Macrophages express a cell surface receptor which mediates phagocytosis and pinocytosis of particles and solutes containing mannose (fucose and N-acetylgalactosamine are also ligands for the receptor). An apparently identical protein has been isolated from human placenta. Proteolytic fragments of the placental receptor were sequenced so that oligonucleotide probes complementary to the receptor cDNA could be generated. These probes were used to isolate cDNA clones covering the entire coding portion of the mRNA for the receptor. Confirmation that these clones encode the mannose receptor was obtained by expression in rat fibroblasts. The expressed protein mediates uptake and degradation of mannose-conjugated serum albumin. The deduced amino acid sequence of the receptor reveals that it is most likely to be a type I transmembrane protein (COOH terminus on the cytoplasmic side of the membrane) since the mature polypeptide is preceded by a signal sequence and a hydrophobic stop transfer sequence is located 45 amino acids from the COOH terminus. The extracellular portion of the receptor polypeptide consists of three types of domains. The first 139 amino acids constitute a cysteine-rich segment which does not resemble other known sequences. There follows a domain which closely resembles fibronectin type II repeats. The remainder of the extracellular portion of the receptor is composed of eight segments homologous with the C-type carbohydrate-recognition domains of the asialoglycoprotein receptor, mannose binding proteins, and other Ca$^{2+}$-dependent animal lectins. This structure suggests that the receptor may contain multiple ligand-binding domains thus accounting for its tight binding to highly multivalent ligands.

*This work was supported by Grant GM37903 from the National Institutes of Health and a Career Scientist Award from the Irma T. Hirsch Foundation (to K. D.), and Grants CA12858 and AI20015 from the National Institutes of Health (to P. D. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J05501.

† Parker B. Francis Fellow in Pulmonary Research. Present address: Center for Lung Research, Vanderbilt University, Nashville, TN 37232.

‡ Recipient of a Faculty Research Award from the American Cancer Society. To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biophysics, Columbia University, 630 W. 168th St., New York, NY 10032.

The mannose receptor of macrophages and hepatic sinusoidal cells is one of a group of cell surface receptors which mediate endocytosis of glycoproteins in which selected terminal residues are exposed (1, 2). Like other receptors in this group, such as the asialoglycoprotein receptor (2), the Kupffer cell galactose/fucose receptor (3), and the chicken hepatic lectin (4), the mannose receptor requires Ca$^{2+}$ for ligand binding (5). It is distinguished from other receptors by a unique spectrum of sugar ligands: mannose and fucose are the most potent monosaccharide ligands, while N-acetylgalactosamine and glucose bind with lower affinity (6). Also, unlike the other glycoprotein receptors, the mannose receptor has the ability to mediate phagocytosis of saccharide-coated particles as well as pinocytosis of soluble glycoconjugates (7). The receptor has been purified from macrophages (8-10) and from placenta (11). The receptor from both sources is composed of a single subunit, the molecular weight of which has been variously estimated as 162,000-180,000. Antibody cross-reactivity suggests that receptors from the two sources are identical (11).

It has been proposed that the mannose receptor may play a role in several recognition and uptake processes performed by macrophages. Since the receptor efficiently internalizes glycoproteins such as α-hexosaminidase and β-glucuronidase, one function may be to scavenge for secreted lysosomal enzymes from which the mannose 6-phosphate recognition marker has been removed, leaving an exposed high mannose structure (12). The presence of a high density of mannose is also common at the surface of potentially pathogenic microorganisms such as yeast and bacteria. Thus, another function of the receptor may be to direct sequestration and destruction of such cells (7). As might be expected for a receptor involved in such a specialized function of the macrophage, its expression is tightly regulated during development. Receptor appears only during terminal differentiation of the macrophage, and is subject to modulation by various agents such as γ-interferon and steroids (13, 14).

