Cloning of a cDNA encoding the human cation-dependent mannose 6-phosphate-specific receptor

(lysosomal enzyme transport/transmembrane protein/chromosomal assignment)

Regina Pohlmann*, Gerhard Nagel*, Bernhard Schmidt*, Martin Stein[†], Gerhard Lorkowski[†], Christiane Krentler*, Jan Cully[†], Helmut E. Meyer[‡], Karl-Heinz Grzeschik[§], Günther Mersmann[†], Andrej Hasilik[†], and Kurt von Figura*[¶]

*Abteilung für Biochemie II, Universität Göttingen, Gossler Str. 12d, D-3400 Göttingen, Federal Republic of Germany; [†]Physiologisch-Chemisches Institut, Universität Münster, Waldeyerstrasse 15, D-4400 Münster, Federal Republic of Germany; [‡]Institut für Physiologische Chemie, Ruhruniversität Bochum, Universitätsstrasse 150, D-4630 Bochum, Federal Republic of Germany; and [§]Institut für Humangenetik, Universität Münster, Vesaliusweg 12-14, D-4400 Münster, Federal Republic of Germany

Communicated by Elizabeth F. Neufeld, April 20, 1987 (received for review March 11, 1987)

ABSTRACT Complementary DNA clones for the human cation-dependent mannose 6-phosphate-specific receptor have been isolated from a human placenta library in λ gt11. The nucleotide sequence of the 2463-base-pair cDNA insert includes a 145-base-pair 5' untranslated region, an open reading frame of 831 base pairs corresponding to 277 amino acids (M_r = 30,993), and a 1487-base-pair 3' untranslated region. The deduced amino acid sequence is colinear with that determined by amino acid sequencing of the N-terminus peptide (41 residues) and nine tryptic peptides (93 additional residues). The receptor is synthesized as a precursor with a signal peptide of 20 amino acids. The hydrophobicity profile of the receptor indicates a single membrane-spanning domain, which separates an N-terminal region containing five potential Nglycosylation sites from a C-terminal region lacking Nglycosylation sites. Thus the N-terminal $(M_r = 18,299)$ and C-terminal ($M_r \leq 7648$) segments of the mature receptor are assumed to be exposed to the extracytosolic and cytosolic sides of the membrane, respectively. Analysis of a panel of somatic cell (mouse-human) hybrids shows that the gene for the receptor is located on human chromosome 12.

Targeting of newly synthesized lysosomal enzymes in many cell types depends on formation of mannose 6-phosphate residues on lysosomal enzymes and recognition of these residues by specific receptors (Man6PR). Two species of Man6PR have been described. First, a receptor with a subunit M_r of 215,000 has been found in a variety of tissues (for reviews, see refs. 1 and 2). This receptor binds its ligands in a pH-dependent manner (3) and participates in targeting of endogenous lysosomal enzymes and in endocytosis of exogenous lysosomal enzymes (4). A second receptor with a subunit M_r of 46,000 has recently been isolated from bovine liver and murine macrophages (5), and its presence has been demonstrated in several human and rat cell types (31). Binding of ligands to this receptor depends on divalent cations (5) and pH (6). This receptor is used for intracellular targeting but not for endocytosis of lysosomal enzymes (32).

Human M_r 46,000 Man6*P*R is a transmembrane protein that consists of two identical subunits, as concluded from crosslinking experiments (33). The deglycosylated receptor has a M_r of 26,500–28,000 (5), and its susceptibility to proteinases indicates that the cytosolic portion has a M_r of \approx 5000 and contains the C terminus (31). The mannose 6-phosphate binding site probably contains a histidine residue in addition to an essential arginine residue (33). In the present study, we report the cloning and sequencing of a cDNA encoding the

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

entire human M_r 46,000 Man6PR^{||} and the chromosomal assignment of the gene for the M_r 46,000 Man6PR to human chromosome 12.

MATERIALS AND METHODS

Materials. The human hepatoma cDNA library (7) in λ gt11 $(8 \times 10^6$ independent clones) was kindly supplied by J. O'Brien and J. de Wet. For construction of a human liver cDNA library in λ gt11 (1 \times 10⁶ independent clones), cDNA synthesis was random-primed (8), and second strand synthesis was performed according to Gubler and Hoffmann (9). The human placenta cDNA library in $\lambda gt11$ (1 \times 10⁶ independent clones) was from Clontech (Palo Alto, CA). Restriction endonucleases, T4 polynucleotide kinase, Klenow fragment, T4 DNA ligase, and M13-derived 15-mer primer were from Boehringer Mannheim: cloned Molonev murine leukemia virus reverse transcriptase was from Bethesda Research Laboratories; $[\gamma^{-32}P]ATP$ (222 TBq/mmol) was from New England Nuclear; and $[\alpha^{-32}P]dATP$ (111 TBq/mmol) and $dATP[\alpha^{-35}S]$ (>22 TBq/mmol) were from Amersham. Dideoxy- and deoxynucleotide triphosphates were from Pharmacia-PL Biochemicals.

