Kell is one of the major blood group systems in human erythrocytes. It is a complex system containing a large number of different antigens. Previously we cloned the Kell cDNA, which was predicted to encode an integral membrane protein with 731 amino acids. Now we have isolated overlapping genomic clones and determined the exon-intron structure of the KEL gene; it spans ~21.5 kb with its coding sequence being organized in 19 exons that range in size from 63 bp to 288 bp. The size of introns ranges from 93 bp to ~6 kb. The donor and acceptor splice sites all conform to the consensus splicing sequences. Exon 1 encodes only the initiation amino acid, methionine, and contains a consensus Sp1 binding site. The single membrane spanning region of Kell protein is encoded in exon 3 and the putative zinc endopeptidase active site is in exon 16. The amino acids encoded by the 19 exons are identical to those of a person with a common Kell phenotype, as determined by RNA polymerase chain reaction of peripheral blood. Amplification of cDNA 5′-ends, derived from human fetal liver, indicated three transcription initiation sites located 30, 81, and 120 bp upstream of the initiation codon. The 5′ flanking region of KEL from −176 does not contain a TATA sequence, but has possible GATA-1 binding sites and has significant promoter activity when determined by chloromphenicol acetyltransferase activity in K562 cells.

## MATERIALS AND METHODS

### Screening of a genomic library.

A human placental genomic DNA library constructed in EMBL3 Sp6/T7 was obtained from Clontech, Inc (Palo Alto, CA). The genomic DNA library was constructed by partially digesting human placenta genomic DNA with Sau 3A I and ligating the fragments into the BamHI site of the vector.

### Isolation of YAC clone containing KEL.

A YAC clone resource, highly enriched for chromosome 7 DNA, was screened using Kell cDNA probes according to standard procedures. The cDNA probes were labeled with 32P by random primer extension using a commercial kit (Boehringer Mannheim, Indianapolis, IN). The specific activity of the probes was ~1 × 109 cpm/μg. Fourteen positive clones were isolated.

### Characterization of exons and introns.

Of the 14 positive clones identified, 5 (A, 2, 8, B, F, and E) were selected for further characterization. The 5 clones were initially mapped by digestion with Xho I plus Pst I and with Xho I plus EcoRI, followed by Southern blot hybridization using different oligonucleotides specific for various locations in the Kell cDNA sequence. The genomic clones were then digested with the following enzymes: Pst I, EcoRI, BamHI, BgIII and Xho I, either individually or in combination. The individual gene fragments were subcloned into PUC18 (GIBCO-BRL, Gaithersburg, MD) or pGEM-3Zf (+) (Promega Co, Madison, WI) and sequenced on an automated 373A DNA sequencer (Applied Biosystems, Foster City, CA). The exons, identified from the Kell cDNA sequence, and their flanking regions, were fully sequenced as were some short introns. The long introns were sized by PCR using primers from the flanking regions of the intron. The longest intron, between exon 10 and 11, was partially sized using primers from known intron sequences obtained from A2 and A6 and using a YAC clone as template. This area of the intron was confirmed by PCR using human genomic DNA as template.

### Restriction enzyme mapping and Southern blot analysis of the genomic clones.

Subclones were mapped with BamHI, EcoRI and Pst I. The different sizes of digested DNA were resolved by agarose gel electrophoresis. The DNA fragments were in some cases further analyzed by Southern blots using 32P-labeled oligonucleotide derived from different locations of Kell cDNA and also identified with available sequences of the subcloned gene fragments. The combination of these procedures allowed a restriction map of EcoRI, BamHI and Pst I to be constructed.

### Transcription initiation site.

The transcription initiation site was determined by the rapid amplification of cDNA ends (5′ RACE) using a human fetal liver (5′ RACE Ready cDNA library, Clontech Laboratories, Palo Alto, CA). The source of poly (A)+ RNA was normal liver pooled from two white female fetuses, 22 and 26 weeks of gestation. The single-stranded cDNA library was made by reverse transcription, using random hexamers as primers. An oligonucleotide anchor (3′-NH2-GGA GAC TTC CAA GGT CTT AGC TAT CAC

---

From the Lindsley F. Kimball Research Institute, New York Blood Center, New York, NY; and the National Center for Human Genome Research, National Institutes of Health, Bethesda, MD.