The structure of the mannose receptor has been investigated by isolating cDNA clones encoding the entire receptor polypeptide. Analysis of the sequence of these clones reveals that the receptor is structurally related to other surface glycoprotein receptors, although it has a unique transmembrane orientation and multiple repetitions of the sequence motif which characterizes Ca$^{2+}$-dependent carbohydrate-recognition domains (CRDs). 1

1 The abbreviation used is CRD, carbohydrate-recognition domain.

EXPERIMENTAL PROCEDURES

Materials—Human placental mRNA and two human placental cDNA libraries were obtained from Clonetech, Inc. (Palo Alto, CA)
Mannose Receptor Primary Structure

Library 1075b was screened first and library 1098b was used for isolation of clones encoding the 5' portion of the mRNA. Nylon filters, radiolabeled nucleotides and amino acids, Na\textsuperscript{32}P, Amplify fluoroxylation solution, and materials for labeling by random priming were purchased from American Corp. Restriction enzymes were from New England Biolabs. Reverse transcriptase was a product of Bethesda Research Laboratories and protease Arg-C was a product of Boehringer Mannheim. Reagents and enzyme for sequencing (Sequence I) were obtained from United States Biochemical Corp. Antigen G14 was a product of Gibco. Mannose-conjugated bovine serum albumin was obtained from Pierce Chemical Co.

**Protein Chemistry**—Receptor was purified from placenta as previously described (11). Approximately 1 mg of receptor in 350 μl of 0.15 M NaCl, 10 mM Tris-Cl, pH 7.4, was diluted with 400 μl of 6 M guanidine hydrochloride, 100 mM Tris-Cl, pH 8.5. Reduction with 2-Mercaptoethanol and alkylation with iodoacetic acid were performed as previously described (15). Digestion with trypsin and chymotrypsin and separation of peptides by size-exclusion high-performance liquid chromatography followed by reversed phase chromatography on C\textsubscript{18} columns were performed by published procedures (15, 16). An additional batch of protein was directly digested with protease Arg-C (two aliquots of 2 μg for 90 μg of receptor over 4 h at 37 °C), and peptides were purified by two cycles of reversed-phase chromatography on a C\textsubscript{18} column, again using gradients of acetonitrile, first in the presence of 0.1% ammonium acetate and then in the presence of 0.1% trifluoroacetic acid. Peptides were subjected to automated Edman degradation on Applied Biosystems 470A and 477A Sequencers, and phenylthiohydantoin derivatives of amino acids were identified by reversed-phase high-performance liquid chromatography.

**Analysis of cDNA clones**—Receptor was identified by pulsed-chase labeling with [35S]methionine as previously described for studies with cultured human monocytes (24). Analysis of uptake and degradation of [125]I-labeled mannose-conjugated bovine serum albumin followed the protocol previously used for analysis of the chicken hepatic lectin (25), except that specificity was measured using an excess of unlabeled ligand.

**RESULTS AND DISCUSSION**

**Isolation of cDNA Clones for Mannose Receptor**—With the quantities of protein available from human placenta, it was possible to determine the NH\textsubscript{2}-terminal amino acid sequence as well as sequences of a number of peptides generated by trypsin, chymotrypsin, and protease Arg C digestion. Based on the sequences of these peptides, a number of oligonucleotide probes were prepared. A 44mer, designated Probe 2 (see below), was used to screen a human placental cDNA library in the vector Lambda ZAP. From 500,000 plaques screened 18 phage were isolated. The sequencing strategy for these phage was closely related. The sequencing strategy for these inserts (subclones 51 and 91) is summarized in Fig. 1. The results are presented in Fig. 2, which also shows the positions of the oligonucleotides used for screening. Translation of the sequence confirmed the presence of the amino acid sequences shown in the center. The extent of individual cDNAs are indicated at the bottom. Segments subcloned for sequencing are indicated at the top, with portions for which sequence was read indicated by solid portions of arrows.

The sequence of 100% of both strands was determined, and sequences overlapping all restriction sites were obtained.

---

**FIG. 1. Receptor cDNA clones and sequencing strategy.** A restriction map of the full-length cDNA is shown in the center. The extent of individual cDNAs are indicated at the bottom. Segments subcloned for sequencing are indicated at the top, with portions for which sequence was read indicated by solid portions of arrows.

The sequence of 100% of both strands was determined, and sequences overlapping all restriction sites were obtained.
FIG. 2. Complete sequence of receptor cDNA. The complete nucleotide sequence is presented, along with the translated open reading frame. In addition, sequences of peptides obtained by Edman degradation are shown.
used as a probe to rescreen the placental cDNA library from which these clones were obtained. The longest of the 13 clones isolated (designated number 901) was subcloned and sequenced. Although the sequence of these clones accounted for a number of additional peptides, no full-length clones were obtained from screening of this library.