Library Screening. Immunological screening was performed with an affinity-purified polyclonal rabbit anti-human M_r 46,000 Man6PR antibody. Detection of fusion proteins was accomplished by using a modification of the procedure described by de Wet et al. (7). After induction of fusion proteins with isopropyl β -thiogalactoside at 37°C, the nitrocellulose filters (Schleicher & Schüll) were saturated with Blotto (10) for 4 hr at 4°C. Incubation with the affinitypurified polyclonal antibodies was for 14-16 hr at 4°C, and subsequent incubation with alkaline phosphatase-conjugated goat anti-rabbit IgG (1:5000 dilution; Tago, Burlingame, CA) was for 2 hr at 4°C. The staining was performed with 5-bromo-4-chloro-3-indolyl phosphate and nitro blue tetrazolium (11). An EcoRI-HindIII fragment was prepared from the 5' end of the insert P4a and was nick-translated by the method of Feinberg and Vogelstein (12). This probe was used to screen 10⁶ plaque-forming units (pfu) of the placenta library (13).

Oligonucleotides. The partial amino acid sequences Asn-Trp-Ile-Met-Leu and Met-Glu-Gln-Phe-Pro of tryptic peptides of the M_r 46,000 Man6*P*R were used to construct the mixed oligonucleotides 280 (5'ARCATDATCCARTT3') and 281 (5'GGRAAYTGYTCCAT3') (where R = an unspecified

Abbreviations: pfu, plaque-forming units; Man6PR, mannose 6-phosphate-specific receptor(s).

To whom reprint requests should be addressed.

[&]quot;This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg, F.R.G.) (accession no. J02937).



purine nucleoside, Y = an unspecified pyrimidine nucleoside, and D = G, A, or T), which specify all possible combinations of codon choices. The oligonucleotides were purified by polyacrylamide gel electrophoresis and were labeled with $[\gamma^{-32}P]$ ATP using T4 polynucleotide kinase (13).

Oligonucleotide Hybridization. λ gt11 clones were digested with *Eco*RI; fragments were separated by electrophoresis in a 1% agarose gel and were stained with ethidium bromide. Hybridization in the gel was performed according to Kidd (14) under high-stringency conditions.

Sequence Analysis. The inserts were subcloned into M13mp18 and M13mp19 vectors (15) and were sequenced by the Sanger *et al.* dideoxy chain termination method (16) with $dATP[\alpha^{-35}S]$ (17) and the Klenow fragment (16) or cloned Moloney murine leukemia virus reverse transcriptase (18). Primers were either a M13-derived 15-mer or sequence-derived oligonucleotides (17-mer).

RNA Blot Hybridization. Poly(A)⁺ RNA from human liver was isolated by the guanidinium thiocyanate method (19) and

FIG. 1. Analysis of recombinant clones for the M_r 46,000 Man6PR and blot-hybridization analysis of poly(A)⁺ RNA. (A) DNA from clones isolated by immunological screening of a human placenta cDNA library was digested with EcoRI, and the fragments were separated by electrophoresis. The DNA was stained with ethidium bromide and after being transferred to a nitro-cellulose filter was hybridized to ³²P-labeled oligonucleotide 281. Positions and sizes (in kb) of DNA markers are indicated. The identification numbers assigned to the placenta clones are given above the lanes. The insert P4 contains an internal EcoRI site and is cleaved into fragments of 1.2 kb (P4a) and 1.5 kb (P4b). The arrow indicates the P4a fragment that hybridizes with oligonucleotide 280. (B) Human liver mRNA (10 μ g) was hybridized with ³²P-labeled P4a and P4b as indicated above the lanes. The arrow indicates the mRNA species hybridizing with P4a.

selected by chromatography on an oligo(dT)-cellulose column (20). The RNA was denaturated with glyoxal (21) and was fractionated by electrophoresis in a 1% agarose gel. Transfer of the RNA to Hybond N membranes (Amersham) was performed according to the supplier's manual. Hybridization was carried out with nick-translated ³²P-labeled cDNA probes (12) by the method of Thomas (21).

