and ligand dissociate (2). The receptor returns to the cell surface, while the ligand is transported to lysosomes where it is degraded (3–5). The ASGP-R from rabbit, rat, and human liver have been purified (6–8). The complete amino acid sequences of the major rat receptor protein, R-1 (9), and of an analogous, chicken N-acetylglucosamine-specific receptor (10) have been determined, as has a partial sequence of a second rat ASGP-R, R-2/3 (9). Many of the functional studies, however, describing the kinetics of ligand binding, internalization, and recycling of the receptor, as well as its biosynthesis, have been performed on the human hepatoma cell line Hep G2 (4, 11, 12). Taking advantage of this model system for receptor-mediated endocytosis, we have cloned and sequenced the cDNA encoding the human ASGP-R.

EXPERIMENTAL PROCEDURES

Materials—Reverse transcriptase was purchased from Life Sciences, Inc., all other enzymes from New England Biolabs and Bethesda Research Laboratories. DNA sequencing reagents were from P-L Biochemicals and sodium [125I]iodide, deoxyctydylate 5'-α[32P] triphosphate, and deoxyadenosine α-5'-[32S]thiotriphosphate were from Amersham. Rat α-tubulin DNA was a gift from Dr. I. R. Lemischka, Massachusetts Institute of Technology.

cDNA Libraries—The maintenance of the human hepatoma cell line Hep G2 has been described earlier (11). Poly(A)⁺ RNA was isolated by the guanidinium isothiocyanate/CsCl gradient method (13) and by chromatography on oligo(dT)-cellulose (14). cDNA was synthesized, essentially following the procedures described in Ref. 14, using oligo(dT) as a primer for reverse transcription. After second-strand synthesis using the Klenow fragment of DNA polymerase I, the cDNA was treated with S1 nuclease, filled-in using Klenow fragment, blunt-end ligated to synthetic EcoRI linkers, and digested with EcoRI. After removal of excess linkers by gel filtration on Sephadex G-100 the cDNA was ligated into EcoRI-cut and phosphatase-treated λ gt11 DNA (15). For the preparation of a second cDNA library with longer inserts the second cDNA strand was synthesized using RNase H and DNA polymerase I as described in Ref. 16. Consequently, S1 nuclease digestion and the fill-in reaction were omitted. Furthermore, endogenous EcoRI sites were modified by

FIG. 1. Northern analysis of ASGP-R RNA. Poly(A)⁺ RNA (20 µg) from HeLa and Hep G2 cells was fractionated on a 1.5% agarose gel containing formaldehyde, transferred to nitrocellulose, and probed (34) with the nick-translated cDNA clone A21 and, as a control, with the rat α-tubulin gene (mRNA size approximately 1800 bp (35)). Hybridization was at 37 °C for 48 h. The positions of the 18 S and 28 S ribosomal RNA are indicated.

A12
Tubulin

-28 S

-18 S
EcoRI methylese before ligation to the linkers. Finally, the cDNA was fractionated by electrophoresis on an agarose gel, and the fraction >600 base pairs (bp) was electroeluted and ligated into the vector DNA. After in vitro packaging of the recombinant DNA into λ particles (14), the resulting libraries were amplified and screened as described in Ref. 17 using an affinity purified polyclonal rabbit anti-human ASGP-R antibody (12) and 125I-iodinated protein A. Approximately 1 out of every 100,000 recombinant phage scored positive, consistent with the expected frequency of ASGP-R cDNA inserted in the correct orientation and reading frame. The majority of these clones hybridized to one another. Clone A12, at 450 bp the largest, was used to probe blots of gels of poly(A)+ RNA from Hep G2 cells, and also from HeLa cells which do not express the ASGP-R (Fig. 1). A single mRNA of 1500-1600 bp, sufficient to encode the ASGP-R, was detected. This RNA was absent from HeLa cells, as expected. A rat α-tubulin clone hybridized to a single 1800-bp species in both RNA preparations, a result demonstrating the integrity of the RNA.

Using A12 as a probe, a second λ gt11 library with longer cDNA inserts (see “Experimental Procedures”) was screened by plaque hybridization, and several clones larger than 1 kilobase were thus isolated. The largest one, A21, was entirely sequenced and contained the full coding sequence of the ASGP-R (Fig. 2). In Fig. 3 the deduced amino acid sequence is compared to the sequences of the rat ASGP receptors.

The 1277 bases of the cDNA contain only one large open reading frame of 933 bases (bases 60 to 873 in Fig. 2). Beginning at the first ATG methionine codon (position 1) the coding region comprises 873 bases. This is preceded by 172 noncoding bases at the 5'-end and followed by 232 noncoding bases at the 3'-end. Clone A21 does not extend into the poly(A) region at the 3'-end of the mRNA, and we do not know if any noncoding bases are missing from the 5'-end.

