erg, a Human *ets*-Related Gene on Chromosome 21: Alternative Splicing, Polyadenylation, and Translation

VEENA N. RAO,* TAKIS S. PAPAS, E. SHYAM P. REDDY

The avian acute leukemia virus E26 induces a mixed erythroid-myeloid leukemia in chickens and carries two distinct oncogenes, v-*myb* and v-*ets*. Recently, a novel gene named *erg*, closely related to the v-*ets* oncogene, was identified in human COLO 320 cells and the nucleotide sequence of its approximately 5.0-kilobase transcript, *erg* 1 was determined. In the present study, the nucleotide sequence of the alternatively spliced transcript, *erg* 2, was found to differ from *erg* 1 by a splicing event that causes a coding frameshift near the amino terminus, resulting in an additional 99–amino acid insertion at the amino-terminus. Expression of complementary DNAs for the two transcripts in vitro resulted in synthesis of polypeptides of approximately 41 and 52 kilodaltons, suggesting the use of alternative translation initiation codons in the case of *erg* proteins. The *erg* gene was localized by somatic cell genetic analysis to human chromosome 21. It is proposed that alternative sites of splicing and polyadenylation, together with alternative sites of translation initiation, allow the synthesis of two related polypeptides from a single *erg* gene transcriptional unit.

PROTOONCOGENES ARE THE CELLULAR counterparts of the oncogenes of acute transforming retroviruses. The replication-defective avian erythroblastosis virus E26 induces a mixed erythroid/myeloid leukemia in chickens and transforms cells of both erythroid and myeloid lineage (*1, 2*). E26 includes elements from two protooncogenes, chicken proto-*myb* and chicken proto-*ets*, and Δ*gag* from the viral *gag* gene (*3, 4*). This tripartite oncogene is expressed as a 135-kD transforming polyprotein ($p135^{gag\text{-}myb\text{-}ets}$) localized in the nucleus of transformed cells. Two distinct *ets* chromosomal loci (Hu-*ets*-1 and Hu-*ets*-2) appear to be present in humans and other vertebrates (*5*). The human *ets* 1 locus on chromosome 11 encodes a single messenger RNA (mRNA) of 6.8 kb; the second locus, *ets* 2, encodes three mRNAs of 4.7, 3.2, and 2.7 kb (*6*). By contrast, the chicken homolog has contiguous *ets* 1 and *ets* 2 sequences and is expressed in normal chicken cells as a major 7.5-kb mRNA (*3*). The Hu-*ets*-1 and Hu-*ets*-2 genes are transposed in certain leukemias (*7, 8*). Because of the significance of *ets* in neoplasia, we searched for other human genes closely related to *ets*. In the chicken, there appears to be a set of proteins that are highly related to each other but have a limited domain of homology with v-*ets* encoded polyprotein (*9*). Recently, several genes that are closely related to, but distinct from, known retroviral oncogenes such as *syn* (*10*), *slk* (*11*), *ral* (*12*), *rho* (*13*), *neu* (*14*), and *arg* (*15*) have been detected in different cells. We previously reported a new class of *ets*-related gene, *erg* (*16*). This gene differs from the Hu-*ets*-2 gene represented by cDNA-14 clone (*6*) in having a more extended domain of homology in the 5′ region of v-*ets*. In this report, we show that *erg* 1 and *erg* 2 mRNAs are the alternative products of *erg*. Results of sequence analysis suggest that these two mRNAs could result from the use of two distinct polyadenylation sites associated with differential splicing patterns. The two related polypeptides may result from the use of different initiation AUG codons. Thus, an increasing number of genes are now understood to invoke controlled mechanisms of alternative RNA processing in the generation of protein diversity. Here we used a panel of rodent × human somatic cell hybrids to assign the chromosomal localization of *erg* to human chromosome 21.

A size-selected cDNA library was made from mRNA of COLO 320 cells by using λgt10 as a vector (*17–19*). This library, when screened with the Hu-*ets*-2 cDNA (*6*), gave three positive clones, λ7, λ12, and λ8. We have reported previously (*16*) the nucleotide sequence of the cDNA clone λ7, which was named *erg* 1 (*ets*-*r*elated *g*ene) because of its homology with the *ets* oncogene. On the basis of restriction analysis, another cDNA clone, λ12, called *erg* 2, was identified (Fig. 1). The *erg* 2 is a full-length cDNA clone about 3.2 kb in length with a polyadenylated stretch at the 3′ end. Complete nucleotide sequence analysis (Fig. 2) revealed that the coding region of *erg* 2 is the same as *erg* 1 except for an additional 218-bp segment located in the 5′-coding region (Fig. 1). A computer search identified two domains of homology with v-*ets*. Both *erg* 1 and *erg* 2 cDNA clones are ~70% homologous at the 3′ region with the 3′ end of v-*ets* and ~40% homologous at the 5′ region with the 5′ end of v-*ets* (Fig. 1). All the three clones appear to be highly related to each other, but display only a limited domain of homology with v-*ets*.

