

mologous to avian and mammalian chromosomal DNA sequences (*c-erbA* and *c-erbB*) (4, 13, 22) that may play a role in the metabolism of normal cells, possibly in the differentiation of immature red blood cells (23).

Wild-type AEV-transformed erythroblasts are tightly blocked in their maturation (at the colony-forming unit stage) (24, 25). Studies involving deletion mutants in *erbA* and *erbB* indicate that *erbB*, in vitro as well as in vivo, yields transformed erythroblast-like cells at different stages of maturation, whereas *erbA* alone induces no transformation.

Thus *erbA* potentiates the transforming activity of *erbB* and appears to be responsible for the early blockage of cell differentiation within the erythroid lineage (26–29).

Although several oncogenes have been extensively studied, “potentializers” such as *erbA* have not been studied. Therefore we analyzed the nucleotide sequence of *erbA*. Computer analysis showed that the deduced amino acid sequence of the *erbA* stretch of P75^{gag-erbA} is clearly different from the sequences of other reported transforming proteins. Thus *erbA* appears to represent a distinct new member of the oncogene families.

represent a distinct new member of the oncogene families.

A molecular clone of AEV (p-AEV 11) (30) with biological activity was used to sequence the *erbA* oncogene by the Maxam and Gilbert procedure (31). Figure 1A shows the strategy used to produce the nucleotide sequence (Fig. 2) defining the *erbA* boundaries (Fig. 1) between the structural *gag* gene upstream and the second oncogene, *erbB*, downstream. The left boundary of *erbA* was assessed by comparison to the nucleotide sequence of the Prague-C strain (Pr-C) of Rous sarcoma virus (32). The