Chromosomal Localization. Mouse-human cell hybrids were generated by fusion of mouse B82, RAG, or A9 cells with human fibroblasts. In the hybrid B82 MS2 1a-14-1, human chromosomes 2, 3, 4, 7, 8, 11, 13, 15, 16, 17, 18, 19, 21, 22, X, and Y were present. The other hybrids—RAG ANLY 1, RAG PI 7-2, RAG PI 5-15, RAG GM194 5-5, RAG GO 4, A9 SU 1-2, A9 1T 2-21-14, RAG GM610 5-23, and RAG GM610 4-5-1—were described previously (22–24). The presence of human chromosomes and chromosome fragments was determined by isozyme and cytogenetic analysis (24). DNA from hybrid cells was digested with EcoRI, separated by electrophoresis in 0.8% agarose, transferred to Biodyne A



FIG. 2. Restriction map and sequencing strategy for the human M_r 46,000 Man6PR clones P4a and P29. Horizontal arrows indicate the direction and extent of sequence determination. Filled arrows refer to sequencing with the M13 primer, and open arrows refer to sequencing with the sequence-derived oligonucleotides.

Biochemistry: Pohlmann et al.

- 14 CGAGO	40 GCGCTA	-13 56666	0 1 A A C G	стсс	120 CCTC	TGAJ	-1' ACT:	GCT	TGGC	100 100	GCC	rcrgo	ao Cooco	cace	-80 CCCT	GCTO	-7 GCCC	70 CCGT(CTCC	-60 Catco	CCAO	-5 G & A G G	GTAT	TCAC	-40 :TGGG	GATT	-3 CTGA	0 GCT	
TTGGCT	-20 TACTCC	AGTTT	10 CCC4	CGAC	ACG) ATG Het	TTC Phe	CCT Pro	10 TTC Phe	TAC Tyr	AGC Ser	20 TGC Cys	TGG Trp	AGG Arg	30 ACT Thr	GGA Gly	CTG Leu	CTA Leu	O CTA Leu	CTA Leu	CTC Leu	50 CTG Leu	GCT Ala	GTC Val	60 GCA Ala	, GTG Val	AGA Ars	G A A G l u	23
70 TCC TC Ser Tr	GG CAG rp Gln	80 ACA Thr	GAA Glu	GAA Glu	90 A A A Lys	ACT Thr	TGC Cys	10 GAC Asp)O TTG Leu	GTA Val	GGA Gly	110 GAA Glu	A A G Lys	GGT Gly	120 A A A Lys	G A G G l u	TCA Ser	13 GAG Glu	0 4 4 A L y s	G A G G l u	TTG Leu	140 GCT Ala	CTA Leu	GTG Val	150 AAG Lys	AGG Arg	CTG Leu	Lys	53
160 CCA CI Pro Le	IG TIT eu Phe	AAT Asn	AAA Lys	AGC Ser	180 TTT Phe	GAG Glu	AGC Ser	ACT Thr	0 GTG Val	GGC Gly	C A G Gl n	200 GGT Gly	TC A Ser	GAC Asp	210 ACA Thr	TAC Tyr	ATC Ile	22 TAC Tyr	20 ATC Ile	TTC Phe	AGG Arg	230 GTG Val	TGC Cys	CGG Arg	240 G A A G 1 u	GCT Ala	GGC Gly	★ AAC Asn	83
250 CAC AC His Tr	CT TCT hr Ser	260 GGG Gly	GCA Ala	GGC Gly	270 CTG Leu	GTG Val	CAA Gln	28 ATC Ile	AAC Asn	AAA Lys	AGT Ser	290 AAT Asn	GGG Gly	AAG Lys	300 G A G G 1 u	ACA Thr	GTG Val	3 GTA Val	0 GGG Gly	AGA Arg	CTC Leu	320* AAC Asn	GAG Glu	ACT Thr	330 CAC His	ATC Ile	TTC Phe		113