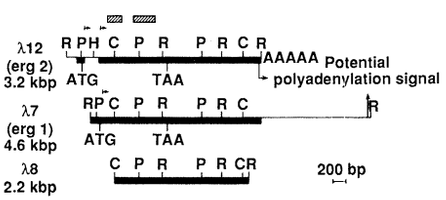

Fig. 1. Comparison of three *erg* cDNA inserts from recombinant λgt10 clones λ7, λ12, and λ8. Locations of restriction endonuclease sites for Eco RI (R), Pvu II (P), Hinc II (H), and Cla I (C) are indicated. The thickened line represents a sequence identical in all the three clones. The hatched areas represent homologous domains corresponding to the 5′ and 3′ regions of v-*ets*. The horizontal arrows (→) represent an 8-bp direct repeat (TCAAGGAA). Initiation (ATG) codon, termination (TAA) codons, and potential polyadenylation signals are also shown.

The longest open reading frame in *erg* 2, starting with a methionine codon at position 264 in the nucleotide sequence (Fig. 2), could encode a 462–amino acid polypeptide with an estimated relative molecular mass of 52,028 daltons. Since three stop codons are present in the same frame further upstream (Fig. 2), the ATG at position 264 may be used as the initiation codon. The only long open reading frame with an initiation codon at nucleotide position 264 is preceded by another initiation codon (nucleotide 192) in the same reading frame upstream of the termination codon that precedes the long open reading frame. Thus, a polypeptide composed of 12 amino acids could also be potentially synthesized from *erg* 2 mRNA in the bicistronic manner proposed for some eukaryotic mRNAs. A similar phenomenon is observed in *erg* 1 mRNA (*16*). In fact, there are roughly nine translation initiation codons at the amino terminal and none of these ATGs are flanked by nucleotides that strictly conform to the consensus for preferred translation initiation sites as deduced by Kozak (*20*), that is $CC\overset{A}{\underset{G}{}}CCAUG(G)$. However, seven of them are preceded by a purine at position −3, a feature that is believed to be important for initiation. The open reading frame is preceded by a 263-bp stretch of noncoding sequences at the 5′ end and followed by an approximately 1564-bp noncoding region at the 3′ end. The 3′ end has a polyadenylated stretch of about 42 bp preceded by a putative polyadenylation signal sequence AAGGAAA, which does not adhere strictly to the consensus sequence, AATAAA. However, the hexanucleotide

V. N. Rao and T. S. Papas, Laboratory of Molecular Oncology, National Cancer Institute, Frederick, MD 21701-1013.
E. S. P. Reddy, Program Resources, Inc., National Cancer Institute, Frederick, MD 21701-1013.

*To whom correspondence should be addressed.

