Received 13 November 1987; accepted 2 February 1988.

1. Medawar, P. B. *Transplantation* **1**, 21-38 (1963).
2. Majoor, G. D. & van Breda Vriesman, P. J. C. *Transplantation* **41**, 42-96 (1986).
3. Wood, K. J. & Morris, P. J. *Transplantation* (in the press.).
4. Nagata, M. *et al. Transplantation* **38**, 522-526 (1984).
5. Faustman, D. L., Lacy, P. E., Davie, J. M. & Hauptfeld, V. *Transplant. Proc.* **15**, 1341-1343 (1983).
6. Wood, M. L. & Monaco, A. P. *Transplantation* **37**, 35-42 (1984).
7. Guttmann, R. D., Forbes, R. D. C., Cramer, D. V. & Gill, T. J. *Transplantation* **30**, 216-218 (1980).
8. Soulillou, J. P., Blandin, F., Günther, E. & Lemoine, V. *Transplantation* **38**, 63-67 (1984).
9. Wood, K. J., Evins, J. & Morris, P. J. *Transplantation* **39**, 56-61 (1985).
10. Wood, P., Horsburgh, T. & Brent, L. *Transplantation* **31**, 8-14 (1981).
11. Townsend, A. R. M., Taylor, P. M., Mellor, A. L. & Askonas, B. A. *Immunogenetics* **17**, 283-294 (1983).
12. Lechler, R. I., Ronchese, F., Braunstein, N. S. & Germain, R. N. *J. exp. Med.* **163**, 678-696 (1986).
13. McKenzie, I. F. C. & Henning, M. M. *J. Immunogen.* **4**, 249-257 (1977).
14. Streilein, J. W. & Klein, J. *J. Immun.* **119**, 2147-2150 (1977).
15. Wallich, R. *et al. Nature* **315**, 301-305 (1985).
16. Lechler, R. I. & Batchelor, J. R. *J. exp. Med.* **155**, 31-41 (1982).
17. Hutchinson, I. V. *Transplantation* **41**, 547-555 (1986).
18. Opelz, G. *et al. Transplantation Proc.* **17**, 2357-2361 (1985).
19. Hutchinson, I. V. & Morris, P. J. *Transplantation* **41**, 166-169 (1986).
20. Spencer, S. C. & Fabre, J. W. *Transplantation* **44**, 141-148 (1987).
21. Kohler, G. *et al. Immune Syst.* **2**, 202-208 (1980).
22. Steinman, R. M. *et al. J. exp. Med.* **159**, 1248-1255 (1980).
23. Ozato, K. & Sachs, D. H. *J. Immun.* **126**, 317-321 (1981).
24. Mellor, A. L. *et al. Nature* **298**, 529-534 (1982).
25. Lechler, R. I., Norcross, M. A. & Germain, R. N. *J. Immun.* **135**, 2914-2922 (1985).
26. Mason, D. W. & Williams, A. F. *Biochem. J.* **187**, 1-20 (1980).
27. Corry, R. J., Winn, H. J. & Russell, P. R. *Transplantn Proc.* **5**, 733-735 (1973).
28. Superina, R. A., Peugh, W. N., Wood, K. J. & Morris, P. J. *Transplantation* **42**, 226-227 (1986).

Haemophilia A resulting from *de novo* insertion of L*1* sequences represents a novel mechanism for mutation in man

Haig H. Kazazian Jr, Corinne Wong, Hagop Youssoufian*, Alan F. Scott†, Deborah G. Phillips & Stylianos E. Antonarakis

Genetics Unit, Department of Pediatrics, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA
† Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