340 GGA AC Gly Se	GT AAT	350 TGG Trp	ATC Ile	ATG Met	360 CTG Leu	ATC Ile	TAT Tyr	31 AAA Lys	ro GGG Gly	GGT Gly	GAT Asp	3,80 GAA Glu	TAT Tyr	GAC Asp	390 A A C A s n	C A C His	TGT Cys	40 GGC Gly)Q AAG Lys	GAG Glu	CAG Gln	410 CGT Arg	CGT Arg	GCA <u>Ala</u>	420 GTG Val	GTG Val	ATG Met	ATC Ile	143
430 TCC TC Ser Cy	I GC AAT /s Asn	CGA Arg	CAC His	ACC Thr	450 CTA Leu	GCG Ala	GAC Asp	4 (A A T A s n	0 TTT Phe	A A C A s n	CCT Pro	470 GTG Val	TCT Ser	GAG Glu	480 GAG Glu	CGT Arg	GGC Gly	49 444 Lys	O GTC Val	CAA Gln	GAT Asp	500 TGT Cys	TTC Phe	TAC Tyr	510 CTC Leu	TTT Phe	GAG Glu	ATG Het	173
520 GAT AG Asp Se	GC AGC er Ser	530 CTG Leu	GCC Ala	TGT Cys	540 TCA Ser	CCA Pro	G A G G l u	59 ATC Ile	0 TCC Ser	C A C H 1 S	CTC Leu	560 AGT Ser	GTG Val	GGT Gly	570 TCC Ser	ATC Ile	TTA Leu	58 CTT Leu	GTC Val	ACG Thr	TTT Phe	590 GCA Ala	TC A Ser	CTG Leu	<u>600</u> GTT Val	GCT Ala	GTT Val	TAT Tyr	203
610 GTT GI Val Va	TT GGG	520 GGG Gly	TTC Phe	CTA Leu	630 TAC Tyr	C A G G l n	CGA Arg	6ª CTG Leu	GTA Val	GTG Val	GGA Gly	GCC Ala	AAA Lys	GGA Gly	660 ATG Het	GAG Glu	CAG Gln	67 TTT Phe	70 CCC Pro	CAC His	tta Leu	680 GCC Ala	TTC Phe	TGG Trp	690 CAG Gln	GAT Asp	CTT Leu	GGC Gly	233
700 AAC CI Asn Le	TG GTA eu Val	710 GCA Ala	GAT Asp	GGC Gly	720 TGT Cys	GAC Asp	TTT Phe	7: GTC Val	O TGC Cys	CGT Arg	TCT Ser	740 AAA Lys	CCT Pro	CGA Årg	750 A A T A s n	GTG Val	CCT Pro	76 GCA Ala	GCA Ala	TAT Tyr	CGT Ary	770 GGT Gly	GTG Val	GGG Gly	780 GAT Asp	GAC Asp	C≱G Gln	CTG Leu	263
790 GGG GA Gly G1	KG GAG Lu Glu	BOO TCA Ser	GAA Glu	GAA Glu	810 Agg Arš	G A T A s p	G A C A s p	82 CAT His	20 TTA Leu	TTA Leu	CCA Pro	330 ATG Net	TAG	ATT	840 GCAC	ΓΤΤΑΊ	850 1 A T G 1		GCCT	860 CTTCC	TCA	870 GTCCC) :CCAA	8 1 A C C A	180 . A A G C	таса	890 C & G C	CAGA	277
900 TTTCTC) Caagca(910 STCTC	AACT	9 CCAG	20 TCCC	TCAT	930 CTC/) VCCC.	9 1 7 A C 1	40 ATT(стс	950 TTGC1) ITTCO	AGT	960 TTGC	TTTT	970 GATT1	J Tgca:	ICTT	980 CTCAC	TAG	990 TAAA) ACTGC	10 CTTC:	000 20077	TGTT	1010 CCTT		•
1020 TCTGT1) [TTTTC	1030 TCTAG	AGAG	10 G T A C	40 AGT1	IGTA;	1050 AGTC/		10 1 T A A 1	060 TATAJ	TAG	1070 GGCC1) rgtgj	1	080 CAGA	GGCTI	1090 TTTGC	D CATTO	1 GTCT	1:00 CTTG#	CAT	1110 CAGA/) NGTT <i>i</i>	11 1 C A A 1	20 7 A G G C	ATAT	1130 GGGC		۱