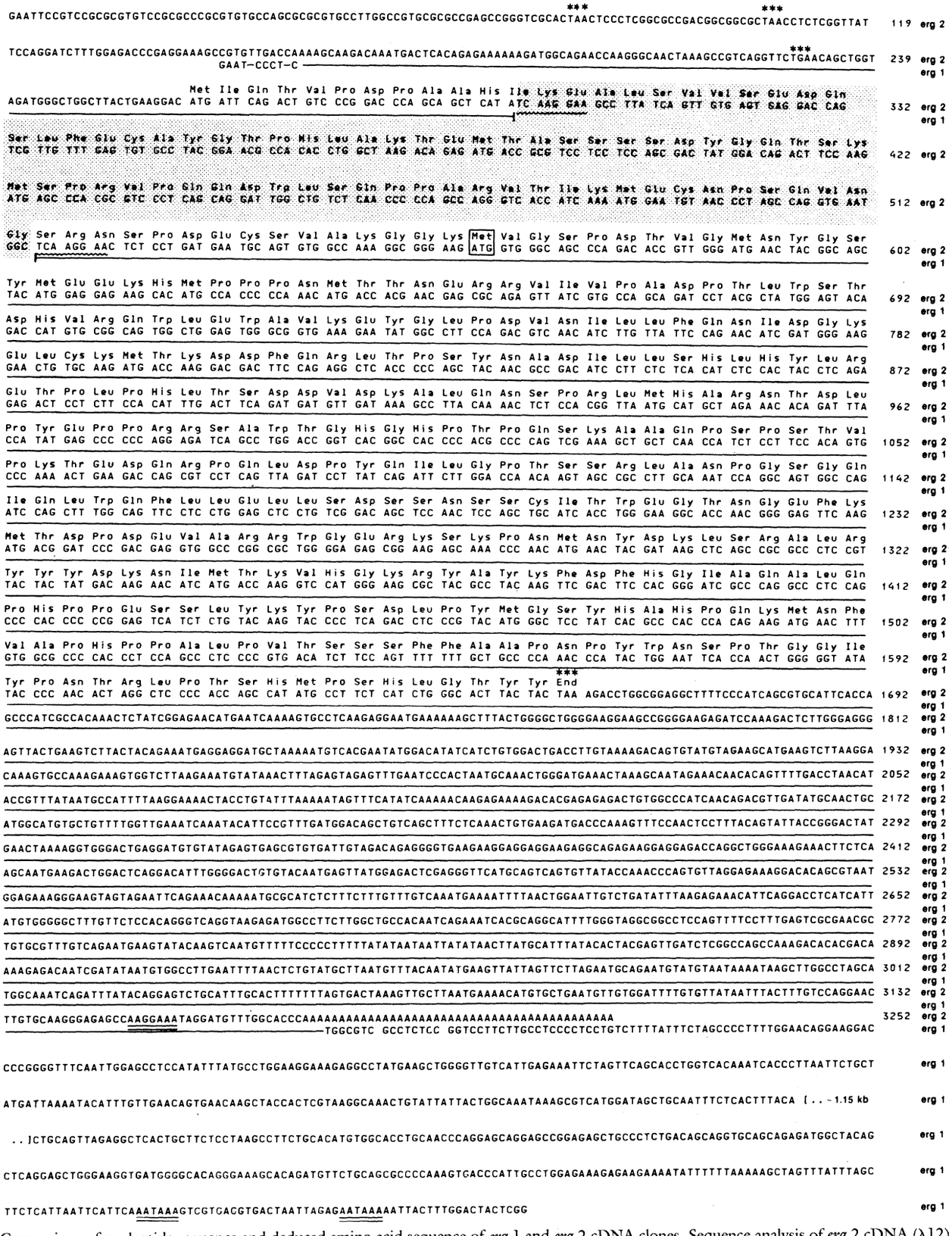

Fig. 2. Comparison of nucleotide sequence and deduced amino acid sequence of *erg* 1 and *erg* 2 cDNA clones. Sequence analysis of *erg* 2 cDNA (λ12) was carried out by the dideoxy chain termination method (*39*) using synthetic oligonucleotide primers and reverse transcriptase. The *erg* 1 cDNA (λ7) sequence is aligned for comparison. The deduced amino acid sequence for the *erg* 2 and *erg* 1 is shown above the corresponding nucleotide sequence. Nucleotide positions for *erg* 2 are indicated on the right, whereas amino acid positions are indicated on the left. The solid line represents the sequence of *erg* 1 that is identical to *erg* 2, unless indicated. The shaded area represents the alternative splice region of *erg* 2 which is absent in *erg* 1. The first seven nucleotides in *erg* 1 and *erg* 2 cDNA reflect a sequence of the synthetic Eco RI linker used for cloning. The four termination codons in-frame with the reading frame upstream and downstream are indicated by asterisks. Wavy lines indicate regions of 8-bp direct repeat sequences. The position of the 8-bp sequence in the *erg* 1 clone with respect to *erg* 2 has been arbitrarily represented. Boxed initiation methionine codon represents the potential methionine initiation codon used in *erg* 1. The double underlines show potential polyadenylation signals in *erg* 1 and *erg* 2.

signal cannot be the only sequence element required for polyadenylation because the same sequence appears in mRNAs without signaling polyadenylation (*21*). The *erg* 1 transcript appears to use a different putative polyadenylation signal, which is 1407 nucleotides downstream from the one used in *erg* 2. Two potential polyadenylation signals appear near the end of 3′ noncoding region of *erg* 1 mRNA (Fig. 2). The other structural feature is that the alternative splice region is flanked on each side by an 8-bp direct repeat, TCAAGGAA. Such direct repeats are contained in many other cellular genes, such as *ral* (*12*) and tropomyosin (*22*). The functional significance, if any, of this repeat remains to be established.

