Molecular Cloning of 114/A10, a Cell Surface Antigen Containing Highly Conserved Repeated Elements, Which Is Expressed by Murine Hemopoietic Progenitor Cells and Interleukin-3-dependent Cell Lines

(Received for publication, November 21, 1988)

Graeme J. Dougherty†‡, Robert J. Kay††, and R. Keith Humphries‡‡

From the †Terry Fox Laboratory, British Columbia Cancer Research Centre, Vancouver, British Columbia V5Z 1L3, Canada and ‡Department of Medicine, University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada

Monoclonal antibody 114/A10, raised against the murine multipotential hemopoietic progenitor cell line B6SUtA, identifies an antigen highly expressed by primary myeloid progenitor cells, the myelomonocytic leukemia cell line WEHI-3, and various interleukin-3 (IL-3)-dependent cell lines. Western blotting studies indicate that the 114/A10 antigen has a mean relative molecular mass of 150,000 but varies greatly in size range between different cell types. cDNA clones encoding this protein were isolated from a plasmid-based expression vector library prepared from B6SUtA RNA. Three clones corresponding in size to the two major mRNA species detected in IL-3-dependent cell lines (3.0 and 2.2 kilobase pairs) and differing in their utilization of alternative polyadenylation signals were obtained. These clones contain a single long open reading frame of 573 amino acids possessing the typical characteristics of an integral membrane protein. A particularly striking feature of this sequence is the presence at the extracellular amino terminus of a series of eight highly conserved 27-amino acid, serine/threonine-rich (66%) tandem repeats that may serve as sites of extensive glycosylation. The extracellular domain also contains three epidermal growth factor-like cysteine-rich repeats. The distribution and structural characteristics of the 114/A10 antigen suggest a possible regulatory role in the cellular response to IL-3.

All mature hemopoietic cell types present in the peripheral blood and tissues are derived initially from a population of pluripotent stem cells, which in the adult are located mainly in the bone marrow (1). These cells divide and differentiate under the control of a complex interacting network of soluble mediators giving rise to progeny which at each stage are increasingly committed to produce cells of only one lineage (2–4). Proteins expressed on the surface of stem cells and their progeny are believed to play an important role in this process, acting for example as receptors for hemopoietic growth factors (2–4) or as mediators of cell-cell or cell-matrix interactions (5).

As a first step toward identifying and characterizing such functionally important molecules, we have raised a panel of monoclonal antibodies (mAbs) against B6SUtA, a murine bone marrow-derived multipotential progenitor cell line that actively proliferates in response to various growth factors and that forms colonies containing several different mature cell types when appropriately stimulated (6, 7). One of these mAbs, designated 114/A10, has recently been shown² to define an antigen highly expressed by primary bone marrow-derived myeloid progenitor cells, various interleukin-3 (IL-3)-dependent cell lines, and the myelomonocytic leukemia cell line WEHI-3, which constitutively produces IL-3. Western blotting studies indicated that the antigen recognized by mAb 114/A10 is extremely heterogeneous in nature and differs greatly in relative molecular mass (M,) depending upon which cell line is examined.

We report now on the isolation and characterization of full length 114/A10 cDNA clones and demonstrate that these clones encode an unusual, extensively glycosylated integral membrane protein containing a series of eight exceptionally well conserved 27-amino acid serine/threonine-rich amino-terminal tandem repeats and three epidermal growth factor (EGF)-like cysteine-rich repeats.

MATERIALS AND METHODS

Construction of the B6SUtA cDNA Library—The characteristics of the CDMS expression vector and details concerning its use in the cloning of cell surface proteins have been described elsewhere (8–10). Briefly, total cellular RNA was isolated from the IL-3-dependent cell line B6SUtA (6, 7) using the guanidinium isothiocyanate/Caci method. Poly(A)⁺ mRNA was selected by passage over an oligo(dT)-cellulose column and cDNA synthesized according to the method of Gubler and Hoffman (11). Non-self-complementary BstXI oligonucleotide linkers (9) were added and the cDNA size-fractionated on a continuous 5–20% potassium acetate gradient. Fractions containing linkerless cDNA larger than 1 kb were pooled, ligated into BstXI-digested "stuffer-free" CDMS, and transformed into competent Escherichia coli strain MC1061/p3. The resulting library consisted of approximately 5 × 10⁸ recombinants and had a mean cDNA insert size of 2 kb.