1140 TGGTG1	D TAGCAG	1150 GCTC;	CTGO	11 1000	60 TTTG	TTTT	1170 TTAA) A C A C	11 ATTT	180 TCAC	AAGT	1190 TTTT) G A G A (1 C A C T	200 GGAT	TCTI	1210 TAA1	0 T T A A	1: • • • • •	2 2 0 A A A A 1	GCC	1230 AAGA;) A A C A 7	12 [Tat1	240 [TATA	CAGG	1250 GTTG	ATTO	;
1260 CTTTC	D A TGTTG	1270 TTATI) TCTG1	12 1200	280 2747	AGTA	129 GCCT	0 C C A T	1: G A G A .	300 ATCT	GGTA	131 TTTC	0 TTGC	1 Tgct	320 TGGA	ACTAC	1330 TTTC	GCAG	1 TGAT	340 TACTI	GGT	1350 TGC A 0) 3TCC/	13 13	160 10701	сстт	1370 TAGT	CTGA	۱
1380 GCCTG0	D GAGATG	1390 TTCT/	GACT	TGC:	100 (101)	CCCA	141 CCTC	0 T G A G	11 4 T T A	420 GGAC	AGGA	143 ****	0 TGTG	1 * * * T	440 TTCC		1450 FACAC	D GGAT	ר ד א ד א	460 CGGT/	CCA	1470 TCAC) A T C A 1	14 17761	180 [ggaa	ATTG	1490 GGGT) GACI	r
1500 GTATA0	D GCTGGG	1510 ATTGO) GGCT/	19 19	520 ACTG	TGGT	153 CTTA	0 TCTG	1º TCCA	540 Cata	CAGC	155 C & & & A	0 Atgc	1 CTAT	560 CCAG	A A A T (1570 CCAG	0 TTCG	1 TTGG	580 A A A G	5 A A A	1590 ATTGO) Staci	16 CCT(00 37GCC	ACAG	1610 :GGGT) [TCC <i>i</i>	۱
1620 GAAAAO	D GGGAAG	1630 TCAC) [T T A (10 CTT0	540 3CGG	TGGT	165 GGGA	о тсст	1 Gatg	660 TCTT	TCAT	167 CCAT	0 TTGT	1 A G T A	680 A A A G	CTGG	1690 FAAA	0 GCTT	1 דדכד	700 TACT(CTG	1710 GTTC) 20110	17 20 a G 1	120 FATTI	CTAA	1730 14C11) ГСТСС	3
1740 CACTT	0 TCTCCA	1750 Caggo	D C A T G T	1: 1507	760 ITTG	ACCT	177 TTTT	0 T T C A	1 ATCT	780 TCTA	G A A A	179 GGGA	0 ACGG	1 A A G C	800 A G A A	GTGG	1810 GACA	D TCGA	1 GGGC	820 TCTG	стат	1830 CCTC	D TGCG(18 2766(340 3tgt(3G A A 1	1850 IGCTC) Ботбо	2
1860 ACCTG	O TCCCTT	1870 CTGC	D TGGC:	11 10 A G	380 36 A A I	GTGT	189 CTTC	0 TTGC	1 CCAC	900 ATTT	стст	191 GGGG	0 A A A G	ן GTTT	920 TT A A	тссте	1930 CTGA	о тсст	1 TCCA	940 TCTT(сстб	1950 TTTA	o GGCCI	19 ATGT(360 3000 <i>1</i>	(G X X X	1970 10070) GGACT	r
1980 Gatet	о ттсттт	1990 AATA	D GTGA	20	000 CTGG	GCCA	201 Ctga	0 A G A G	2 T & A C	020 A T G G	стсс	203 ACTG	0 G A C A	2 C A A A	040 AGAG	GGAT	2050 36 A A	0 T C A A	2 Cagg	060 C A G G	GGGC	2070 CTTT	0 TATA	20 AGCC')80 TT+G(3 4 4 4 4	2090 1611;	D A A T G A	A
2100 AACTA	0 TTTCAT	211) CTTT	GGAC'	2 TTTT	120 CAAT	ACTA	213 TTGG	0 Agtg	2 ATTT	140 TTTT	сттт	215 CT A A	0 A C A G	2 G G A A	160 A A T A	ATGT	217) TACA	0 A A A G	2 C & T C	180 TTTT	TTGT	219 TATT	0 TGTT	2 2 TGC A	200 TCCC7	rcccc	2210 :CAC;	accc:	г
2220 GGTGT	0 TTT AA A	2 2 3 4 A T G A	D A G A A .	2	240 A A C C	ATCA	225 CCTT	о ттст	2 A C A A	260 A A A C	тстт	227 A A T G	0 ATTA	2 * * * *	280 ACAA		229	0	2 * * * *	300 A A A A	A A A A	231	0	A A A A .	A				