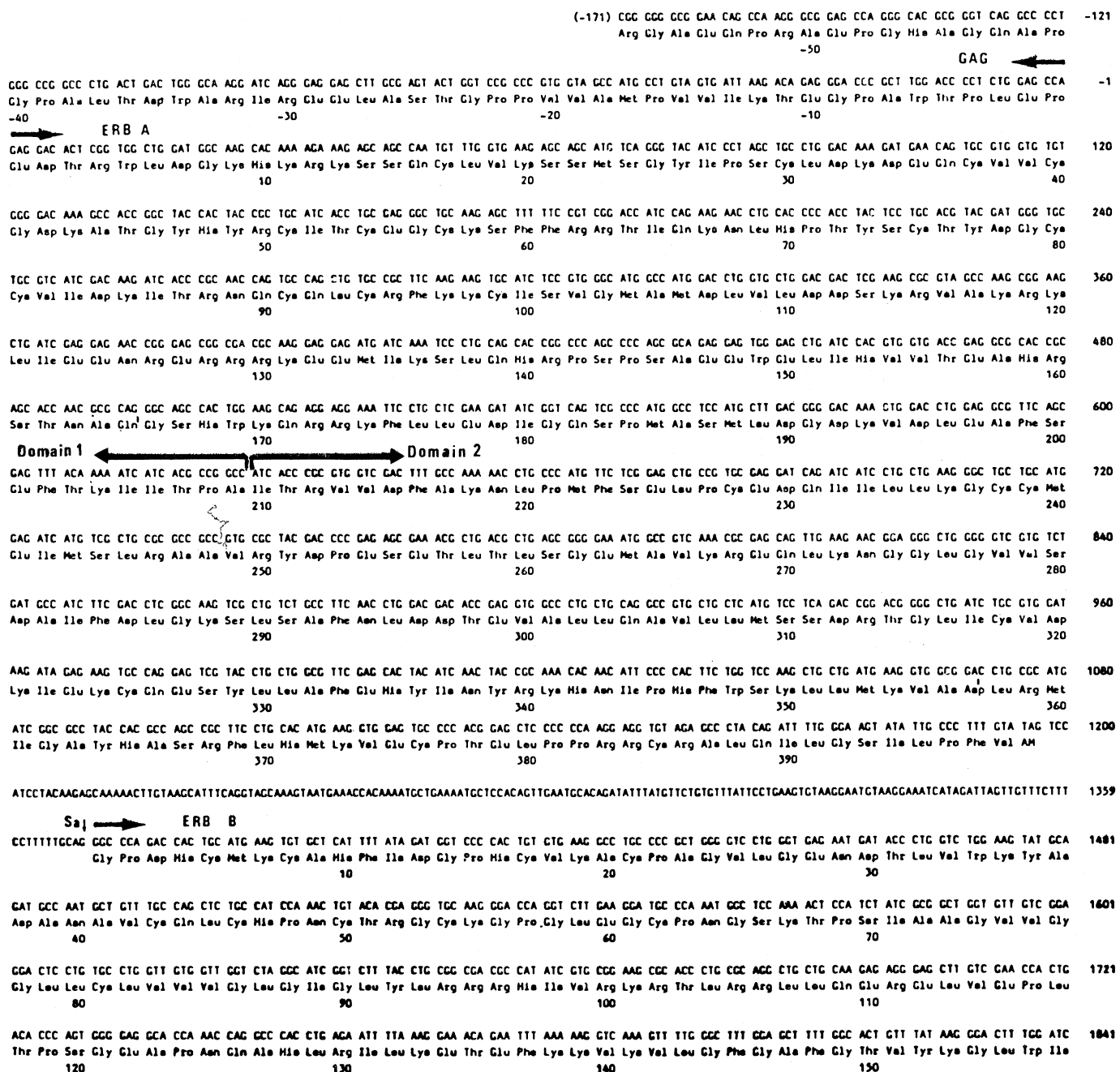


Fig. 2. Nucleotide sequence of the *erbA* gene and of the 5' end of *erbB*. The sequence of 2012 nucleotides encompassing the *erbA* gene is shown. The deduced amino acid sequence of the *erbA* domain of P75^{gag-erbA} is indicated from nucleotides 1 to 1194 and the deduced amino acid sequence of the 5' end of *erbB* is indicated from the putative splice acceptor site (Sa, nucleotide 1370) to nucleotide 1841. Nucleotides are numbered in the right column and every tenth amino acid is numbered.

Ava I site at the extreme left of our sequence (Fig. 1) is within the coding domain for the virion protein p10^{gag}. The nucleotide sequences of Pr-C and AEV are almost identical from this position rightward to residue 1 in the AEV sequence, where complete divergence of the two sequences marks the point of insertion of *erbA*. This insertion is located within the coding domain for the virion protein p27^{gag} (32).

The reading frame of *erbA* continues uninterrupted from the p27^{gag} reading frame at position +1 until it is terminated by an amber codon at position 1195; this is followed by a stretch of noncoding sequences up to a putative splice acceptor (Sa in Fig. 2) site (33), TTTCC-TTTTTCAG.G (T, thymine; C, cytosine; G, guanine; A, adenine) for the *erbB* gene at nucleotide 1370. It is unlikely that the other two *erbA* reading frames are used because they are frequently closed by termination codons (vertical bars in Fig. 1B). The consensus splice acceptor site at position 1370 could generate the subgenomic *erbB* mRNA in an open reading frame different from that of *erbA*, and we present the deduced partial amino acid sequence of this reading frame (Fig. 2). Thus the *erbB* product may start at the AUG (U, uracil) codon of *gag* used to produce P75^{gag-erbA} (assuming that AEV and Pr-C use the same splice donor site in *gag*); the two proteins then share a few common amino acids at their NH₂ terminus. Alternatively, the *erbB* product could initiate at the in-frame AUG codon at position 1386 in

our sequence (Fig. 2). The structure of *erbB* is virtually identical, in the region we studied, to the one recently described for another independent isolate of AEV, namely AEV-H, which lacks the *erbA* oncogene (34). Importantly, *erbB* is closely related to the *src* gene of avian sarcoma virus [(34) and our sequence data] and to a portion of the epidermal growth factor receptor (35).

The deduced amino acid sequence of *erbA* (398 amino acid residues) is shown under the nucleotide sequence in Fig. 2. The calculated molecular size of the *erbA* polypeptide is 45.4 kD, yielding for P75^{gag-erbA} (36) a deduced molecular size of 72 kD. Cellular adenosine 3',5'-monophosphate-dependent protein kinases phosphorylate serine or threonine residues within sequences x-y-z-(Ser or Thr) (37), where x and y are basic residues; two such serine residues are found in the *erbA* stretch at positions 14 and 15. There are no potential glycosylation sites Asn-x-(Thr or Ser) (38) in this polypeptide. Thus *erbA* may be phosphorylated but not glycosylated. Two domains can be defined within the *erbA* protein; notable features of the amino acid composition are found in the first domain of *erbA* (amino acids 1 to 209 in Fig. 2) with elevated levels of cysteine and basic residues (10 percent and 21 percent, respectively, within amino acids 1 to 131), in strong contrast with the rest of the molecule. Hydrophilicity studies with the Hopp-Woods procedure (39) indicate that *erbA* encodes a relatively hydrophilic product (the value for the *erbA* poly-

peptide is 0.1, whereas the average protein has a net hydrophilicity of 0.07), as expected for a cytoplasmic protein (21).