Submitted May 23, 1994; accepted October 21, 1994.

Supported in part by National Institutes of Health Grant No. HL35841 and the Starr Foundation.

Address reprint requests to Colvin Redman, PhD, The New York Blood Center, 310 E 67th St, New York, NY 10021.

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. section 1734 solely to indicate this fact.

© 1995 by The American Society of Hematology.

From www.bloodjournal.org by guest on October 23, 2017. For personal use only.
TTA AGC AC-p-5') was ligated to the cDNA. For the 5' RACE, two sets of nested PCR reactions were performed, using primers from two different locations on Kell cDNA (nucleotide (nt) 132 and nt 178). In the first PCR, 5' RACE Ready cDNA was used as template. The primers used were the anchor primer provided by Clontech (5'C-CTG GTT CCG CCC ACC TCT GAA GGT TCC AGA ATC GAT AG-3') and Kell antisense primer (5'-CTC GGC TCT TCA CTT TGG TCC-3', nt 132) in one set (PCR 1) and the Clontech anchor primer and Kell antisense primer (5'C-CTC TGG GCT CCA GAG AGT TCC CAT-3', nt 178) in the other set (PCR 2). For the second nested PCR, the products of the first PCR reactions were used as template and the Clontech anchor primer was again used as sense primer for both parallel reactions. In PCR 1, Kell antisense primer (5'-CCC ACC TTC ATG TCT ATC TTC-3', nt 109) was used. For PCR 2, Kell antisense, nt 132, (listed above) was used.

PCR was performed in an automated thermocycler (Minicycler, MJ Research Inc, Watertown, MA) using an initial primer of 94°C for 3 minutes, 62°C for 1 minute, and 72°C for 30 seconds. In cycles 2 to 30, the conditions were 94°C for 30 seconds, 62°C for 30 seconds and 72°C for 30 seconds. In the last cycle, the polymerization step was extended to 72°C. The final concentrations of the reagents were 50 mmoVL KCl, 10 mmoVL TRIS-HCl (pH 9.0), 3 mmoVL MgCl2, 400 mmoVL of each primer, 200 mmoVL of each deoxynucleotide triphosphate, 0.1% Triton X100 and 2.5 U of taq polymerase in a final volume of 100 μL. The "hot start" method using Ampli wax from Perkin Elmer Cetus (Branchburg, NJ) was used. The PCR products were separated by electrophoresis in 0.8% low melting agarose gels, eluted and directly ligated to pT7-Blue(R) plasmid vector (Novagen, Madison, WI) and transformed in DH5αF' strain E coli.

Analysis of 5' flanking region. The 5' flanking region was obtained by DNA sequencing of sub-clone λB, Pst I (see Fig 1). A PCR product spanning -176 to -1 relative to the first cap site was subcloned into pCAT basic vector (Promega, Madison, WI) and K562 cells were transfected cells using the lipofectin method. Construction of the chloramphenicol acetyltransferase (CAT) vector, transfection, and CAT activity were assayed following the protocol provided by Promega. CAT enzymatic activity was measured using [14C]-chloramphenicol (50 to 60 mCi/mmol, Amersham). The reaction products were measured by liquid scintillation spectrometry and analyzing the reaction products by thin-layer chromatography. As a negative control, the pCAT basic vector without promoter was used. All cell extracts were normalized by protein analysis to compare values. A unit is defined as the amount of enzyme required to transfer 1 nmol of acetate to chloramphenicol in 1 minute at 37°C.

Analysis of the 3' end. RNA was isolated from peripheral blood and cDNA prepared by reverse transcription and PCR amplification using a Perkin Elmer RNA PCR kit (Roche Molecular Systems, Inc, Branchburg, NJ). First-strand synthesis was initiated using an anchored oligo d(T)n primer. PCR amplification of the 3' end of Kell cDNA was performed using the anchor primer and an oligonucleotide primer from the coding sequence of Kell cDNA. The anchor antisense primer used was oligo 5'CAGTCCAGTGGACACATGCTT(GT)n-3' and the sense primer from the 3' end coding sequence of Kell cDNA was 5'-ATGGGAGACTGTCCTG-3'.