 10 FIG. 3. Nucleotide sequence and deduced amino acid sequence of human M_r 46,000 Man6PR cDNA. The nucleotide sequence and the corresponding amino acid sequence of M_r 46,000 Man6PR were obtained by analysis of P4a and P29 as shown in Fig. 2. The numbering of the nucleotides starts at the adenosine of the initiator methionine. The arrow points to the N terminal of the mature M_r 46,000 Man6PR. The arrowhead marks the N terminus of the presumptive propeptide of the receptor. The potential N-glycosylation sites are indicated by stars. The lines indicate amino acid sequences that were determined by amino acid sequencing of the receptor and its tryptic fragments. The triple asterisk indicates the stop codon. The presumed membrane-spanning domain is boxed. The neighboring arginine and histidine residues presumably involved in the ligand binding are marked by a dotted line.

membranes (Pall Filtrationstechnik, Dreieich, F.R.G.), and hybridized to nick-translated P4a (13).

Computer Analysis. The hydrophobicity profiles of the amino acid sequences were generated by a program obtained

from Novotny and Auffray (25) using the hydrophobicity parameters given by Rose and Roy (26). Three smoothing cycles were used with a window size of 7. An homology search was performed with a program using the algorithm of Wilbur and Lipman (27), which evaluates direct amino acid matches and the coincidence of related amino acids as defined by the logarithm of odds score of Dayhoff *et al.* (28). The predicted amino acid sequence was compared with that of all the human receptor protein sequences stored in the Protein Sequence Database.**

Amino Acid Sequencing. Human M_r 46,000 Man6PR purified from liver (33) was reduced and carboxymethylated. Tryptic digestion of the purified protein was performed at a trypsin-to-protein ratio of 1:100. Peptides were separated by HPLC using a Baker RP18 wide-pore column and a solvent gradient of 0% to 70% acetonitrile in 0.1% trifluoroacetic acid. Single fractions were rechromatographed with the same acetonitrile gradient prepared in 0.1% ammonium acetate. The sequence analysis of the peptides was performed using an Applied Biosystems gas-phase sequencer with an on-line PTH analyzer (29).

RESULTS

Isolation of a Partial cDNA Clone for the Human Mr 46,000 Man6PR. Human cDNA libraries constructed in the expression vector λ gt11 were screened with affinity-purified polyclonal antibodies against the human M_r 46,000 Man6PR. Screening of placenta (2×10^6 pfu), hepatoma (1.2×10^7 pfu), and liver $(8 \times 10^5 \text{ pfu})$ cDNA libraries yielded 10, 5, and 13 clones, respectively. The DNA of the clones was digested with EcoRI and was examined for hybridization with two 14-mer oligonucleotide probes (8- and 12-fold degenerate) that had been deduced from the partial sequences of two tryptic peptides of the receptor. Under high-stringency conditions, both probes hybridized to a 1.2-kilobase (kb) fragment of clone P4 from the human placenta cDNA library (shown for oligonucleotide 280 in Fig. 1A). The insert of clone P4 had an internal EcoRI site yielding EcoRI fragments of 1.2 kb (P4a) and 1.5 kb (P4b). In blot-hybridization analysis of poly(A)⁺ RNA from human liver, the nick-translated inserts P4a and P4b hybridized to RNA species of 2.3 and 5.2 kb, respectively (Fig. 1B). Therefore, P4a and P4b are considered to be unrelated fragments that were ligated during construction of the cDNA library. P4a was subjected to sequencing (see below). In the longest open reading frame, 134 deduced amino acid residues were colinear with the N-terminus peptide and with nine tryptic peptides of the human liver M_r 46,000 receptor. This established the authenticity of P4a as the M_r 46,000 Man6PR.

Isolation of a Clone Containing the Entire M_r 46,000 Man6PR Coding Sequence. The placenta cDNA library was rescreened using the nick-translated EcoRI-HindIII fragment from the 5' end of P4a (Fig. 2). Plaque hybridization of 1×10^{6} pfu yielded 28 positive clones. The restriction map and sequencing strategy for P29, the longest cDNA isolated from the placenta library, and for P4a are shown in Fig. 2. The nucleotide sequence and the deduced amino acid sequence are shown in Fig. 3. The P29 cDNA is 2463 base pairs (bp) long. It contains a 145-bp 5' untranslated region and an 831-bp open reading frame that is initiated by a methionine-encoding triplet; it is terminated by a TAG stop codon and contains a 1487-bp 3' untranslated region. P4a corresponds to nucleotides 36-1211 of P29. The 5' untranslated region contains an in-frame termination codon at nucleotides -31 to -33. The 3' untranslated region contains a poly(A) tail. A presumptive

polyadenylylation signal, ATTAAA, is present 14 bp upstream of the poly(A) tail.

Predicted Amino Acid Sequence of Human M_r 46,000 Man6PR. The amino acid sequence predicted by the nucleotide sequence is shown in Fig. 3. The 20 residues at the N terminus contain a hydrophobic core and are assumed to function as a signal sequence for insertion of the nascent receptor polypeptide into the endoplasmic reticulum membrane. The hexapeptide located between the putative signal peptide and Thr-27, the amino terminal of the purified receptor, is removed during maturation *in vivo* or during purification. With the exception of asparagine/aspartic acid and glutamine/glutamic acid, the deduced amino acid composition of the receptor closely corresponds to that obtained by analysis of purified human liver receptor (Table 1).

The hydrophobicity profile of the sequence revealed a single hydrophobic stretch of 19 amino acids (residues 191–209) that is not interrupted by charged amino acids and is considered to be the membrane-spanning domain (Fig. 4). It separates the receptor into an N-terminal portion of 164 residues and a C-terminal portion of 68 residues.

Comparison of the amino acid sequence of the M_r 46,000 Man6*P*R with that of other human receptors revealed no homology greater than that seen for nonreceptor proteins.