Attempts to estimate the size of the receptor mRNA by Northern blot analysis using the partial clones were not successful, presumably due to the low abundance of the receptor mRNA in placental poly(A)-positive RNA. As an alternative method of estimating the size of the missing portion of the mRNA, a primer extension experiment was performed using an oligonucleotide complementary to the sequence indicated in Fig. 2. The longest cDNA obtained in the primer extension experiment was approximately 1200 bases in length, suggesting that significantly longer clones had to be isolated.

The 5′ EcoRI-PstI fragment of subclone 901 was used as a probe to screen a second cDNA library. Of the three positive phage isolated, two proved to have inserts longer than those already analyzed. Subcloning and sequencing of the longest insert (number 1001) provided what appears to be nearly the full-length cDNA sequence, based on two criteria. First, the distance from the primer extension oligonucleotide to the 5′ end of this clone is 1227 nucleotides, which agrees well with the size of the longest cDNA observed by primer extension analysis. Second, the clone encodes the NH2-terminal sequence of the protein obtained by direct Edman degradation, preceded by a signal sequence (see below). These data indicate that the full coding sequence of the mRNA has been deduced. One discrepancy observed between the sequence of this clone and those previously characterized is that a sequence of 39 bases found in the other clones is not present in this cDNA (Fig. 1). The possible origin of this deletion is discussed below.

Expression of Receptor in Fibroblasts—Previous labeling studies have strongly suggested that the 162,000–180,000-Da polypeptide from macrophages is the receptor responsible for endocytosis, and immunological evidence indicates that the placental protein is similar or identical to the macrophage protein. The comparison of peptide and cDNA sequence data indicate that the correct placental protein has been cloned, but it was desirable to have direct evidence that this protein is capable of mediating endocytosis. For this reason, the protein was expressed in fibroblasts using a retroviral expression system. Pulse-chase labeling experiments (Fig. 3) were performed to document that expression of the appropriate protein was achieved. Following 1 h of labeling, the major species detected with antibodies to the purified placental protein had an apparent molecular weight of 170,000. During the chase period, this species disappeared and a new form accumulated at 190,000 Da. This pattern of maturation, presumably due to the processing of carbohydrate side chains, is essentially identical to that previously observed for the receptor in cultured monocytes (24).

The cells expressing receptor were tested for the ability to take up and degrade a mannos-containing ligand (Fig. 4). The ligand used, mannos-conjugated serum albumin with 30–40 mol of mannose attached per mol of protein, is known to be a good ligand for the placental and macrophage forms of the receptor (8, 26). Cell-associated ligand increases rapidly.
on warming and reaches a plateau, while degraded ligand, measured as trichloroacetic acid-soluble radioactivity, appears in the medium after a brief lag period. Similar results were obtained for a second, independently isolated cell line expressing the receptor, and no uptake was observed for cells not expressing the receptor (data not shown). These results provide confirmation that the cDNAs isolated encode a functional receptor.

**Predicted Receptor Structure**—The sequence of the receptor deduced from the mRNA sequence shown in Fig. 2 accounts for the 16 unambiguous peptide sequences obtained by Edman degradation, as well as the NH2-terminal sequence. These sequences are distributed throughout the coding segment of the mRNA. The open reading frame predicts a total protein of 1438 amino acids (after removal of a signal sequence; see below). The predicted molecular weight of this polypeptide (155,000) is close to the previous estimate of 150,000 for the mature protein by gel electrophoresis and following growth of cells in the presence of glycosylation inhibitors (24).

The major features of the mannose receptor structure which emerge from analysis of the sequence are summarized in Fig. 5. The presence and arrangement of two hydrophobic sequences suggests that the protein is likely to be disposed as a type I transmembrane protein. The NH2-terminal 18 residues of the initial translation product have the hydrophobic core characteristic of signal sequences (27). These residues are removed from the mature protein by processing at a typical recognition sequence for signal peptidase, Ala-Val-Leu (27). At the other end of the molecule, the hydrophobic sequence extending from residue 1366 through residue 1393, followed by four positively charged amino acid residues, resembles the stop transfer sequence of the type I class of transmembrane proteins (27).