Isolation of cDNA Clones Encoding the 114/A10 Antigen—Bacterial colonies comprising the primary cDNA library were pooled and used to transfect approximately 1–2 × 10⁸ COSI cells (12) by protoplast

---

¹ The abbreviations used are: mAbs, monoclonal antibodies; IL-3, interleukin-3; EGF, epidermal growth factor; TGF-α, transforming growth factor type 1; bp, base pairs; kb, kilobase pairs.

transfection (13). After a 48-h incubation period to allow replication of the transfected plasmids and expression of their cDNA inserts, the cells were harvested using phosphate-buffered saline containing 2 mM EDTA and stained with mAb 114/A10 and a fluorescein isothiocyanate-conjugated goat anti-rat IgG second antibody (Cooper Biomedical, West Chester, PA). The 0.5–1% of COS cells expressing the highest fluorescence intensity were sorted on a FACS 440 (Becton Dickinson, Sunnyvale, CA). Replicated plasmid DNA was isolated from these cells by Hirt extraction (14), transferred back into competent MC1061/p8, and subjected to further rounds of fusion and selection. Eventually, plasmid DNA isolated from individual bacterial clones was introduced into COS cells by DEAE-dextran-mediated transfection (15).

Isolation of Additional 114/A10 cDNA Clones by Colony Hybridization—Ten thousand clones from the primary cDNA library were plated, lifted onto nitrocellulose, and lysed as described (16). The filters were hybridized with the XmnI fragment of the CDM8.Al0.1 cDNA which extends from nucleotide number 201-929 (Fig. 2). Putative positive colonies were cloned and rescreened by Southern blot analysis using the same probe.

To compare restriction enzyme sites, BglII-HindIII fragments prepared from the cDNAs were cloned into a derivative pUC vector and digested with DraI, HindIII, SphI, XmnI, BglII, StuI, and AluI.

DNA Sequence Determination and Comparisons—Restriction fragments of cDNA inserts were subcloned into pUC vectors (17) and sequenced by the dideoxy chain-termination method, using denatured double-stranded templates (18) and T7 DNA polymerase (Pharmacia (Canada) Inc., Dorval, Quebec) with reaction conditions as described by the supplier. Both strands of the CDM8.Al0.1 cDNA and the 3'-900 bp of the CDM8.Al0.2 cDNA were fully sequenced, with overlapping sequence obtained at all fragment termini.

Comparisons of the predicted 114/A10 antigen sequence to other proteins in the Genbank, EMBL, and NBRF databases utilized the WORDSEARCH and BESTFIT programs of the University of Wisconsin Genetics Computer Group (19).

Northern Blot Analysis—Ten μg of total cellular RNA isolated from the IL-3-dependent cell lines B6SUTA, 32D, FDC-P1, and DA-1, the T cell line MBL-2, the myeloma cell line SP2/0, and the mastocytoma cell line P815 were electrophoresed through a 1% agarose gel containing 5% (v/v) formaldehyde, transferred to a nylon membrane by centrifugation (20), and cross-linked by a 1-min exposure to ultraviolet irradiation (Philips TUV 15W germicidal tube at 7.2,1 mM EDTA, 1 mg/ml bovine serum albumin and then hybridized with the 114/A10 antigen. 32P-Labeled probes were prepared by oligonucleotide priming (23). Filters were washed as described (24).

Western Blot Analysis—Western blot analysis of 114/A10 antigen expression in COS cells transfected with plasmid DNA using DEAE-dextran and in various IL-3-dependent cell lines was carried out exactly as described previously.

RESULTS

Isolation of cDNA Clones Encoding the 114/A10 Antigen—A library of 5 × 10⁶ cDNAs derived from mRNA isolated from the multipotential hemopoietic progenitor cell line B6SUTA was constructed in the plasmid expression vector CDM8 and introduced into COS cells by protoplast fusion. Cells expressing the 114/A10 antigen on their surface were identified by indirect immunofluorescence staining and purified using a FACs. The replicated vectors present within the sorted cell population were recovered, transformed back into bacteria, and subjected to further rounds of protoplast fusion and selection. Following the third round of protoplast fusion, a distinct population of COS cells expressing high levels of the 114/A10 antigen was observed. Plasmids isolated from these cells were transformed into bacteria and individual colonies picked. Restriction enzyme analysis of plasmid DNA prepared from these colonies revealed that 14 of the 20 clones examined contained an apparently identical 2.2-kb cDNA insert. One of these clones, designated CDM8.Al0.1, is described in detail in this report.