**RESULTS AND DISCUSSION**

In contrast to the rat and rabbit receptor, the human ASGP-R migrates on sodium dodecyl sulfate-polyacrylamide gels as a single species of Mr = 46,000. In the presence of tunicamycin an unglycosylated polypeptide of approximately 34,000 is synthesized. In Hep G2 cells approximately 0.01% of the [35S]methionine incorporated into newly synthesized protein was in the ASGP-R (12). Thus a rough estimate for the frequency of the ASGP-R mRNA is 1 in 10,000.

Double-stranded cDNA was synthesized from poly(A)+ RNA from Hep G2 cells and ligated into the bacteriophage expression vector λ gt11 (15). About 2 million recombinant phage lines were obtained. The terminal EcoRI sites, originally synthetic linkers ligated to the cDNA, are indicated by an underlined. Sequence data were analyzed using the computer-assisted method by Staden (22).

**Fig. 2. Sequence of cdNA clone A21 coding for the human ASGP-R.** The nucleotide sequence and the corresponding amino acid sequence of the coding region are shown. The hydrophobic, membrane-spanning region is underlined. The terminal EcoRI sites, originally synthetic linkers ligated to the cDNA, are indicated by broken lines.
The corresponding polypeptide of 291 residues would have a Mr of 33,122 which agrees well with the experimental value of 34,000 for the unglycosylated receptor as determined on SDS-gels (12).

The rat R-1 and the human ASGP receptors are strikingly homologous; 79% of the residues are identical. (For comparison, the human and mouse &globin sequences have 82% identical residues (23).) The rat R-1 sequence lacks the initiator methionine, 6 residues at the C terminus, and 1 amino acid at position 15 in the human sequence. Both receptor sequences consist of two hydrophilic domains (residues 1-40 and 60-291) separated by a very hydrophobic, in all likelihood membrane-spanning, segment (residues 41-59). The homology extends throughout the whole sequence with 75 and 79% identical residues in the N-terminal and the C-terminal domains, respectively. It is noteworthy that the hydrophobic region is also very well conserved, with 16 identical residues out of 19 and 3 conservative replacements (Leu-Phe at residue 42, Leu-Ile at 50, and Cys-Ser at position 57). This might indicate that this segment serves a more specific function than just a hydrophobic membrane anchor (see below). Although only a partial sequence of R-2/3 is available, the human ASGP-R appears more homologous to R-1 than to R-2/3 (Fig. 3).

Of the 3 sites for N-linked glycosylation (Asn-X-Thr/Ser) in the rat R-1 sequence (positions 76, 79, and 147 in Fig. 3) only 2 are conserved in the human sequence (positions 79 and 147). In place of asparagine 76 in the rat R-1 sequence, a threonine is found in the human one. Experiments on the biosynthesis of the human receptor in the presence of inhibitors of glycosylation as well as with endoglycosidases confirm the presence of two N-linked oligosaccharides (12, 24). Since N-linked carbohydrates are invariably found on the extracytoplasmic segments of membrane proteins (25), it was concluded that the carboxyl terminus of the rat receptor faces the outside of the cell and contains the ligand binding site (9). The same reasoning would apply to the human receptor. Furthermore, the human ASGP-R is phosphorylated at a serine residue whose position, however, is not known. Since phosphorylation is a cytoplasmic activity, this observation and the finding that the N-terminal domain of the chicken hepatic lectin is phosphorylated at Ser 7 (26) strongly suggest that the amino terminus of the ASGP-R faces the cytoplasm. Thus the transmembrane orientation of the ASGP-R, carboxyl terminus outside, amino terminus inside, would be opposite to that of such transmembrane proteins as the vesicular stomatitis virus G protein, glycophorin, or the class I MHC proteins (27).

Our most significant finding is that the comparison of the coding sequence of the human ASGP-R with the sequence of

---

Schwartz, A. L. (1985) Biochem. J., in press.
the mature rat protein indicates that there is no significant post-translational proteolytic processing either at the amino- or carboxyl terminus. There is clearly no hydrophobic leader sequence, cleaved or otherwise, at the N terminus such as the kind described for a large number of other membrane or secretory proteins (28), including the low-density lipoprotein receptor (29, 30). These sequences initiate membrane insertion of the nascent polypeptide. Unlike these proteins, the amino terminus of the ASGP-R is not extruded through the membrane. A similar situation obtains for the transferrin receptor (36) and the invariant chain of the HLA-DR protein (31), which have a transmembrane orientation the same as the ASGP-R. In the case of the red blood cell protein Band III, a long (450 residues) amino-terminal hydrophilic segment faces the cytoplasm, and the carboxyl terminus is inserted into the membrane. An internal, presumably hydrophobic sequence functions as a signal for insertion of the nascent protein into the rough endoplasmic reticulum (32). As judged by the synthesis and modifications of its N-linked oligosaccharides, the ASGP-R is also inserted into the rough endoplasmic reticulum and matures through the Golgi en route to the plasma membrane (12). We suggest that the well conserved membrane-spanning sequence might function as an internal signal sequence, and initiate the transport of the carboxyl terminus of the nascent ASGP-R protein across the endoplasmic reticulum membrane. As proposed by Engelman and Steitz (33), this sequence might form half of a helical hairpin hydrophobic enough to insert into and span the membrane, and initiate subsequent extrusion of the C-terminal portion of the polypeptide across the bilayer.