A search for similarities between the *erbA* protein and the other proteins contained in the protein data bases (40, 41) revealed no obvious relationships with other oncogenes. Thus *erbA* may be a genuinely new member of this class of proteins. Domain 1 showed no salient homology with other known proteins, but a relatedness was found between the carboxyl terminal half of *erbA* (domain 2 in Fig. 2) and the carbonic anhydrase family (Fig. 3). The homology begins with *erbA* amino acid residue 219 (amino acid residue 10 for the carbonic anhydrases) and extends across 180 residues up to the carboxyl terminus of the molecule (residue 398 of *erbA* or residue 195 of the carbonic anhydrases, which continue for another 64 residues). When compared with the four mammalian carbonic anhydrases II available in the data banks we used, the overall homology was 27 percent for 180 amino acids, clustered in some portions of the molecule. For example, close to one of the known active sites of carbonic anhydrases (42), the homology increases to more than 51 percent (*erbA* residues 248 to 274), although the two residues reported to be important in the active site (His⁶³ and Asn⁶⁶) (42) are not found in the *erbA* protein. This may imply a structural, but not a functional relationship. Another example is *erbA* residues 310 to 330, with more than 40 percent homology. This homology between the



Fig. 3. Relatedness of *erbA* domain 2 to the carbonic anhydrase family. Using the sequence of the putative *erbA* polypeptide, we performed exhaustive homology searches on the NBRF (41) and NEWAT (40) protein data banks with our computer system (47). These searches revealed a homology encompassing 180 amino acid residues between *erbA* domain 2 and the carbonic anhydrase family. The significance of this homology was further assessed by an alignment program (48) and adapted by one of us (J.M.C.). The amino acid sequences of the predicted *erbA* polypeptide and of sheep (CRSH2) (49), rabbit (CRRB2) (50), bovine (CRBO2) (51), and human (CRHU2) (42) carbonic anhydrases II were compared. The optimal alignment was found when we used a deletion weight of 1 for each gap plus 0.2 times the number of residues in each gap. In this case, the Needleman-Wunsch similarity value (52) was found 3.8 standard deviation units above the mean of 30 shuffled sequences. Scores that are 3 or more standard deviation units above the mean can reasonably be expected to represent authentic relationships (53). Common residues are boxed. Closed circles indicate amino acids known to be important in the active site of carbonic anhydrases. The one-letter symbols for the amino acids are A, alanine; R, arginine; N, asparagine; D, aspartic acid; C, cysteine; Q, glutamine; E, glutamic acid; G, glycine; H, histidine; I, isoleucine; L, leucine; K, lysine; M, methionine; F, phenylalanine; P, proline; S, serine; T, threonine; W, tryptophan; Y, tyrosine; V, valine; B, asparagine or aspartic acid; and Z, glutamine or glutamic acid.

erbA product and, for example, human carbonic anhydrase II (CRHU 2 in Fig. 3) is 3.8 standard deviation units above the mean of 30 shuffled sequences, considerably higher than the cutoff value of 3, which indicates, with statistical significance, an evolutionary relatedness (40) (for details of the computer program see Fig. 3). Moreover, in human carbonic anhydrase I, which has a 61 percent homology with human carbonic anhydrase II, most of the homologies depicted in Fig. 3 are conserved (not shown). Although *erbA* domain 2 is related to carbonic anhydrases, we showed that it is not the enzyme itself. Using a viral *erbA* DNA probe corresponding to domain 2, we screened both chicken and human genomic DNA libraries and found that normal DNA in both species contained a similar locus. The human cellular *erbA* domain 2 was cloned, and a nucleotide sequence was determined (corresponding to nucleotides 675 to 936 in the *erbA* domain 2 of Fig. 2). Homology at the nucleotide level was 83 percent, and the deduced amino acid sequence in the corresponding open reading frame showed 96 percent homology (the two other reading frames were closed by stop codons) (43). The homologies of the human *erbA* domain 2 with human carbonic anhydrase remained but were not augmented. Thus the *erbA* gene is related to but not identical to known carbonic anhydrases.