COS cells transfected with CDM8.Al0.1 plasmid DNA reacted strongly with mAb 114/A10 (Fig. 1a) but not with any of nine other antibodies directed against antigens expressed by B6SUTA (data not shown). Untransfected COS cells or COS cells transfected with the CDM8 vector alone were not stained with mAb 114/A10 (Fig. 1a). Western blot analysis confirmed the presence within COS cells transfected with the CDM8.Al0.1 clone of a protein of Mr, 150,000, similar in size to that recognized by mAb 114/A10 in various IL-3-dependent cell lines (Fig. 1b). No proteins reactive with mAb 114/A10 were detected in untransfected COS cells or COS cells transfected with the CDM8 vector (Fig. 1b).

Sequence Analysis of 114/A10 cDNAs—The cDNA insert present within the CDM8.Al0.1 clone was sequenced and found to encode a single long open reading frame of 573 amino acids having the typical features of an integral membrane protein including a 17-amino acid hydrophobic signal sequence, an extracellular domain containing three potential N-linked glycosylation sites (Asn-X-Ser/Thr), a hydrophobic membrane-spanning domain, and a 65-amino acid cytoplasmic tail (Fig. 2). Assignment of the signal sequence cleavage site according to the method of Von Heijne (25) predicts a mature protein of 556 amino acids beginning at residue 18 and having an unglycosylated Mr, of approximately 56,000.

Three ATG trinucleotides are found upstream of the initiating ATG of the open reading frame of CDM8.Al0.1. All three of these ATGs, however, are in a poor sequence context for translation initiation, with pyrimidines at positions –3 relative to the ATG (26, 27). In contrast, the ATG initiating

Fig. 1. 114/A10 antigen expression in transfected COS cells and the IL-3-dependent cell line B6SUTA. a, COS cells transfected with clone CDM8.Al0.1 plasmid DNA (—) or CDM8 vector DNA (-----) were stained with mAb 114/A10 and a fluorescein isothiocyanate-conjugated goat anti-rat IgG second antibody and analyzed on the FACs. b, cell lysates prepared from IL-3-dependent cell lines or COS cells transfected with CDM8.Al0.1 plasmid DNA or CDM8 vector DNA were analyzed for 114/A10 antigen expression by Western blotting.
FIG. 2. Nucleotide sequence and deduced amino acid sequence of 114/A10 cDNAs. The sequence is a composite of the complete CDM8.A10.1 cDNA together with the 5' and 3' ends of CDM8.A10.2 and CDM8.A10.3. The position of the poly(A) tracts of each cDNA are noted above the nucleotide sequence. ATGs preceding the long open reading frame and putative polyadenylation signals following it are boxed. The hydrophobic signal sequence and transmembrane domain are also boxed. Vertical lines above the nucleotide sequence indicate the boundaries of the serine/threonine-rich tandem repeats designated R1 to R8. Potential N-linked glycosylation sites are underlined and cysteine residues are circled.

the long open reading frame, with purines at positions -3 and +4, would be expected to be efficiently utilized as an initiation codon (26, 27). Deletion of the first 166 bp of the CDM8.A10.1 cDNA results in an estimated 3-fold increase in the production of the 114/A10 antigen in transfected COS cells (data not shown). Thus, the upstream ATGs appear to have a repressive effect on 114/A10 expression, probably by diverting a proportion of the ribosomes associating with the 114/A10 mRNA leader to the synthesis of short nonfunctional peptides. Upstream ATGs have been shown to have an equivalent repressive effect on the expression of some oncogenes (28).