RESULTS

Isolation and characterization of EMBL3 clones encoding the human Kell blood group protein. Fourteen positive clones were isolated by screening the human genomic DNA library with Kell cDNA probes. Two clones (X2 and E) span most of the Kell gene and cover about 21.5 Kb (Fig 1). Three other clones were also analyzed. Clone λμ was studied because it extended more of the 5' flanking region and also served to reconfirm the sequences determined from clone X2. Clones λB and λF were used because they overlapped with λE, and in addition, a small segment of λF overlapped with the 3' end of X2 (see Fig 1). Sequence analysis of the subcloned Kell gene fragments from these 5 genomic clones showed that the human Kell gene contains 19 exons ranging in size from 63 bp (exon 7) to 288 bp (exon 19) which includes a 3' untranslated region. All exon/intron splice junctions were found to contain the 5' donor-gt and the 3' acceptor-ag sequences. The introns ranged in size from 93 bp to 1036 bp. There were 6 introns which were longer than 1 kb (Figs 1 and 2). The long introns were not fully sequenced but were sized by PCR. There was ambiguity in analyzing the intron between exons 10 and 11.
Fig. 2. Sequence of the individual exons encoding human Kell protein with the immediate intron flanking splice junction sequences.

The base sequences of the individual exons are shown in capital letters and the flanking intron sequences in small letters. The amino acid sequences are shown in bold, and the base substitution in the genomic DNA library was constructed is unknown, we

because the 5' region of clones λB and λF which were expected to overlap with the 3' region of clones 2 and 8, did not. These ambiguous areas of clones λB and λF are shown as dotted lines in Fig 1. Because of this uncertainty, the small gap of the gene not covered by clones λB and λF was bridged by PCR amplification of YAC clone yWSS679 containing the Kell gene (Fig 1). The size of this PCR-amplified region was further confirmed by PCR of genomic DNA obtained from a person of common Kell phenotype using the same primers used for the YAC clone.

All of the exons were sequenced and compared with that of a full-length Kell cDNA isolated from a human bone marrow library. Differences were noted in four bases in exon 3 as compared with the published sequence for Kell cDNA. These differences were caused by sequencing errors in the original study. The corrected sequences were submitted to EMBL/GENBANK Updates (National Center for Biotechnology Information, Bethesda, MD) under accession No. M64934. One notable difference is a base substitution in the membrane-spanning region, which encodes a leucine instead of proline. The corrected base sequences are shown in bold type in Fig 2.

Because the Kell phenotype of the person from whom the genomic DNA library was constructed is unknown, we
isolated RNA from peripheral blood of a person of known common phenotype (K: -1, 2, -3, 4, -6, 7). cDNA was prepared by RNA-PCR and sequenced. The deduced amino acid sequence was identical to that shown in Fig 2 as obtained from the Clontech genomic library. In the person of known phenotype, there were two locations at which C-to-T base differences (nts 1656 and 1664) occurred, but these substitutions did not change the predicted amino acids.

Of interest is that exon 1 includes the 5' untranslated region and codes only for initiation methionine. The single membrane spanning region is located in exon 3 and the pentameric sequence (HELIH) that conforms with a consensus sequence (HEXXH) found in the active sites of zinc neutral endopeptidases is in exon 16 (Fig 2).

Transcription initiation sites. Transcription initiation sites were determined using the 5' RACE procedure and a human fetal liver cDNA library. Two different locations on Kell cDNA were used to initiate the reaction. Each primer yielded 3 different size products. These products were subcloned into Mlu I and Xho I or Mlu I and Hind III digestion restriction mapping and sequencing.

**Fig 2. (Cont'd)**
At least two GATA-1 sites are close to a CACCC box. Sp1 and GATA-1 sequences are present in exon 1. The 5' flanking region contains purine-rich regions. These areas of interest are shown in Fig 3.

