Cloning of the Gene Encoding the Human Erythropoietin Receptor

By Leila Maouche, Christophe Tournamille, Claude Hattab, Georges Boffa, Jean-Pierre Cartron, and Stany Chrétien

The genomic and complementary DNAs of the human erythropoietin receptor (hEpo-R) have been isolated and characterized from a genomic placental library and from two cDNA libraries prepared from bone marrow and fetal liver. The five different partial cDNAs isolated were aberrant in the predicted reading frames as compared with the Epo-R protein sequence, because all retained insert sequences that may represent splicing intermediates (three clones), cloning artifact (one clone), or a new sequence at a splice junction (one clone) of the gene. The cDNAs were used to isolate several genomic clones encompassing the complete hEpo-R gene. This gene, which encodes a 500-amino acid polypeptide chain of predicted Mr 55,000, is organized into eight exons spread over 6 kb of DNA and exhibited a high degree of sequence homology (81.8% in the coding region) and structural organization with its murine counterpart. Primer extension analysis indicated that the transcription initiation site is located 141 bp upstream of the initiation codon. Sequence homology 320 bp upstream of the cap site was significantly lower (60%) and diverged completely further upstream as compared with the murine gene. Similarly, the human and murine sequences were largely divergent downstream of the stop codon, indicating that a strong conservation during evolution was restricted to the coding sequence of the Epo-R protein. The 320-bp region upstream of the cap site does not contain the typical TATA or CAAT boxes present in many tissue-specific genes, but does include potential binding sites for the ubiquitous Sp1 and the erythroid-specific GATA-1 trans-activating factors. These boxes are well conserved in sequence and position relative to the cap site within the promoter region of the human and murine genes, but the CACCC boxes present in the murine gene are absent in the human gene.

© 1991 by The American Society of Hematology.

Materials and Methods

Oligonucleotides and primers. Oligonucleotides were synthesized on a Milligen Biosearch 8700 DNA synthetizer (Millipore, Bedford, MA) and purified on a 20% acrylamide-urea gel.

cDNA cloning. The mEpo-R cDNA was cloned by polymerase chain reaction (PCR) amplification using available sequence information, as previously described. The Epo-R cDNA was radiolabeled by the random priming method (Boehringer, Mannheim, Germany) using (32P)dCTP (3,000 Ci/mmol) from Amersham (Bucks, UK) and hybridized at low stringency (prehybridization and hybridization in 6X SSC, 5X Denhardt, 0.05 mol/L Napi, and salmon sperm DNA 0.1 mg/L at 52°C overnight and washings in 0.5X SSC, 0.05 mol/L Napi at 52°C) to one million bacteriophage Agt11 plaques plated from a human bone marrow cDNA library (Clontech, Palo Alto, CA) and transferred to nylon membranes (Hybond N+; Amersham). Of the 15 positive clones analyzed, two (6RE and 11RE) containing the longest inserts were selected and sequenced, but were not full length. Using a new probe derived from the 5' region of the 6RE cDNA, we were able to isolate 13 new positive clones from a fetal liver cDNA library (Clontech, Palo Alto, CA) and transferred to nylon membranes (Hybond N+; Amersham). Of the 15 positive clones analyzed, two (6RE and 11RE) containing the longest inserts (2RE, 9RE, and 15RE) were purified and sequenced.

PCR cloning of the 5' end of the hEpo-R cDNA. Total cellular RNA from human erythroid spleen erythroblasts was extracted according to Chirgwin et al and 10 μg was used for cDNA preparation (RNAgen 40 U, Tris 100 mmol/L, pH 8.3, KCI 140 mmol/L, MgCl2 10 mmol/L, pH 8.3, β-mercaptoethanol 28 mmol/L, dNTPs 1 mmol/L, avian myeloblastosis virus (AMV) reverse transcriptase 10 U, 40 minutes at 42°C). This cDNA was used as template for PCR amplification of Epo-R cDNA, using primers INTS-527 (GGGGCTGT-ATCATGGACCA) and INTS-530 (GGGCTGT-ATCATGGACCA) and INTS-530 (TGAAGCAGAAGCTCT) deduced from the cloned genomic hEpo-R DNA, under conditions suggested by Perkin Elmer Cetus (Norwalk, CT). The probe INTS-527 was localized in exon 1 at

From the Institut National de Transfusion Sanguine and the Unité INSERM U76, Paris, France.

Submitted April 24, 1991; accepted July 12, 1991.

Supported by the Ligue Nationale Contre le Cancer (Comité de Paris, Grant No. 89CC/209).

Address reprint requests to Jean-Pierre Cartron, MD, Unite INSERM U76, INTS, 6 rue Alexandre Cabanel, 75015 Paris, France.