In vitro mutagenesis of the cDNA clone, followed by expression either in transfected cells or in cell-free systems, may define the regions of the protein essential for membrane insertion, ligand binding, and receptor internalization and recycling.

Acknowledgment—We thank Dr. Cécile Wandersman for her help in the early stages of this project.

REFERENCES
1. Ashwell, G., and Harford, J. (1982) Annu. Rev. Biochem. 51, 531–554
2. Geuze, H. J., Slot, J. W., Strous, G. J. A. M., Lodish, H. F., and Schwartz, A. L. (1983) Cell 32, 277–287
3. Hubbard, A. L., Wilson, G., Ashwell, G., and Stukenbrok, H. (1979) J. Cell Biol. 83, 47–64
4. Schwartz, A. L., Fridovich, S. E., and Lodish, H. F. (1982) J. Biol. Chem. 257, 4230–4237
5. Ciechanover, A., Schwartz, A. L., and Lodish, H. F. (1983) Cell 32, 267–275
6. Hudgin, R. L., Price, W. E., Ashwell, G., Stockert, R. J., and Morell, A. G. (1974) J. Biol. Chem. 249, 5336–5434
7. Tanabe, T., Price, W. E., and Ashwell, G. (1979) J. Biol. Chem. 254, 1038–1043
8. Baenziger, J. U., and Maynard, Y. (1980) J. Biol. Chem. 255, 4607–4613
9. Drickamer, K., Mamon, J. F., Binns, G., and Leung, J. O. (1984) J. Biol. Chem. 259, 770–778
10. Drickamer, K. (1981) J. Biol. Chem. 256, 5827–5839
11. Schwartz, A. L., Fridovich, S. E., Knowles, B. B., and Lodish, H. F. (1981) J. Biol. Chem. 256, 8875–8881
12. Schwartz, A. L., and Rup, D. (1983) J. Biol. Chem. 258, 11249–11255
13. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294–5299
14. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY
15. Young, R. A., and Davis, R. W. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 1194–1198
16. Okayama, H., and Berg, P. (1982) Cell 2, 161–170
17. Young, R. A., and Davis, R. W. (1983) Science (Wash. D. C.) 222, 778–782
18. Vieira, J., and Messing, J. (1982) Gene 19, 259–268
19. Guo, L.-H., Yang, R. C. A., and Wu, R. (1983) Nucleic Acids Res. 11, 5521–5540
20. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5468
21. Biggin, M. D., Gibson, T. J., and Houg, G. F. (1984) Nucleic Acids Res. 12, 4731–4751
22. Pasman, G. D. (ed) (1975) Handbook of Biochemistry and Molecular Biology, Proteins, Vol. III, 3rd Ed., CRC Press, Cleveland
23. Breitfeld, P. P., Rup, D., and Schwartz, A. L. (1984) J. Biol. Chem. 259, 10414–10421
24. Rothman, J. E., and Lenard, J. (1977) Science (Wash. D. C.) 195, 743–753
25. Drickamer, K., and Mamon, J. F. (1982) J. Biol. Chem. 257, 15156–15161
26. Lodish, H. F., Braell, W. A., Schwartz, A. L., Strous, G. J. A. M., and Zilberstein, A. (1981) Int. Rev. Cytol. 12, (suppl.) 247–307
27. Krell, G. (1981) Annu. Rev. Biochem. 50, 317–348
28. Schneider, W. J., Slaughter, C. J., Goldstein, J. L., Anderson, R. G. W., Capra, D. J., and Brown, M. S. (1983) J. Cell Biol. 97, 1635–1640
29. Russell, D. W., Schneider, W. J., Yamamoto, T., Laskey, K. L., Brown, M. S., and Goldstein, J. L. (1984) Cell 37, 577–585
30. Strubin, M., Bach, B., and Long, E. O. (1984) EMBO J. 3, 869–872
31. Braell, W. A., and Lodish, H. F. (1982) Cell 28, 23–31
32. Engelman, D. M., and Steitz, T. A. (1981) Cell 23, 411–422
33. Zaker, C., and Lodish, H. F. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 5386–5390
34. Lemischka, I. R., Farmer, S., Rocaniello, V. R., and Sharp, P. A. (1981) J. Mol. Biol. 151, 101–120
35. Schneider, C., Owen, M. J., Banville, D., and Williams, J. G. (1984) Nature (Lond.) 311, 670–678