The extracellular domain of the 114/A10 antigen contains several areas of particular interest. First of all, immediately following the predicted signal sequence are a series of eight serine/threonine-rich tandem repeats (Fig. 3). In contrast to most previously described mammalian repeat sequences, five of these eight repeats are absolutely identical at both the protein and nucleotide level, while two of the three remaining repeats show only one or two conservative changes from the consensus sequence. The extracellular domain of the 114/A10 antigen also contains a series of three cysteine-rich repeats. A fourth partial repeat is also present. Within these repeats, both the number and spacing of cysteine residues is well conserved (Fig. 4). A comparison with the Protein Sequence Database of the Na-
Expression Cloning of a Hemopoietic Cell Surface Molecule

The CDM8.A10.1 cDNA appears similar in size to the smaller 2.2-kb mRNA transcript found in IL-3-dependent cell lines. In order to characterize further the larger 3.0-kb transcript, additional 114/A10 cDNA clones were isolated from the CDM8-B6SUtA cDNA library by colony hybridization. Two clones were obtained, both of which appeared from restriction enzyme analysis to include the complete protein coding region of CDM8.A10.1. Sequence analysis confirmed that the smaller of the two cDNAs, designated CDM8.A10.3, differed from CDM8.A10.1 only in that it lacked the first 140 bp found within the 5'-untranslated region of CDM8.A10.1 and ended 12 bp 3' of the putative CDM8.A10.1 polyadenylation signal. The larger of the two cDNA clones obtained was approximately 3.0 kb in size and differed from CDM8.A10.1 only in that it lacked the first 9 bp found within the 5'-untranslated region of CDM8.A10.1 and contained an additional 808 bp of 3'-untranslated sequence which included a stretch of 21 CA nucleotides and a second putative polyadenylation signal (Fig. 2). When Northern blots were re-hybridized with a probe prepared from the unique 3' end of the CDM8.A10.2 clone, only the larger of the two mRNA species was detected (Fig. 5). Thus, it seems likely that the two mRNA species present in IL-3-dependent cell lines are generated by the utilization in the 114/A10 gene of two distinct polyadenylation signals. The functional significance of cell-specific differences in the relative levels of the two transcripts remains to be determined.

**DISCUSSION**

In the present study, we describe the molecular cloning of 114/A10, an antigen highly expressed on the surface of murine hemopoietic progenitor cells, the myelomonocytic leukemia cell line WEHI-3, and IL-3-dependent cell lines. Sequencing analysis of the three cDNA clones obtained revealed several interesting features of the molecule including the presence at the extracellular amino terminus of a series

![Fig. 3. Comparison of the amino acid sequence of the eight amino-terminal serine/threonine-rich 114/A10 tandem repeats. Gaps, indicated by hyphens, have been introduced into the last repeat to maximize alignment. Amino acids differing from the consensus sequence are boxed. The sequence motif serine-glycine which could serve as the site of glycosaminoglycan attachment (44, 45) is underlined.](image)

![Fig. 4. Optimal alignment of 114/A10 cysteine-rich EGF-like repeats and comparison with EGF-like sequences present in other proteins. Gaps, indicated by hyphens, have been introduced into the sequences to maximize homology. Conserved cysteine residues are boxed while residues shared by one or more of the other EGF-like sequences are underlined. Sequences were obtained from published data: C. elegans lin 12 (33), Drosophila Notch locus (32), human factor IX (34), human factor X (35), human factor XII (36), bovine protein C (37), human EGF (29), murine EGF (30), and rat TGF-α (31).](image)
been shown to contain near-identical repeats are present in various cell types. While many other proteins have been shown to contain repeat elements, the exactness and length of the serine/threonine-rich repeats found in the 114/A10 antigen is highly unusual. Among the limited number of proteins which have been shown to contain similar near-identical repeats are porcine submaxillary gland apomucin (41) and the human epithelial cell mucin associated with breast and other carcinomas (42, 43). Interestingly, the repeated elements present within these related proteins are also serine/threonine-rich and are located at the amino terminus of the molecule (41-43).

Previous studies have suggested that the majority of the serine and threonine residues present in both porcine submaxillary gland apomucin and human epithelial cell mucin carry O-linked carbohydrate moieties which range in size from monosaccharides to pentasaccharides (43-45). The serine/threonine-rich repeats of the 114/A10 antigen also appear to be extensively glycosylated. Studies involving the deletion of various numbers of repeats from the CDMS.A10.1 expression clone and the addition of these repeats to an indicator gene suggest that in transfected COS cells each repeat has a 5-fold higher than that predicted from its amino acid composition.