The transcriptional activity of the 5' flanking region from -176 to -1 (Fig 3) was determined by CAT assay in transfected K562 cells. As a negative control, the CAT vector was used without a promoter. The -176 to -1 fragment induced high levels of CAT activity in transfected K562 cells as compared with the promoter-less pCAT vector (Fig 4). pCAT vector with the -176 to -1 5' flanking region expressed ~0.8 U of CAT activity per milligram of cell extract protein.

Characterization of the 3' end. Exon 19 is the largest exon, encoding the C-terminal 53 amino acids and containing the 3' untranslated region with a polyadenylation signal 100 bp downstream of the termination codon. Although smaller amounts of larger mRNAs, notably 6.6kb, also were noted, Northern blots showed that the major Kell transcript in bone marrow and fetal liver is 2.5 kb. In cloning the Kell gene, we isolated a cDNA with a large (3 kb) 3' untranslated region. To determine the 3' end structure of the Kell gene, we isolated a cDNA with a large (3 kb) 3' untranslated region. The 3' end sequences were obtained in three different Kell cDNAs obtained from the bone marrow library. The base sequences, before the start of the poly A sequence, differ slightly in length before the start of the poly A sequence. Similar differences in 3' end sequences were obtained in three different Kell cDNAs obtained from a human bone marrow cDNA library. Only one of the cDNA clones, the original full-length cDNA, obtained from the bone marrow library contained a large 3' untranslated region. The 3-kb untranslated fragment from this cDNA clone (no. 191) does not hybridize with human genomic DNA indicating that it is foreign DNA. This foreign fragment is preceded by an EcoRI site and a possible linker and may have been artificially inserted during the preparation of the library.

DISCUSSION

The Kell blood group system is distinguished by its antigenic complexity. Over 20 different antigens have been ascribed to the Kell system and are organized in five antigenic sets: high- and low-frequency antigens, with other Kell antigens being independently expressed. The molecular basis for this antigenic diversity is not yet understood. The current study, which describes the organization of the gene and the boundaries of its 19 exons, sets the stage for the molecular characterization of the different Kell phenotypes.

The coding region of the native protein is present in 18 of the 19 exons; the first exon contains the 5' untranslated region and the initiation methionine. Other examples having only the initiation codon in the first exon are known. The single transmembrane region is encoded in exon 3 with exons

---

[Figures and tables are not transcribed due to the nature of the content.]
4 to 19 encoding most of the extracellular portion. The HEXXH motif, which is unique to zinc metallopeptidases, is encoded in a 68-bp exon (exon 16). In addition to the consensus pentameric sequence, Kell has sequence and structural homology with a family of membrane-associated zinc-neutral endopeptidases (EC24.11) of which the enkephalines and CALLA are examples. The CALLA gene is about 80 kb, much larger than the Kell gene, and is composed of 24 exons. In CALLA, the putative enzyme active site is encoded in exon 19 and a comparison of base sequences of exons 18 and 19 of CALLA with exons 15 and 16 of Kell shows 54.5% base identity.

Both the bone marrow cDNA library from which Kell cDNA was obtained and the human genomic library were purchased commercially. The Kell phenotypes were unknown. For this reason, a sequence was obtained by PCR of peripheral blood RNA from a person of common Kell phenotype. The exons deduced from the genomic library encode identical amino acids as the cDNA obtained from peripheral blood indicating that both share common Kell phenotypes. There was no amino acid polymorphism noted in the two samples and the only difference were two C-to-T changes that do not affect the primary structure.

Two of the genomic clones from the commercial genomic library, clones λB and E, gave inconsistent sequences when compared with other clones (λ2 and λ8). The 5' segments of clones λB and λE should be identical to the 3' segments of 2 and 8, but sequence analysis showed that they were different. The reasons for this are not known, but it could be because of ligation of DNA material during the preparation of the genomic library. A small segment of the large intron not covered by clones λ2, λ8, and λE was determined by sequencing PCR products from YAC clone yWSS679 and from genomic DNA from a person of known Kell phenotype instead of the clones obtained from the commercially purchased genomic library. PCR primers from the 3' end of λ2 and the 5' end of λE were used in PCR amplification of the YAC clone and the genomic DNA. This strategy was used even though the λE clone covers the 3' end of the λ2 clone. This allowed us to cover the entire Kell gene and to size the intron between exons 10 and 11 without ambiguity.