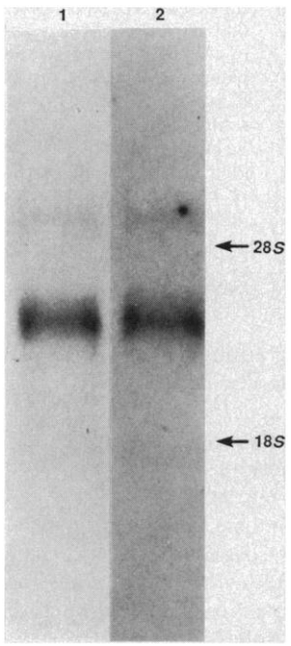

Fig. 3. Analysis of *erg* mRNA in COLO 320 cells. Polyadenylated RNA from COLO 320 cells was separated by denaturing gel electrophoresis in formaldehyde (*19*) and transferred to Nytran filters. Hybridization of the filters with the *erg* probe was performed as suggested by the manufacturer (Schleicher & Schuell). Hybridization of the *erg* probe (^{32}P-labeled 0.95-kb Eco RI fragment) to polyadenylated RNA from COLO 320 cells (lane 2). Same blot was washed and reprobed with *erg* 2–specific probe, a ^{32}P-labeled synthetic oligomer (nucleotides 116 to 215) obtained from the alternative splice region of *erg* 2 cDNA clone (lane 1). RNA sizes were determined by comparison with the ribosomal RNA markers (28*S* and 18*S*) and BRL RNA ladder as size standards.

Polyadenylated RNA from COLO 320 cells was analyzed by Northern blot hybridization to an oligonucleotide probe specific to *erg* 2. A major 3.2- to 3.6-kb band and a minor ~5-kb band was observed (Fig. 3, lane 1). Similar bands were obtained when the probe was a 0.95-kb Eco RI fragment common to *erg* 1 and *erg* 2 (Fig. 3, lane 2). This result suggests that *erg* 2 may form the major species and may occur in different sizes.

The insertion of 218 bp in *erg* 2 because of alternative splicing causes a shift in the open reading frame at the 5′ end, thus facilitating the use of an alternative initiation codon at nucleotide position 264 in *erg* 2. The *erg* 1 transcript appears to use the translation initiation codon, which is 98 amino acids downstream from the potential initiation codon used in *erg* 2 (Fig. 2). The biosynthesis of two similar forms of a given protein as the result of different splicing events producing different mRNA species has been described for a number of genes, both in viral and nonviral systems. The resulting mRNAs produced can differ either in their leader sequence (*23*), in a region coding for an internal exon (*24*), or in their 3′ coding exons (*25*). To confirm the presence of the deduced open reading frame for the *erg* gene products, we subcloned the entire insert (of *erg* 1 and *erg* 2) into GEM riboprobe vectors (Promega Biotec), downstream from the T7 RNA polymerase promoter (Fig. 4), in an orientation intended to generate transcripts with the same chemical polarity as *erg* mRNAs. To show that *erg* 1 uses an alternative initiation codon, we deleted the sequences that code for the upstream methionine initiation codon (used in *erg* 2) and cloned into the GEM-3 vector (Fig. 4).

The RNAs synthesized in vitro by the T7 RNA polymerase were then translated in a rabbit reticulocyte lysate. The largest protein bands detected migrated in SDS–polyacrylamide gel electrophoresis with an apparent molecular mass of approximately 52 kD in the case of *erg* 2 (Fig. 5, lanes 4 and 5) and 41 kD in the case of *erg* 1 (Fig. 5, lane 2); these are close to the sizes predicted from the nucleotide sequence. The translation of RNAs transcribed in vitro from pT7*erg* 2, pT7*erg* 1, and pT7*erg* 1 minus ATG (Fig. 5) yielded a series of polypeptides ranging from the full-length translation product to several smaller species. In the case of pT7*erg* 2, these appear to be derived from initiation of translation at each of the downstream methionine residues (residues 40, 54, 75, 100, 109, 115, 120, and 125). The pT7*erg* 1 minus ATG yielded products of the same size as pT7*erg* 1, further supporting the use of the methionine codon at residue 100 for translational initiation in the case of *erg* 1 protein.

The deduced amino acid sequence of *erg* 2 protein shows that it has a primary length of 462 amino acids, is rich in proline (50 residues) and serine (45 residues), slightly hydrophobic at its amino terminus, and