Previously we demonstrated that the 114/A10 antigen varies greatly in M, range depending upon which IL-3-dependent cell line is examined. Human tumor-associated epithelial cell mucin is similarly polymorphic and varies considerably in M, in different individuals (42, 43, 46). In this instance, such heterogeneity has been ascribed to individual differences in the number of tandem repeats present within the mucin core protein and both the RNA and DNA encoding this molecule (42, 43, 46). In contrast, Northern blot analysis of RNA isolated from various IL-3-dependent cell lines using a probe that included the serine/threonine-rich repeat region of 114/A10 identified two similarly sized mRNA transcripts in all 114/A10-positive cells examined regardless of the M, range of the 114/A10 antigen expressed by these cells. Moreover, each of the three 114/A10 cDNA clones obtained in this study was found to contain the same number of repeats. In addition, Southern blot analysis revealed an identical pattern of bands on DNA isolated from several different mouse strains (data not shown). Therefore, it seems unlikely that the heterogeneous nature of the 114/A10 antigen is due to differences in the number of serine/threonine-rich repeats present in the core protein of the molecule. Instead cell-specific differences in glycosylation may account for the observed heterogeneity as has been reported for other cell surface proteins (8).

The sequence motif Ser-Gly is conserved in all eight 114/A10 serine/threonine-rich repeats. In other studies this particular sequence has been shown to serve as the site of attachment of glycosaminoglycan side chains to the core proteins of various cell surface and secreted proteoglycans, although additional flanking acidic residues may also be required (47, 48). Initial experiments have suggested that the 114/A10 antigen may indeed contain glycosaminoglycan moieties. Treatment of the CDMS.A10 cell line with para-nitrophenyl-β-D-xylopyranoside, an inhibitor of glycosaminoglycan attachment (49), significantly reduced the size of the 114/A10 antigen detected on Western blots.

Previous studies have demonstrated that IL-3-dependent cell lines derived from murine bone marrow, spleen, lymph node, and fetal liver resemble mucosal type mast cells in that they preferentially synthesize chondroitin sulfate E proteoglycan rather than the heparin proteoglycan associated with connective tissue or seminal type mast cells (50, 51). Interestingly, within IL-3-dependent cell lines, both the 114/A10 antigen and chondroitin sulfate E proteoglycan are most highly expressed within large cytoplasmic granules (50, 51). Recently, a chondroitin sulfate proteoglycan core protein expressed by human fibroblasts has been cloned (52). Although this molecule lacks amino-terminal serine/threonine-rich repeats, it does resemble the 114/A10 antigen in that it too contains EGF-like cysteine-rich domains. Although the function of such EGF-like domains remains to be determined, their conservation in many otherwise distinct proteins suggests that they are of some importance. In this regard, it may be relevant that proteolytic cleavage has been shown to release functionally active ectodomains from many of the cell surface proteins that contain EGF-like cysteine-rich repeats, including EGF itself (29, 30), TGF-α (31), and several different factors involved in the coagulation cascade (34, 36, 37). Such cleavage has been suggested to occur at protease-sensitive arginine-containing sites (30, 36). It is interesting to note then that the 114/A10 antigen also contains a group of arginine residues clustered closely together in an area between the first and second EGF-like cysteine-rich domains. Cleavage at this site would release an ectodomain containing the serine/threonine-rich repeat region attached to a single EGF-like domain. By analogy with EGF and TGF-α such a molecule could perhaps function to transmit regulatory signals to other cell types. The availability of the CDMS.A10.1 expression clone offers the opportunity to study this possibility further.

Acknowledgments—We are grateful to Dr. Brian Seed, Department of Molecular Biology, Massachusetts General Hospital (Boston, MA) for supplying both the CDMS expression vector and advice on cDNA library construction and screening. We thank Darlene Nipius and Patncia Rosten for expert technical assistance.

REFERENCES


FIG. 5. Northern blot analysis of 114/A10 mRNA expression in various cell types. Ten µg of total cellular RNA were run in each lane. Filters were hybridized with a fragment extending from the 5' end of clone CDM8.A10.1 to the PvuII site at position 1643 (a) or the SacI site at position 2266 to the 3' end of clone CDM8.A10.2 (b).