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.

© 1991 by The American Society of Hematology.

0006-4971/91/7810-0028$3.00/0
nucleotide 131 and INTS-530 is localized in exon 2 at nucleotide 301.

**hEpo-R genomic clones.** A human genomic placental library (Clontech) cloned in λEMBL3 was screened with the cloned hEpo-R cDNAs. The first probe used was the restriction fragment Bgl II-BamHI containing exon 8 and a part of exon 7. This probe was 32P-labeled by random priming and hybridized under stringent conditions as above. Two recombinant clones, λ1 and λ2, were purified and the DNA was extracted and analyzed. The second probe used for library screening was a PCR probe specific for exon 1 isolated as described above. Using this 32P-labeled probe, a further genomic clone, λ3, encompassing the 5' region and the promoter of the Epo-R gene was isolated.

**DNA sequencing.** cDNA and genomic DNA inserts were subcloned into the pUC18 vector and sequenced on both strands by the chain-termination method with a Pharmacia T7 sequencing kit (Pharmacia, Uppsala, Sweden), using universal primers or specific oligonucleotides as internal primers.

**Primer extension analysis.** Primer extension analysis was performed using synthetic oligonucleotides INTS-680 (CGGCA-CAGTCCACAGCTGGGT) and INTS-602 (TGGTCCATGATACGCC) and 20 μg of RNAs isolated from a human erythroleukemic cell line and human spleen erythroblasts. Hybridization and reverse transcription were performed as described and extended products were sized by electrophoresis in a 6% polyacrylamide/7 mol/L urea gel and detected by autoradiography. The Epo-R promoter sequence determined with oligoprobes INTS-680 and INTS-602 was used as a size marker.

**RESULTS AND DISCUSSION**

**Epo-R genomic clones.** Three overlapping genomic clones encompassing the complete coding sequence of the Epo-R were isolated from a human placental λEMBL3 library using the hEpo-R cDNA probes isolated as described below. After subcloning in pUC18 vector, the 8 exons encoding the hEpo-R gene and all intron-exon junctions were sequenced (Fig 1A and Table 1). The organization of the human gene that spreads over 6 kb of DNA, as well as the intron-exon junctions have been compared with the structure and organization of the murine counterpart published recently. There is a striking homology between the two genes that are composed of 8 exons of identical size, except for exon 2, which is 3 bp longer in the human gene, and for slight differences in exons 1 and 8. The introns are also identically located within the two genes and the intron-exon junctions conform
### Table 1. Intron-Exon Junctions of the hEpo-R and mEpo-R

<table>
<thead>
<tr>
<th>Intron</th>
<th>Exon</th>
<th>Intron</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>256</td>
<td>255</td>
</tr>
<tr>
<td>5' C</td>
<td>TAG A</td>
<td>1 GCA A</td>
</tr>
<tr>
<td>257</td>
<td>392</td>
<td></td>
</tr>
<tr>
<td>1 C</td>
<td>TAG A</td>
<td>1 GCA A</td>
</tr>
<tr>
<td>393</td>
<td>568</td>
<td></td>
</tr>
<tr>
<td>1 T G A A</td>
<td>1 GCA A</td>
<td></td>
</tr>
<tr>
<td>569</td>
<td>726</td>
<td></td>
</tr>
<tr>
<td>1 C</td>
<td>TAG A</td>
<td>1 GCA A</td>
</tr>
<tr>
<td>727</td>
<td>880</td>
<td></td>
</tr>
<tr>
<td>1 C</td>
<td>TAG A</td>
<td>1 GCA A</td>
</tr>
<tr>
<td>881</td>
<td>968</td>
<td></td>
</tr>
<tr>
<td>1 C</td>
<td>TAG A</td>
<td>1 GCA A</td>
</tr>
<tr>
<td>969</td>
<td>1056</td>
<td></td>
</tr>
<tr>
<td>1 C</td>
<td>TAG A</td>
<td>1 GCA A</td>
</tr>
<tr>
<td>1057</td>
<td>1881</td>
<td></td>
</tr>
</tbody>
</table>

The position of each intron with respect to the amino acid codon sequence is indicated, as well as each intron-junction in the hEpo-R (boldface) and the mEpo-R gene. Numbers refer to nucleotide positions on the human sequence taken from the transcription initiation site (+1). Nucleotide differences are boxed. Exons are in capitals.

The overall homology between the nucleotide sequences encoding the hEpo-R and mEpo-R proteins is 81.6%, we found no homology in the 3' region of the two genes downstream of the stop codon, at least for the 230 bp of the human gene sequenced (not shown). Except for an AT-rich region common to the human and murine genes, neither a canonical polyadenylation signal (AATAAA) nor a signal of efficient 3' terminus processing (YGTG'TYY) could be identified in the 3' region of the gene. Because there is also a significant divergence between the two genes upstream of the transcription initiation site (Fig 2 and see below), only the nucleotide region encoding the Epo-R protein itself is highly conserved, a finding that correlates well with the lack of species specificity of the ligand Epo.