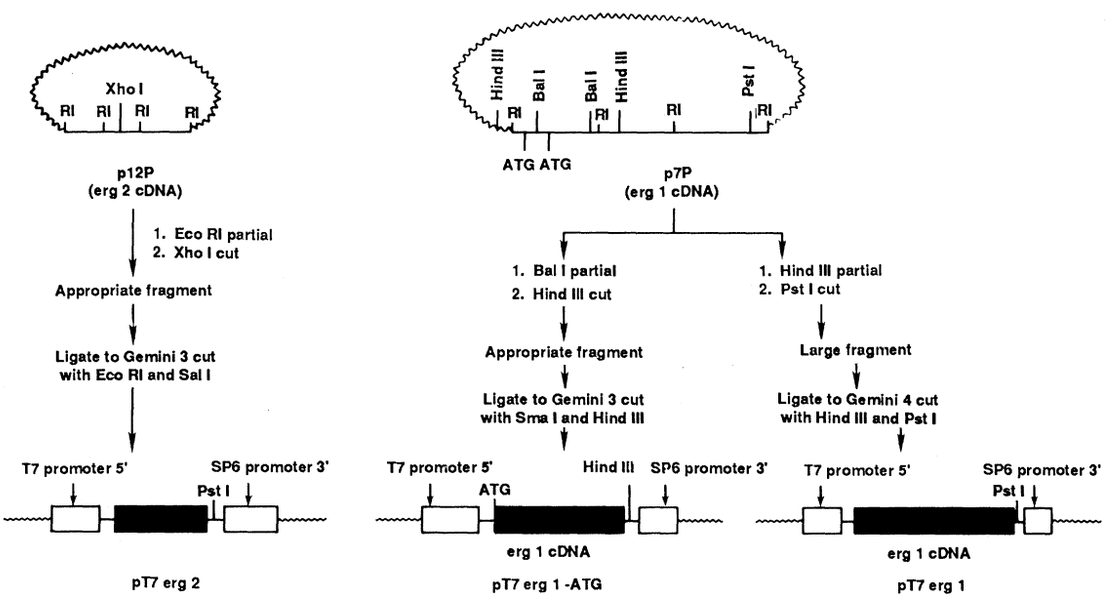

Fig. 4. Schematic representation of the construction of plasmids pT7*erg* 2, pT7*erg* 1 minus ATG, and pT7*erg* 1 used for in vitro transcription-translation.

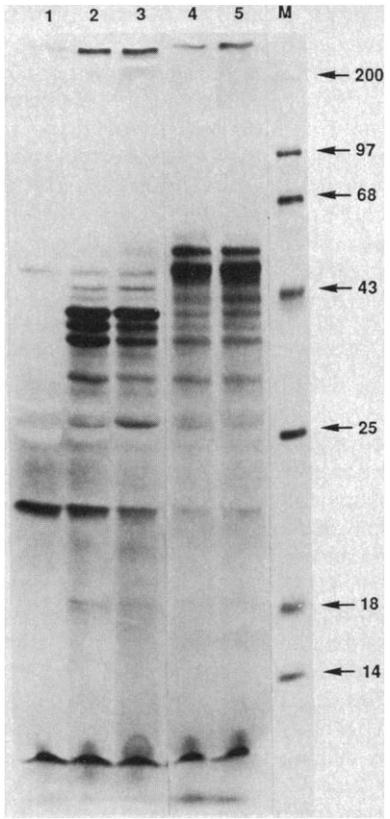

Fig. 5. In vitro synthesis of *erg* proteins. The *erg* 1 and *erg* 2 cDNA clones (cloned in riboprobe GEM vectors; pT7*erg* 2, pT7*erg* 1 minus ATG, and pT7*erg* 1) were linearized by digestion with Pst I and transcribed in vitro with T7 polymerase (Promega) according to the manufacturer's protocols. The uncapped RNAs produced were translated in vitro with a rabbit reticulocyte lysate (Promega) in the presence of [^{35}S]methionine. The labeled proteins were subjected to electrophoresis on a 12% SDS-polyacrylamide gel (*40*) and revealed by fluorography. Lane 1, control minus RNA. Lanes 2 and 3, translation products of RNAs transcribed from pT7*erg* 1 and pT7*erg* 1 minus ATG, respectively. Lanes 4 and 5, translational products of RNA transcribed from pT7*erg* 2. Lane M, molecular mass markers.

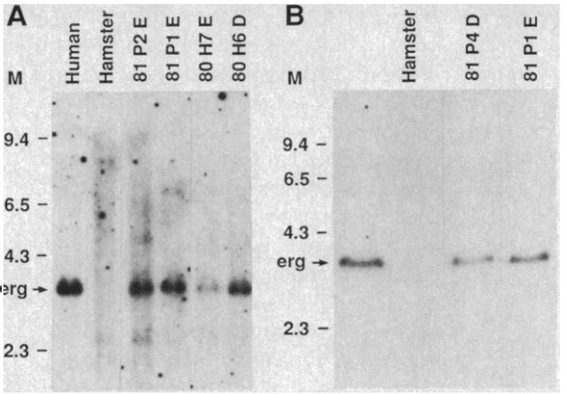

Fig. 6. (**A** and **B**) Hybridization of labeled *erg* DNA to Pst I restriction digests of human, hamster, and human × hamster DNAs (80 and 81 series). Restriction digests containing 5 to 10 μg of DNA were subjected to electrophoresis in 0.8% agarose gels and transferred to Nytran filters. Hybridization with a ^{32}P-labeled *erg* probe (Fig. 3) was performed as described by manufacturer (Schleicher & Schuell) at 42°C in 50% formamide. The arrow indicates a ~3.5-kb fragment specific to an *erg* gene derived from human chromosome 21. Lanes M on the left of (A) and (B) represent DNA size markers in kilobases.

bears four potential glycosylation sites at amino acid positions 78, 83, 124, and 309. The hydrophobicity and potential glycosylation sites may indicate membrane association or secretion.