Three possible transcription initiation sites were found using the 5' RACE procedure and poly (A) RNA from fetal liver. The first of these cap sites, located 120 bp upstream from the initiation ATG, is also the 5' end of a cDNA cloned from a human bone cDNA library of the other two sites are 81 and 30 bp upstream from the initiation codon. All three sites were obtained using two different Kell cDNA antisense primers. Although all three locations are purine bases and could act as transcription initiation sites. It is also possible that those at 81 and 30 bp upstream from the ATG are artificial because of incomplete reverse transcription. This is unlikely given that random hexamers were used to prepare the cDNA library. However secondary RNA structures can also cause premature termination of reverse transcription.

Introduced in front of a CAT reporter gene, the 5' flanking region -176 to -1 nt exhibited promoter activity in the erythroleukemic cell line, K562. The putative promoter region does not contain the typical TATA box which is usually located -25 to 30 nt relative to the cap site. However, as in several erythroid specific genes, consensus sequences for GATA-1 factor was found in the promoter region. In addition, possible Sp1 and GATA-1 binding sites were noted in exon 1. We do not know whether GATA-1 and Sp1 regulate KEL gene expression, but GATA-1 is common in erythroid genes and is known to be expressed at low levels in hematopoietic progenitor cells and upregulated during erythroid maturation. If these transcription factors define erythroid tissue specificity, it will be in agreement with our Northern blot studies that detected Kell transcripts in bone marrow and fetal liver, but not in several nonerythroid tissues.

The previous Northern blot analysis with human bone marrow and fetal liver showed a major 2.5-kb transcript and smaller amounts of larger transcripts. We had originally obtained a cDNA from a human bone marrow library with a large 3' untranslated region. The untranslated 3' region follows an EcoRI site and is probably an artifact of preparation of the cDNA library. Sequences of the 3' segments of transcripts in peripheral blood by RNA PCR only detected short 3' untranslated regions that varied slightly in length before the poly A tail. Similar sequences were obtained when other cDNAs from the human bone marrow library were analyzed. The distance between the polyadenylation signal (AATAAA) and the cleavage site is known to be variable. The larger Kell transcripts could not be amplified using the RNA PCR method because of their length, but Northern analysis indicates their presence. The occurrence of multiple transcripts with larger 3' untranslated region is not uncommon, and although their function is not well understood, the 3' untranslated regions are thought to play roles in regulation of expression.

ACKNOWLEDGMENT

We are indebted to Don Lee of the Microchemistry Laboratory at the New York Blood Center for his valuable help in DNA sequencing and the synthesis of primers, Cheng-Han Huang for helpful suggestions during the preparation of the manuscript, Helene Currie for the preparation of the manuscript, Tellervo Huima for illustrations, and Xu Wu for technical assistance. This article is dedicated to the memory of Eleni Zambas.

REFERENCES

12. Murphy MT, Morrison N, Miles JS, Fraser RH, Spurr NK, Boyd E: Regional chromosomal assignment of the Kell blood group locus (KEL) to chromosome 7q35-q35 by fluorescence in situ hybridization: Evidence for the polypeptide nature of antigenic variations. Hum Genet 91:585, 1993
Organization of the gene encoding the human Kell blood group protein [published erratum appears in Blood 1996 Jun 1;87(11):4922]

S Lee, E Zambas, ED Green and C Redman

Updated information and services can be found at:
http://www.bloodjournal.org/content/85/5/1364.full.html

Information about reproducing this article in parts or in its entirety may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#repub_requests

Information about ordering reprints may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#reprints

Information about subscriptions and ASH membership may be found online at:
http://www.bloodjournal.org/site/subscriptions/index.xhtml