Structure of the human cDNA clones. Two different human cDNA libraries (bone marrow and fetal liver) were used to isolate five Epo-R cDNA clones by cross-hybridization with an mEpo-R cDNA probe. After sequencing and comparison with the mEpo-R cDNA and the recently published sequence of the hEpo-R cDNA, we found that all our human cDNAs were abnormal. Indeed, as schematically presented in Fig 1B, all cDNAs contain additional sequences that correspond either to unspliced introns, aberrant insertional sequences, or new sequences at normal intron-exon junctions, as deduced from the Epo-R genomic clones analysis. In fact, no full-length cDNA could be isolated from these libraries. The 5'-region of the hEpo-R messenger RNA (mRNA) was amplified by PCR from the reverse transcribed RNAs extracted from a human erythro-leukemic cell line, using oligonucleotide probes deduced from the genomic sequence. Primer extension analysis performed with primers INTS-680 and INTS-602 and RNAs from human erythroleukemic cells or spleen erythroblasts yielded identical extension products of 81 and 149 nucleotides, respectively (Fig 3), therefore indicating a major transcriptional start site at 141 nucleotides upstream of the translation initiation codon (Fig 2). Alignment of nucleotide sequences indicates that the human and murine cap site defined by Youssoufian et al and Lacombe et al fall in a region of homology. However, a different cap site for the murine gene (43 to 46 bp upstream of the initiation ATG) has been identified independently, using mRNAs from the SKT6 murine erythroleukemia (MEL) cell line. It is possible that this different cap site results from a rearrangement of the Epo-R promoter region, as we found recently for the murine T3C1-2 cell line. From all these data, we have deduced the complete sequence of the hEpo-R cDNA and of the Epo-R protein.

In clone 11RE, one of the additional sequence (93 bp) corresponds to the unspliced intron 7 and the other to a fragment of intron 6. Similarly, the extra sequences present in clone 9RE correspond, respectively, to an unspliced intron 3 (80 bp) and to a fragment of intron 2. All these new sequences are located at the expected exon-intron junctions. Clone 15RE retains intron 3 and a fragment of intron 1, again at correct splice junctions. The cDNA clone 6RE contains an unknown sequence (called S1 on Fig 1B) inserted into exon 6, very likely attributable to a cloning artifact. The cDNA clone 2RE is intriguing because, in addition to the unspliced intron 7 (93 bp), there is an...
we found no homology of S2 with any reported nucleotide sequence prepared from the human erythroleukemic cell line OCIM-1. The amino acid sequence of the Epo-R predicted from the genomic DNA.35 46 The amino acid sequence of the Epo-R protein deduced from the cDNA cloning recently published by Jones et al22 but differ at four positions (amino acids 102, 189, 190, and 244 of the protein of 508 residues, highly homologous (81.6%) to the protein of the Epo-R sequence shown in Fig 1 is M76595), and search in the GenBank database (the accession number of the human Epo-R sequence shown in Fig 1 is M76595), 27,28 Rh blood group protein,27 and the transcription factors hGATA-1 and hGATA-3.36 It is possible, therefore, that stable unspliced Epo-R RNAs are produced in these human tissues and that splicing events represent an efficient expression regulation of the Epo-R translation protein. Using one of the hEpo-R cDNA in Northern blot hybridization to poly(A)⁺ RNAs from human spleen erythroblasts, we found a major band at 2 kb as well as some larger transcripts, which might account for new mRNA species like those described in Fig 1B.

We have finally identified aberrant hEpo-R transcripts carrying unspliced intron sequences 1, 2, 3, 6, and 7, but not intron sequences 4 or 5. Kuramochi et al have recently described an mEpo-R cDNA clone containing two insert sequences, one of 78 bp that most likely represents the unspliced intron 5 and the second of 77 bp, located downstream, that represents a fragment of intron 6 and might result from alternative splicing (use of a cryptic site of splicing in intron 6). As a result, this clone encodes a 29,000 Mr soluble Epo-R still binding Epo, because the first insert introduced a new 20-amino acid sequence and a termination codon before exon 6, which encodes the transmembrane domain of Epo-R. Unspliced introns such as those we have described in this report might also generate different forms of receptors, for example, membrane-bound forms of receptors truncated within their cytoplasmic domain, but whether such molecules do exist remains to be proven. Another illustration of this situation has been described recently for the receptor of interleukin-7 (IL-7).29 It is clear from these data that these abnormal splicing events are not artifactuals, but may rather indicate that the splicing process fidelity is not so absolute. This idea is supported by the recent identification of abnormally spliced transcripts, in which exons from a tumor suppressor gene (DCC) have been scrambled relative to their order in genomic DNA.30