L*1* sequences are a human-specific family of long, interspersed, repetitive elements, present as ~10^5 copies dispersed throughout the genome[1]. The full-length L*1* sequence is 6.1 kilobases, but the majority of L*1* elements are truncated at the 5′ end, resulting in a fivefold higher copy number of 3′ sequences[1]. The nucleotide sequence of L*1* elements includes an A-rich 3′ end and two long open reading frames (*orf*-1 and *orf*-2), the second of which encodes a potential polypeptide having sequence homology with the reverse transcriptases[1–4]. This structure suggests that L*1* elements represent a class of non-viral retrotransposons[1,2]. A number of L*1* complementary DNAs, including a nearly full-length element, have been isolated from an undifferentiated teratocarcinoma cell line[5]. We now report insertions of L*1* elements into exon 14 of the factor VIII gene in two of 240 unrelated patients with haemophilia A. Both of these insertions (3.8 and 2.3 kilobases respectively) contain 3′ portions of the L*1* sequence, including the poly (A) tract, and create target site duplications of at least 12 and 13 nucleotides of the factor VIII gene. In addition, their 3′-trailer sequences following *orf*-2 are nearly identical to the consensus sequence of L*1* cDNAs (ref. 6). These results indicate that certain L*1* sequences in man can be dispersed, presumably by an RNA intermediate, and cause disease by insertional mutation.

Haemophilia A is an X-linked disorder of blood coagulation caused by deficiency of factor VIII. The factor VIII gene has been cloned and characterized, and a number of gene defects producing haemophilia A have been identified[7–11]. We have screened the DNA of 240 unrelated males with haemophilia A for any abnormalities in the factor VIII gene detectable by restriction analysis. Significant deletions were found in 15 patients and point mutations in 12 (ref. 11 and unpublished results). In addition, the DNA of two unrelated patients had insertions in exon 14 of the factor VIII gene. In patient 1 (JH-27), exon 14 fragments were increased in size by 3.8 kilobases (kb) and in patient 2 (JH-28) by 2.3 kb (Fig. 1). No abnormalities were detected in the factor VIII genes in the parents of both patients (Fig. 1), suggesting that in both cases the mutations had arisen *de novo*, although maternal mosaicism cannot be excluded. In both patients the disease was severe and no factor VIII activity was detectable.

To characterize these insertions, abnormal 3.5-kb and 6.5-kb *Eco*RI fragments containing the 5′ and 3′ ends of the insertion of JH-27, and the abnormal 6.9-kb *Eco*RI fragment containing the entire insertion of JH-28, were cloned in λgt10 (ref. 12). Pertinent fragments containing the breakpoints and internal portions of the insertions were subcloned and sequenced[13] (Fig. 2).

The insertions of JH-27 contained a 3.8-kb truncated L*1* sequence from nucleotide 2,363 of the genomic L*1* consensus sequence[2] to the 3′ end at nucleotide 6,161, followed by a 57-nucleotide tract of A residues. The insertional event resulted in duplication of at least 12 nucleotides of the target site in the factor VIII gene from nucleotides 3,054 to 3,065 of the factor VIII cDNA sequence. Interestingly, this 12-nucleotide target site is rich in A residues on the coding strand (8 of 12 residues) (Fig. 2).

The insertion in JH-28 also included the 3′ end of an L*1* sequence (nucleotides 4,020 to 6,161), but it is more complex (Fig. 2). This insertion contained two blocks of sequence, from nucleotides, 4,020–5,114 and 5,115 to 6,161 in a head-to-head arrangement. Like the insertion in JH-27, the block of sequence 5,115–6,161 contained a highly A-rich 3′ sequence of 77 nucleotides, the last 30 of which were A residues. Although this L*1* insertion is rearranged, a clean break between nucleotides 5,114 and 5,115, a polarity change and gap repair occurred without loss of L*1* or factor VIII sequences. The L*1* insertion also created a target site duplication of at least 13 nucleotides of the factor VIII sequence from nucleotides 3,653–3,665 of the factor VIII cDNA sequence. Again, the target site sequence is A-rich in the coding strand (9 of 13 nucleotides).