A computer-assisted search of the NBRF protein database identified several regions of homology between the protein deduced from the sequence of the open reading frame of *erg* 1 and *erg* 2 with other proteins [see (*16, 26*)].

The chromosomal localization of the *erg* gene was determined by Southern blot analysis of DNAs prepared from human × rodent somatic cell hybrids with a ^{32}P-labeled *erg*-specific probe (0.95-kb Eco RI fragment) generated from the *erg* 2 clone. The somatic cell hybrids were constructed as described previously (*27–29*). The *erg* probe used had no homology with Hu-*ets*-2. The hybrids carrying human chromosome 21 all contained the ~3.5-kb Pst I fragment (Fig. 6, A and B) that is characteristic of the *erg* gene locus. The frequency of concordance between *erg* for each human chromosome was calculated with computer assistance (*30*), and the results enabled us to assign *erg* to human chromosome 21. The Hu-*ets*-2 gene is also on chromosome 21. Many human chromosomes are nonrandomly associated with certain types of human cancers (*31–34*). Klein (*33*) has suggested that certain human cancers may result from translocation of a transforming gene to a chromosomal position where the regulation of transcription is altered. Chromosome 21 has been implicated in some diseases such as Down's syndrome and Alzheimer's disease (*35*). Thus, finding the precise location of *erg* and analyzing the *erg* locus in different human cancers, Down's syndrome, and Alzheimer's disease should make it possible to determine if amplification, translocation, or rearrangement of this gene can be linked to any human disease.

The generation of protein diversity from a single gene by alternative RNA processing has been demonstrated previously. For example, certain membrane and secreted forms of immunoglobulins and the calcitonin gene may be derived by the use of a single initiation site for transcription but multiple polyadenylation sites associated with different 3′-end splicing (*36*); myosin light chains may be generated by the use of alternative promoters and a single polyadenylation signal but different 5′ or internal splice choices (*23*); troponin T may be derived by use of a single promoter and single polyadenylation site, but different internal splice choices (*37*); and certain other proteins may be generated by a single transcription initiation site and multiple polyadenylation sites associated with different 3′-end splicing and alternative internal splice choices (*38*). Here we have shown that a fourth type of genomic organization may generate protein diversity via alternative RNA processing. The *erg* gene appears to contain single or multiple transcription initiation sites and multiple polyadenylation sites associated with an alternative internal splice at the 5′ end. Thus, the *erg* proteins may be generated by three distinct mechanisms operating at the splicing, polyadenylation, and translational level. Whether the pattern of expression of the two *erg* mRNAs shows tissue specificity remains to be determined.

REFERENCES AND NOTES

1. K. Radke, H. Beug, S. Kornfeld, T. Graf, *Cell* **31**, 643 (1982).
2. M. G. Moscovici *et al.*, *Virology* **129**, 65 (1983).
3. D. Leprince *et al.*, *Nature (London)* **306**, 395 (1983).
4. M. F. Nunn, P. M. Seeburg, C. Moscovici, P. H. Duesberg, *ibid.* **306**, 391 (1983).
5. D. K. Watson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1792 (1986).
6. ———, *ibid.* **82**, 7294 (1985).
7. M. O. Diaz, M. M. LeBeau, P. Pitha, J. D. Rowley, *Science* **231**, 265 (1986).
8. N. Sacchi *et al.*, *ibid.*, p. 379.
9. J. Ghysdael *et al.*, *EMBO J.* **5**, 2251 (1986).
10. S. Kentaro *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5459 (1986).
11. T. Kawakami, C. Y. Pennington, K. C. Robbins, *Mol. Cell. Biol.* **6**, 4195 (1986).
12. P. Chardin and A. Tavitian, *EMBO J.* **5**, 2203 (1986).
13. P. Madaule and R. Axel, *Cell* **41**, 31 (1985).
14. A. L. Schechter *et al.*, *Nature (London)* **312**, 513 (1984).
15. G. D. Kruh *et al.*, *Science* **234**, 1545 (1986).
16. E. S. P. Reddy, V. N. Rao, T. S. Papas, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
17. T. V. Huynh, R. A. Young, R. W. Davis, in *DNA Cloning*, vol. 1, *A Practical Approach*, D. M. Glover, Ed. (IRL Press, Washington, DC, 1985), pp. 49–78).
18. U. Gubler and B. J. Hoffman, *Gene* **5**, 263 (1983).
19. T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982).
20. M. Kozak, *Nucleic Acids Res.* **12**, 857 (1984).
21. M. Tosi, R. A. Young, O. Hagenbuchle, V. Schibler, *ibid.* **9**, 2313 (1981).
22. D. M. Helfman *et al.*, *Mol. Cell. Biol.* **6**, 3582 (1986).
23. Y. Nabeshima, Y. Fujii-Kuriyama, M. Muramatsu, K. Ogata, *Nature (London)* **308**, 333 (1984).
24. C. R. King and J. Piatigorsky, *Cell* **32**, 707 (1983).
25. M. Kress, D. Glaros, G. Khoury, J. Gilbert, *Nature (London)* **306**, 602 (1983).
26. A computer search on the 99 amino acids specific to the *erg* 2 polypeptide, which has no homology with the v-*ets* oncogene, gave a low homology score with several proteins, 18% homology with human T cell