Chromosomal Assignment. Nick-translated P4a hybridized with a 6.8-kb fragment of *Eco*RI-digested human DNA. Under the same conditions, P4a did not hybridize to mouse DNA (data not shown). The 6.8-kb fragment was present in all hybrids that contained human chromosome 12 (RAG ANLY 1, RAG PI 7-2, RAG PI 5-15, RAG GM194 5-5, A9 1T 2-21-14, and RAG GM610 4-5-1). It was absent from hybrids that lacked human chromosome 12.

DISCUSSION

A potentially full-length cDNA clone for the human M_r 46,000 Man6*P*R was isolated. The nucleotide sequence predicts a precursor receptor polypeptide of 277 amino acids. The first

Table 1.	Amino	acid	composition	of	the	human	М
46,000 Ma	1n6PR						

Amino	Residues per M _r 46,000 Man6PR								
acid	Protein*	Deduced from cDNA [†]							
Asp	18 0 (Aav)	14							
Asn	10.0 (ASX)	13							
Glu	21.6 (Clv)	21							
Gln	21.0 (UIX)	8							
Ser	16.3	18							
Gly	22.0	24							
Thr	7.9	9							
Ala	12.2	13							
His	5.9	7							
Pro	6.8	7							
Arg	11.1	13							
Tyr	6.8	8							
Val	21.1	23							
Met	3.5	5							
lle	9.7	9							
Leu	21.4	23							
Phe	13.0	12							
Lys	11.9	14							
Cys	ND	8							
Trp	2	2							

ND, not determined.

*Composition of the purified human liver receptor protein taken from ref. 33.

[†]Amino acid composition calculated from human placenta cDNA sequences for mature M_r 46,000 Man6*P*R (residues 27–277, Fig. 3).

^{**}Protein Identification Resource (1986) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 9.0.



20 residues of the N terminus have the characteristics of a signal peptide and terminate with Ala-Val-Ala. The latter conforms to the sequence Ala-Xaa-Ala, after which prepeptides are often cleaved by the signal peptidase (30). A hexapeptide is located between the signal peptide and the N terminus of the purified receptor. The gene for the M_r 46,000 Man6PR was localized to human chromosome 12.

The deduced amino acid sequence predicts a M_r of 27,910 for the mature receptor. The M_r of the deglycosylated receptor subunit determined in NaDodSO₄/polyacrylamide gel electrophoresis was 26,500–28,000 (5, 31), which is in good agreement with the predicted value. Therefore, it is unlikely that maturation of the receptor involves gross changes at the C terminus. Based on a decrease in M_r of 18,000–19,500 due to deglycosylation, it has been proposed that the receptor contains 5–8 N-linked oligosaccharides (5, 31). This may indicate that most, if not all, of the five potential N-glycosylation sites in the predicted amino acid sequence (Fig. 3) are utilized.

The exclusive occurrence of N-glycosylation sites in the N-terminal portion locates this part of the receptor on the extracytosolic side of the membrane. The calculated M_r values for the extracytosolic N-terminal, the membrane-spanning, and the C-terminal segments are 18,299, 1963, and 7648, respectively. In a study on the susceptibility of the M_r 46,000 Man6PR to proteinases, it was observed that the receptor is a transmembrane protein that has a large N-terminal portion exposed at the extracytosolic face and a smaller C-terminal portion at the cytosolic face of the membrane (31). The M_r of the C-terminal extension and that of the receptor portion containing the C-terminal extension plus the membrane-spanning domain were estimated to be \geq 5000 and \leq 8500, respectively.

Modification of arginine destroys the binding activity of the M_r 46,000 Man6PR (33). The binding activity is preserved if the treatment is done in the presence of mannose 6-phosphate, suggesting that arginine is a part of the binding site in the M_r 46,000 Man6PR. In the N-terminal portion of the receptor, which is assumed to contain the ligand binding site, eight arginine residues are found. One of these residues (Arg-147) is adjacent to a histidine (His-148). Interestingly, modification of histidine has also been shown to destroy the binding property of the receptor (33). However, presence of mannose 6-phosphate during modification of histidine does not protect the binding property of the receptor. Peptidespecific antibodies based on the sequence information provided by the receptor cDNA and site-directed mutagenesis may enable the identification of the ligand binding site in the Mr 46,000 Man6PR.

Most mammalian cells express a second Man6PR, which is cation-independent and has a subunit M_r of 215,000 (3). The two receptors are immunologically unrelated (33). The partial amino acid sequence of the M_r 215,000 receptor available to us (eight peptides comprising 111 amino acid residues) shows no homology to the sequence predicted for the M_r 46,000 Man6PR.