**NH2-Terminal Cysteine-rich Domains**—Much of the extracellular portion of the receptor is cysteine rich. Although the first 139 residues of the protein follow this pattern, extensive sequence comparison with known protein sequences in several databases failed to detect significant similarities with any of the deposited sequences. It is, therefore, difficult to predict the function of this portion of the receptor. The next segment of amino acids (residues 139-192) bears a close resemblance to the type II repeats of fibronectin (28). The closest similarity, highlighted in Fig. 6, is with the second repeat in fibronectin. The observed homology suggests that this domain has been incorporated into several proteins as a result of exon shuffling. This suggestion is consistent with the known gene organization of fibronectin, since the two type II repeats are present on exons separate from the other portions of the gene (28). Unfortunately, no specific function has been ascribed to the type II repeats in fibronectin or the related sequence present in coagulation factor XII, so that no prediction about the role of this domain in the mannose receptor can be presented.

**Potential Ligand-binding Domains**—The sequence of the remainder of the extracellular portion of the receptor is much more revealing. As summarized in Fig. 7, this portion of the molecule consists of eight segments which appear to have diverged from a common ancestral domain. All of the repeats are related in sequence to the C-type carbohydrate-recognition domains of animal lectins (30). Since the mannose receptor binds saccharide ligands in a Ca2+-dependent manner, it seems very likely that one or more of these CRD-like repeats constitute the ligand-binding portion of the protein. The presence of repeated binding motifs (only some of which may be functional) has been observed in two other endocytic receptors: the low density lipoprotein receptor (31) and the cation-independent mannose 6-phosphate receptor (32). However, it should be noted that although the latter protein binds carbohydrate, the repeating motif does not appear to be related to the C-type CRDs.

C-type CRDs have now been found in a number of distinct positions within various membrane, extracellular matrix, and soluble molecules (30). The mammalian asialoglycoprotein receptors are prototypes for this large family of proteins, and, like the mannose receptor, they mediate endocytosis of glycoproteins. However, the asialoglycoprotein receptors (33), their chicken homolog (the N-acetylgalactosamine-binding chicken hepatic lectin) and a fucose/galactose binding protein from Kupffer cells (34) are all type II transmembrane proteins, with orientation opposite to that of the mannose receptor. In these cases, the CRDs lie at the COOH termini of the polypeptides. In contrast, the lymphocyte homing receptor and several related proteins which contain C-type carbohydrate-recognition domains along with other distinct domains are type I transmembrane proteins (55). However, in these cases the CRDs are located at the extreme NH2-terminal ends of the molecules, with other types of domains closer to the membrane. Thus, it appears that C-type CRDs can function at either end or at internal positions within polypeptides.

A detailed analysis of the sequences of a number of C-type CRDs has revealed that these domains can be divided into two evolutionarily divergent subgroups (36). One group is distinguished by the presence of introns within the coding...
sequence for the CRD, and the other group lacks such interruptions. Each group is characterized by certain conserved amino acid residues. For example, all of the CRDs in the intron-positive subgroup contain an extra pair of cysteine residues toward the NH$_2$-terminal end of the domain. Alignment of the CRDs of the mannose receptor with the characteristic residues is shown in Fig. 7. The CRDs are grouped to emphasize that five contain the extra pair of cysteine residues, although all eight appear to be more closely related to the intron-positive subgroup than to the intron-negative class.

The structure of one of the cDNA clones isolated is of interest, because it suggests that the coding sequence for the third CRD may be interrupted by an intron. The cDNA of interest lacks nucleotides 1737 through 1775; the sequence at the 5' end of this segment (GATATGAACAAG) is a possible splice donor sequence. These observations can be reconciled by hypothesizing that the gene for the receptor is likely to contain an intron at position 1775, and that an aberrant (or alternative) splicing event must have generated the unused mRNA. The position of this putative intron would correspond to the position of introns in other C-type CRDs (36).

It is interesting that the mannose receptor CRDs appear to fall into the intron-positive subgroup. The ligand selectivity of the receptor is actually closest to the soluble mannose-binding proteins, which also display affinity for mannose, fucose, and N-acetylglucosamine (37). The mannose-binding proteins fall into the intron-negative subgroup (36). In contrast, most of the members of the intron-positive subgroup show specificity for galactose and/or fucose (for example, the asialoglycoprotein receptor, proteoglycan core protein, and the Kupffer cell lectin). However, we have recently shown that the gene for the chicken hepatic lectin, which can bind N-acetylglucosamine and mannose, is interrupted by introns in the CRD-encoding segment. Thus, it would appear that the soluble mannose-binding proteins and the mannose receptor, although derived from a common ancestor through separate evolutionary paths, have converged to form similar ligand-binding sites.