### hEpo-R protein

Information deduced from the cDNAs and the gene structure indicates that exons 1 to 5 encode the extracellular domain of Epo-R, exon 6 the transmembrane domain, and exons 7 and 8 the intracellular domain. The amino acid sequence of the Epo-R predicted from the nucleotide sequence data (not shown) is a single-chain protein of 508 residues, highly homologous (81.6%) to the murine protein. These findings are in excellent agreement with the sequence of the Epo-R protein deduced from cDNA cloning recently published by Jones et al but differ at four positions (amino acids 102, 189, 190, and 244 of the preprotein) from the sequence of Winkelmann et al. However, these differences arose from unfortunate publication of an uncorrected sequence (B. Forget, personal communication, November 1990) and, therefore, all groups agree with the same data.
**hEpo-R promoter.** It is interesting that the homology between the mEpo-R and hEpo-R genes is less pronounced for the region upstream of the translation initiation site over approximately 460 nucleotides (Fig 2) than for the region encoding the receptor protein itself (60% and 81.6%, respectively). Further upstream (not shown), the human sequence is rich in AG repeats and diverges completely from the murine sequence. The promoter region of the Epo-R gene encompassing 320 bp upstream of the transcription initiation site (Fig 2) has no typical TATA or CAAT boxes, but includes a conserved Sp1 motif as well as a GATA-1 (NF-E1, GF-1, Eryf1) inverted consensus sequence identified as a potential binding site for trans-activating factor(s) responsible for erythroid-specific gene expression.23,32 Kuramochi et al23 have detected five GATA-1 motifs within the 1,640 bp upstream of the transcription initiation site. Except for one of them, these motifs are also present (but not outlined) in the sequence published by Youssoufian et al.7 In the promoter region shown in Fig 2, there are two CACCC boxes within the murine gene that are apparently lost in the human counterpart. However, the regions surrounding the initiation transcription site as well as the Sp1 and GATA-1 motifs of both genes are well conserved. The relative distance between these boxes is also

**Fig 3.** Transcriptional start site of the hEpo-R in erythroleukemic cell and spleen erythroblast RNAs. Primer extension analysis was performed with the 5' end-labeled primers INTS-602 and INTS-680. RNA templates from human erythroleukemic cells were extended with INTS-602 (lane 1); adult liver RNA and human spleen erythroblasts were extended with INTS-602 (respectively, lane 2 and 3) and INTS-680 (respectively, lane 4 and 5). A schematic drawing of the mRNAs, probes, and amplification products used in the primer extension analysis is given. Sequencing ladders (lane A with INTS-602, lane B with INTS-680) were used as size markers.
conserved and this might be important for the transcriptional activity of the Epo-R promoter and the choice of the cap site. Three other GATA-1 motifs have been found downstream in the gene, respectively, in exon 3, intron 4, and the 3' noncoding region. Further confirmation of these potential regulatory elements is required because the two published murine sequences exhibited a number of differences, possibly related to sequence errors. In addition, gel-shift assay, DNase footprinting, and deletion analysis should establish which of these motifs is used and involved in the gene regulation. Other erythroid-specific genes that have no TATA or CAAT boxes include human glycoporphins A, B, and E, murine band 3, rat pyruvate kinase R-type, and Rh protein (Chérif-Zahar et al, in preparation). The promoter of these genes all carry GATA-1 motifs, one being often located in lieu of the TATA box, suggesting that GATA-1 protein(s) might be also involved in transcriptional activation and cap site determination, perhaps in association with TFIID or other factors. In addition, as noted previously, the GATA-1 motifs are frequently positioned in close vicinity to a CACCC box (as found also for the mEpo-R but not the hEpo-R), suggesting further interaction with CACCC-binding factors. A functional characterization of the hEpo-R promoter should be performed to address these issues and provide new insights into erythroid differentiation under normal and pathologic conditions.

ACKNOWLEDGMENT

We thank Dr P.-H. Romeo, W. Vainchenker (INSERM U91, Hôpital Henri Mondor, Créteil), Dr C. Lacombe, and P. Mayeux (INSERM U152, Hôpital Cochin, Paris) for help and critical reading.

REFERENCES

Cloning of the gene encoding the human erythropoietin receptor

L Maouche, C Tournamille, C Hattab, G Boffa, JP Cartron and S Chretien

Updated information and services can be found at:
http://www.bloodjournal.org/content/78/10/2557.full.html
Articles on similar topics can be found in the following Blood collections

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