receptor α chain over a range of 34 amino acids, 21% homology over the entire 99–amino acid stretch with the *gag* polyprotein of the human immunodeficiency virus (HIV), about 24% homology (over a 62–amino acid range) with the genome polyprotein of poliovirus types 1 and 3, and 29% homology with the human *fos* transforming protein over a range of 31 amino acids, the significance of which, if any, remains to be established.
27. S. J. O'Brien *et al.*, *Nature (London)* **303**, 74 (1983).
28. S. J. O'Brien, J. M. Simonson, M. A. Eichelberger, in *Techniques in Somatic Cell Genetics*, J. W. Shay, Ed. (Plenum, New York, 1982), pp. 342–370.
29. W. G. Nash and S. J. O'Brien, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6631 (1982).
30. S. J. O'Brien and W. G. Nash, *Science* **216**, 257 (1982).
31. J. J. Mulvihill and S. M. Robinette, *Genet. Maps* **2**, 356 (1982).
32. R. Berger, A. Bernheim, A. Chapelle, *Cytogenet. Cell Genet.* **32**, 205 (1982).
33. G. Klein, *Nature (London)* **294**, 313 (1981).
34. J. J. Yunis *et al.*, *N. Engl. J. Med.* **307**, 1231 (1982).
35. D. J. Selkoe, D. S. Bell, M. B. Podlisny, D. L. Price, L. C. Cork, *Science* **235**, 873 (1987); D. Goldgaber *et al.*, *ibid.*, p. 877; R. E. Tanzi *et al.*, *ibid.*, p. 880; P. H. St George-Hyslop *et al.*, *ibid.*, p. 885.
36. P. Early *et al.*, *Cell* **20**, 313 (1980); M. G. Rosenfeld *et al.*, *Nature (London)* **304**, 129 (1983).
37. R. M. Medford *et al.*, *Cell* **38**, 409 (1984).
38. D. M. Helfman *et al.*, *Mol. Cell. Biol.* **6**, 3582 (1986).
39. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
40. U. K. Laemmli, *Nature (London)* **222**, 680 (1970).
41. We thank S. J. O'Brien for discussion and for providing somatic cell hybrid panels, D. Watson for the Hu-*ets* 2 cDNA clone, L. Lee for synthesizing oligonucleotide primers, K. McNitt and G. Smythers for assistance with the computer analysis, S. L. Chen for technical assistance, and K. Cannon for typing the manuscript.

9 March 1987; accepted 12 June 1987

Naturally Acquired Antibodies to Sporozoites Do Not Prevent Malaria: Vaccine Development Implications

STEPHEN L. HOFFMAN,* CHARLES N. OSTER, CHRISTOPHER V. PLOWE, GILLIAN R. WOOLLETT, JOHN C. BEIER, JEFFREY D. CHULAY, ROBERT A. WIRTZ, MICHAEL R. HOLLINGDALE, MUTUMA MUGAMBI

The first human vaccines against the malaria parasite have been designed to elicit antibodies to the circumsporozoite protein of *Plasmodium falciparum*. However, it is not known whether any level of naturally acquired antibodies to the circumsporozoite protein can predict resistance to *Plasmodium falciparum* malaria. In this study, 83 adults in a malaria-endemic region of Kenya were tested for circumsporozoite antibodies and then treated for malaria. They were monitored for the development of new malaria infections for 98 days. Antibody levels, as determined by four assays in vitro, were indistinguishable between the 60 individuals who did and the 23 who did not develop parasitemia during follow-up, and there was no apparent relation between day of onset of parasitemia and level of antibodies to circumsporozoite protein. Unless immunization with sporozoite vaccines induces antibodies that are quantitatively or qualitatively superior to the circumsporozoite antibodies in these adults, it is unlikely that such antibodies will prevent infection in areas with as intense malaria transmission as western Kenya.