FIG. 4. Profile of relative hydrophobicity of the M_r 46,000 Man6*P*R and the distribution of positively and negatively charged amino acid residues. The numbers on the abscissa refer to the centers of the heptapeptides for which the relative hydrophobicity is shown on the ordinate.

We would like to thank Drs. J. de Wet and J. O'Brien (University of California, San Diego) for providing us with their cDNA library and also Drs. H. Funke (Münster), K. Beyreuther (Köln), and D. Gallwitz (Göttingen) for the synthesis of oligonucleotides. This work was supported by the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

- 1. Sahagian, G. G. (1984) Biol. Cell. 51, 207-214.
- 2. von Figura, K. & Hasilik, A. (1986) Annu. Rev. Biochem. 55, 167–193.
- 3. Sahagian, G. G., Distler, J. & Jourdian, G. W. (1981) Proc. Natl. Acad. Sci. USA 78, 4289-4293.
- 4. von Figura, K., Gieselmann, V. & Hasilik, A. (1984) *EMBO J.* 3, 1281–1286.
- 5. Hoflack, B. & Kornfeld, S. (1985) J. Biol. Chem. 260, 12008-12014.
- Hoflack, B., Fujimoto, K. & Kornfeld, S. (1987) J. Biol. Chem. 262, 123-129.
- 7. de Wet, J., Fukushima, H., Dewji, N., Wilcox, E., O'Brien, J. & Helinski, P. R. (1984) DNA 3, 437-447.
- Goelet, P., Lomonossoff, G. P., Butler, P. J. G., Akam, M. E., Gait, M. J. & Karn, J. (1982) Proc. Natl. Acad. Sci. USA 79, 5818-5822.
- 9. Gubler, U. & Hoffmann, B. (1983) Gene 25, 263-269.
- Johnson, D. A., Gantsch, J. W., Sportsman, J. R. & Edler, J. H. (1984) Gene Anal. Tech. 1, 3-8.
- Blake, M. S., Johnston, K. H., Russel-Jones, G. J. & Gotschlich, E. C. (1984) Anal. Biochem. 136, 175–179.
- 12. Feinberg, A. P. & Vogelstein, B. (1984) Anal. Biochem. 132, 6-13.
- 13. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- 14. Kidd, V. J. (1984) BRL-Focus 6 (3), 3-4.
- Messing, J., Crea, R. & Seeburg, P. H. (1981) Nucleic Acids Res. 9, 309-321.
 Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
 Birgin M D. Gibson T I & Hong G F (1983) Proc. Natl.
- 17. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
- Graham, A., Steven, J., McKechnie, D. & Harris, W. J. (1986) BRL-Focus 8 (2), 4-5.
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* 18, 5294-5299.
- 20. Aviv, H. & Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412.
- 21. Thomas, P. S. (1983) Methods Enzymol. 100, 255-266.
- Winquist, R., Mäkelä, T. P., Seppänen, P., Jänne, O. A., Alhonen-Hongisto, L., Jänne, J., Grzeschik, K.-H. & Alitalo, K. (1986) Cytogenet. Cell Genet. 421, 133-140.
- 23. Braun, T., Grzeschik, K.-H., Bober, E., Singh, S., Agerwal, D. P. & Goedde, H. W. (1986) Hum. Genet. 73, 365-367.
- 24. Balazs, I., Purello, M., Kurnit, D. M., Grzeschik, K. H. & Siniscalco, M. (1984) Somatic Cell Mol. Genet. 10, 385–397.
- 25. Novotny, J. & Auffray, C. (1984) Nucleic Acids Res. 12, 243-255.
- 26. Rose, G. D. & Roy, S. (1980) Proc. Natl. Acad. Sci. USA 77, 4643-4647.
- Wilbur, W. J. & Lipman, D. J. (1983) Proc. Natl. Acad. Sci. USA 80, 726-730.
- Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in Atlas of Protein Sequence and Structure, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 345-352.
- Hewick, R. M., Hunkapiller, M. W., Hood, L. E. & Dreyer, W. (1981) J. Biol. Chem. 256, 7990-7997.
- 30. von Heijne, G. (1983) Eur. J. Biochem. 133, 17-21.
- 31. Stein, M., Braulke, T. Krentler, C., Hasilik, A. & von Figura, K. (1987) Biol. Chem. Hoppe-Seyler, in press.
- 32. Stein, M., Zijderhand-Bleeleemolen, J. E., Geŭze, H., Hasilik, A. & von Figura, K. (1987) *EMBO J.*, in press.
- 33. Stein, M., Meyer, H. E., Hasilik, A. & von Figura, K. (1987) Biol. Chem. Hoppe-Seyler, in press.