**Potential Glycosylation**—A complete analysis of the carbohydrate substituents covalently associated with the mannose receptor remains to be undertaken. However, the DNA sequence combined with peptide sequence information allows some conclusions to be drawn. It is known from experiments with glycosylation inhibitors and endoglycosidases that the receptor contains (probably several) N-linked oligosaccharides (24). In addition, reactivity with pea lectin is detected following neuraminidase treatment, indicating that O-linked sugars are also present. Eight potential N-linked glycosylation sites (Asn-$\times$Ser/Thr) are found in the sequence. Sequencing of two peptides containing possible acceptor sites (at positions 326 and 1142) produced no identifiable amino acids at the asparagine positions, suggesting that these are likely to be glycosylated.

No consensus sequence for O-linked carbohydrate attachment has been established, although the presence of proline residues adjacent to or near glycosylated serine and threonine residues has been noted (38). The possibility of glycosylation of certain serine and threonine residues found in this type of context in the receptor is excluded by peptide sequence results in which unmodified residues were identified. However, threonine-rich sites (Asn-$\times$Ser/Thr) are found in the sequence. Sequencing of two peptides containing possible acceptor sites (at positions 326 and 1142) produced no identifiable amino acids at the asparagine positions, suggesting that these are likely to be glycosylated.

*J. M. Rose, M. E. Taylor, and K. Drickamer, unpublished observations.*
strongly suggesting that this is a site of carbohydrate attachment. It is interesting that the proposed sites of O-linked glycosylation lie in the linker regions between candidate CRDs, since it has been suggested that late attachment of sugars may be a cause of the delay between peptide synthesis and attainment of ligand-binding activity in cultured monocytes (39). It is possible that glycosylation may affect the conformation of this portion of the receptor.

Cytoplasmic Domain—One of the striking features of endocytic receptors is the lack of obvious relationships between the primary structures of the cytoplasmic portions of the constituent polypeptides (40). It is, therefore, perhaps not surprising that the COOH-terminal hydrophilic portion of the mannose receptor is not homologous to any of the other known receptors, including those expressed on the surface of macrophages, such as the mannose 6-phosphate receptors and the immunoglobulin Fc receptor. The only consistent feature of the cytoplasmic domains of endocytic receptors is the presence of a tyrosine residue, which appears to be critical for high efficiency localization in clathrin-coated pits (41). The cytoplasmic domains of endocytic receptors is the presence of a tyrosine residue, which appears to be critical for high efficiency localization in clathrin-coated pits (41). The only consistent feature of the cytoplasmic domains of endocytic receptors is the presence of a tyrosine residue, which appears to be critical for high efficiency localization in clathrin-coated pits (41). The only consistent feature of the cytoplasmic domains of endocytic receptors is the presence of a tyrosine residue, which appears to be critical for high efficiency localization in clathrin-coated pits (41).

The presence of multiple saccharide-binding sites, which may represent multiple ligand-binding domains. The primary structure of the mannose receptor reveals several notable features which may be important in the function of this protein. Most striking are the repeated CRD-like motifs, which may represent multiple ligand-binding domains. The presence of multiple saccharide-binding sites, which may differ in their affinities for various complex oligosaccharides, may account for the range of natural and synthetic sugars, which are recognized by this receptor (6,42). Multiple ligand-binding sites may also be responsible for the strong dependence of binding affinity on valency of neoglycoprotein ligands (26). On the other hand, the primary structure alone does not provide obvious clues to how the mannose receptor is routed during endocytosis or how binding of ligand activates other macrophage functions, such as lysosomal enzyme secretion (43). A clearer understanding of these functions must await experiments in which individual segments of the polypeptide are altered and tested for effects on receptor function.

\[ \text{Acknowledgments—We wish to thank Nydia Ruiz and Rita Ros-} \]
\[ \text{hans for technical assistance, and Ned Siegel of Monsanto Corpora-} \]
\[ \text{tion for sequencing some of the peptides.} \]

REFERENCES