WHEN A FEMALE ANOPHELINE mosquito carrying malaria sporozoites in her salivary gland feeds on a human, sporozoites pass into the bloodstream. Some of these sporozoites are transported to the liver, where they invade hepatocytes, initiating the cycle that culminates in malarial disease.

Sporozoites are covered by a polypeptide called the circumsporozoite (CS) protein. The CS protein of *Plasmodium falciparum* contains a large, immunodominant central domain consisting of 41 tandemly repeated tetrapeptides (37 Asn-Ala-Asn-Pro and 4 Asn-Val-Asp-Pro) (*1*). Development of human subunit sporozoite vaccines has been based on the hypothesis that antibodies to the repeat region of the *P. falciparum* CS protein will prevent sporozoite infection of hepatocytes and the subsequent development of malaria (*1–4*). The primary support for this hypothesis came from murine studies that showed that passive transfer of monoclonal antibodies to *P. berghei* sporozoites and polyclonal antibodies to synthetic peptides from the *P. berghei* repeat region protected mice against sporozoite challenge (*5, 6*). In a recent clinical trial of a recombinant DNA-produced malaria vaccine, human volunteers with vaccine-induced antibodies to the repeat region of the *P. falciparum* CS protein were challenged by the bite of five infected mosquitoes. The individual with the highest level of antibodies was protected against malaria infection, and two individuals with lower levels of antibodies had a delayed onset of parasitemia compared to controls (*7*). This study confirmed the hypothesis that humans can be protected from sporozoite infection by immunization with a subunit vaccine, and strongly suggested that the protection was mediated by antibodies.

Epidemiologic surveys in malaria-endemic areas have repeatedly shown that the prevalence of malaria decreases with increasing age, indicating that adults develop some protective immunity. Since most adults in such areas have antibodies to sporozoites (*8*), including antibodies to the CS protein repeat region (*9*), these antibodies have been considered to be possible mediators of protective immunity. In a retrospective study of sera from Indonesians living in a malaria-endemic area, the age-specific prevalence and titers of antibodies to the CS protein were inversely correlated with the age-specific prevalence of *P. falciparum* malaria. Affinity-purified antibodies to the repeat region of the CS protein derived from these sera mediated the circumsporozoite precipitation (CSP) reaction and inhibited sporozoite invasion of hepatoma cells (ISI) (*9*). These results were interpreted as supporting the hypothesis that antibodies to the CS protein repeat region protect against sporozoite infection. We now report the results of a prospective study designed to determine whether the level of naturally acquired antibodies to the CS protein would be predictive of protection against *P. falciparum* infection during a 98-day period when the rate of malaria transmission was high.

The study was conducted in Saradidi, western Kenya. In April 1986, 4 weeks prior to the peak malaria transmission season (*10*), 93 adult, male, lifelong residents of Saradidi volunteered for the study. Blood samples for serum and malaria smears were obtained from each volunteer, each of whom was then treated with a single dose of three tablets of pyrimethamine/sulfadoxine (Fansidar, Roche), followed by 100 mg of doxycycline twice daily for 7 days. On the basis of local parasite susceptibility to Fansidar alone, radical cures were expected for all volunteers. The home of each volunteer was visited daily for the next 98 days and blood smears were obtained on days 7, 14, 28, 42, 56, 70, 84, 98, and on any day a volunteer complained of illness. If malaria parasites

S. L. Hoffman, Infectious Diseases Department, Naval Medical Research Institute, Bethesda, MD 20814-5055, and Department of Immunology, Walter Reed Army Institute of Research, Washington, DC 20307-5100.
C. N. Oster and J. C. Beier, Kenya Medical Research Institute, U.S. Army Medical Research Unit–Kenya, Nairobi, Kenya.
C. V. Plowe, Cornell University Medical College, New York, NY 10021.
G. R. Woollett and M. R. Hollingdale, Biomedical Research Institute, Rockville, MD 20852.
J. D. Chulay, Department of Immunology, Walter Reed Army Institute of Research, Washington, DC 20307-5100.
R. A. Wirtz, Department of Entomology, Walter Reed Army Institute of Research, Washington, DC 20307-5100.
M. Mugambi, Kenya Medical Research Institute, Nairobi, Kenya.

*To whom correspondence should be addressed.