In conclusion, AEV is an unusual retrovirus in that it can specifically block the maturation of erythroid cells at an immature stage of differentiation. This is achieved by the synergistic action of two distinct oncogenes, *erbA* and *erbB*. The *erbB* product is a membrane glycoprotein (20, 21) that is homologous to a portion of the epidermal growth factor receptor (35) and shares extensive homology with the *src* oncogene family [(34) and our results]. This latter homology is also functional, since both *erbB* in the absence of *erbA* and other *src* family viral genes such as *src* or *fps* (44) can transform erythroblasts (28). Such cells require complex growth conditions and partially differentiate into mature erythrocytes in vitro (28). In contrast, *erbA* in combination with *erbB* is capable of ar-

resting erythroid leukemic cells at an early stage of differentiation where they are able to grow in simple tissue culture media unsuitable for normal erythroid precursors (45, 46).

The finding that the deduced *erbA* protein bears statistically significant relatedness, but not identity, with carbonic anhydrases is interesting because such enzymes play a fundamental role in the CO₂ transport by erythrocytes, the precursors of which are precisely the main targets of AEV.

Our finding of this new type of oncogene leads to the idea that some *src*-related oncogenes found—with the exception of AEV—to be single transforming genes in retroviruses and to transform mainly fibroblasts can affect early hematopoietic differentiation when acting in synergy with a specific activated cell-derived gene such as *erbA*.

B. DEBUIRE*

C. HENRY

M. BENAÏSSA

G. BISERTE

U 124 INSERM, Institut de
Recherches sur le Cancer de Lille,
Faculté de Médecine, Place de Verdun,
59045 Lille Cedex France

J. M. CLAVERIE

Unité d'Informatique, Scientifique
Institut Pasteur, 25 Rue du Docteur
Roux, 75724 Paris Cedex 15 France

S. SAULE

P. MARTIN

D. STEHELIN

U 186 INSERM, Institut Pasteur,
15 Rue C. Guérin, 59019
Lille Cedex France

References and Notes

1. H. Land, L. F. Parada, R. A. Weinberg, *Nature (London)* **304**, 596 (1983).
2. H. E. Ruley, *ibid.*, p. 602.
3. R. F. Newbold and R. W. Overell, *ibid.*, p. 648.
4. B. Vennstrom and J. M. Bishop, *Cell* **28**, 135 (1982).
5. D. LePrince, A. Gégonne, J. Coll, C. de Taisne, A. Schneeberger, C. Lagron, D. Stehelin, *Nature (London)* **306**, 395 (1983).
6. J. Coll *et al.*, *EMBO J.* **2**, 2189 (1983).
7. T. Graf, B. Royer-Pokora, G. E. Schubert, H. Beug, *Virology* **71**, 423 (1976).
8. T. Graf and H. Beug, *Biochim. Biophys. Acta* **516**, 269 (1978).
9. K. Bister and P. H. Duesberg, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5023 (1979).
10. M. M. C. Lai, S. S. F. Hu, P. K. Vogt, *Virology* **97**, 366 (1979).
11. M. Roussel *et al.*, *Nature (London)* **281**, 452 (1979).
12. B. Vennstrom, L. Fanshier, C. Moscovici, J. M. Bishop, *J. Virol.* **36**, 575 (1980).
13. S. Saule, M. Roussel, C. Lagrou, D. Stehelin, *ibid.* **38**, 409 (1981).
14. J. M. Coffin *et al.*, *ibid.* **40**, 953 (1981).
15. D. Sheiness, B. Vennstrom, J. M. Bishop, *Cell* **23**, 291 (1981).
16. S. M. Anderson, W. S. Hayward, B. G. Neel, H. Hanafusa, *J. Virol.* **36**, 676 (1980).
17. M. M. C. Lai, J. C. Neil, P. K. Vogt, *Virology* **100**, 475 (1980).
18. T. Pawson and G. S. Martin, *J. Virol.* **34**, 280 (1980).
19. M. L. Privalsky and J. M. Bishop, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3958 (1982).
20. M. L. Privalsky, L. Sealy, J. M. Bishop, J. P. McGrath, A. D. Levinson, *Cell* **32**, 1257 (1983).
21. M. J. Hayman *et al.*, *ibid.* p. 579.
22. A. Sergeant *et al.*, *EMBO J.* **1**, 237 (1982).
23. J. Coll *et al.*, *Exp. Cell. Res.* **149**, 151 (1983).
24. Y. Samarut and L. Gazzolo, *Cell* **28**, 921 (1982).
25. H. Beug, A. Von Kirchbach, G. Doderlein, J. F. Conscience, T. Graf, *ibid.* **18**, 375 (1979).
26. L. Gazzolo, J. Samarut, M. Bouadbelli, J. P. Blanchet, *ibid.* **22**, 683 (1980).
27. J. K. T. Fung, W. G. Lewis, L. B. Crittenden, H. J. Kung, *ibid.* **33**, 357 (1983).
28. L. Frykberg *et al.*, *ibid.* **32**, 227 (1983).
29. T. Graf and H. Beug, *ibid.* **34**, 7 (1983).
30. B. Vennstrom, L. Fanshier, C. Moscovici, J. M. Bishop, *J. Virol.* **36**, 575 (1980).
31. A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
32. D. Schwartz, R. Tizard, W. Gilbert, *Cell* **32**, 853 (1983).
33. S. M. Mount, *Nucleic Acids Res.* **10**, 459 (1982).
34. T. Yamamoto *et al.*, *Cell* **35**, 71 (1983).
35. J. Downward *et al.*, *Nature (London)* **307**, 521 (1984).
36. M. J. Hayman, B. Royer-Pokora, T. Graf, *Virology* **92**, 31 (1979).
37. E. G. Krebs and J. A. Beavo, *Annu. Rev. Biochem.* **48**, 923 (1979).
38. R. D. Marshall, *Biochem. Soc. Symp.* **40**, 17 (1974).
39. T. P. Hopp and K. R. Woods, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3824 (1981).
40. R. F. Doolittle, *Science* **214**, 149 (1981).
41. M. O. Dayhoff, R. M. Schwartz, B. C. Orcutt, *Protein Sequence Database Release 83* (National Biomedical Research Foundation, Washington, D.C., 1983), document PRSD-0183C.
42. A. Liljas *et al.*, *Nature (London) New Biol.* **235**, 131 (1972).
43. S. Saule, C. Henry, P. Martin, in preparation.
44. P. Kahn and T. Graf, personal communication.
45. H. Beug *et al.*, *Cell* **28**, 907 (1982).
46. H. Beug and M. J. Hayman, *ibid.*, in press.
47. J. M. Claverie, *Nucleic Acids Res.* **12**, 397 (1984).
48. T. F. Smith, M. S. Waterman, W. M. Fitch, *J. Mol. Evol.* **18**, 385 (1981).
49. R. J. Tanis, R. E. Ferrell, R. E. Tashian, *Biochim. Biophys. Acta* **371**, 534 (1974).
50. R. E. Ferrell, S. K. Stroup, R. J. Tanis, R. E. Tashian, *ibid.* **533**, 1 (1978).
51. M. Sciaky, N. Limozin, D. Filippi-Foveau, J. M. Gulian, G. Laurent-Tabusse, *Biochimie* **58**, 1071 (1976).
52. S. B. Needleman and C. D. Wunsch, *J. Mol. Biol.* **48**, 443 (1970).
53. W. C. Barker and M. O. Dayhoff, in *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C., 1972), vol. 5, pp. 101–110.
54. We thank B. Vennstrom and J. M. Bishop for providing the clone p-AEV 11, A. Begue and A. Schneeberger for helpful advice and discussion, and N. Devassine for manuscript preparation. Supported by the Institut National de la Santé et de la Recherche Médicale (U 124 CRL 822020 and U 186 CRL 822033); Institut de Recherches sur le Cancer de Lille, Institut Pasteur de Lille, and private foundations (Association pour la Recherche sur le Cancer and Fondation pour la Recherche médicale).

* To whom requests for reprints should be addressed.

13 February 1984; accepted 8 May 1984