A consensus sequence for the 3′-trailer region of L*1* cDNA has been published[6]. This sequence, obtained after analysis of 19 L*1* cDNA isolated separately, differs in 15 positions out of 186 nucleotides from the consensus sequence derived from human genomic L*1* elements[2]. We have compared the 186 3′ nucleotides of our two inserted L*1* sequences to both the cDNA and genomic consensus sequences. The inserted sequences differ from each other by only a single nucleotide in this region, and they have the cDNA consensus nucleotide changes at 14 of 15 sites (Fig. 3). Furthermore, 4 of the 19 cDNAs (subset Ta) differ from the cDNA consensus at four addtional nucleotides (Fig. 3). The insert of JH-27 had these four additional changes, and the insert of JH-28 had three of the four changes. Both insertions contained a G residue, 11 nucleotides upstream of the A-rich region. This G has been observed in one of the published cDNA sequences which has the same sequence in the 3′-trailer region as the insert of JH-27. These data provide evidence that our L*1* insertions are the result of invasion of an RNA intermediate or of cDNA into the coding sequences.

These L*1* insertions are the first large non-viral insertions described in man which are not due to expansion of short repeats by unequal crossing-over events. A small segment of an L*1* sequence (36–41 nucletoides) has been found between the breakpoints of a deletion involving the β-globin gene cluster[14]. Insertions of species-specific elements related to human L*1* elements have been observed in other organisms. An analogous

* Present address: Hematology-Oncology Unit, Cox 6, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA.

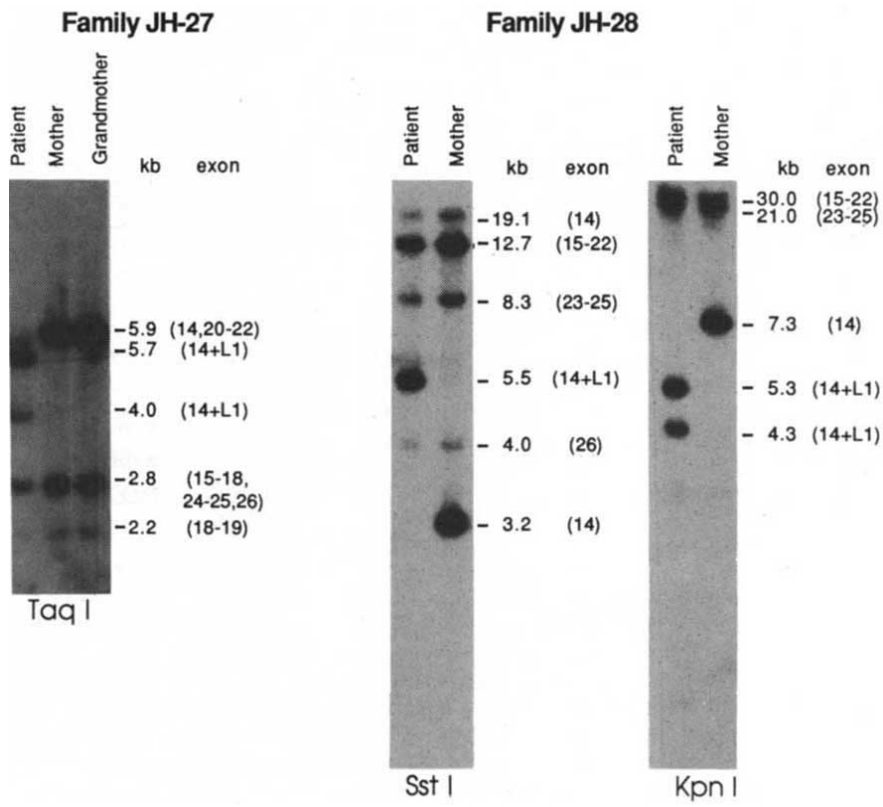

Fig. 1 Restriction endonuclease analysis of DNA from pertinent members of families JH-27 and JH-28. The probe is a 3′ factor VIII cDNA fragment spanning exons 14–26. In the left panel, *Taq*I digest of DNA from JH-27 shows loss of the major 5.9-kb fragment, which includes exon 14, and its replacement with fragments of 5.7 kb and 4.0 kb. The minor fragment (~5.9 kb) which remains in patients DNA is a normal fragment derived from exons 20 to 22. Further analysis showed insertion of nucleotides 2,363–6,161 of an L*1* element, plus an A-rich tract, into exon 14 (Fig. 2). Two new *Taq*I fragments are expected from patient DNA as the L*1* consensus sequence in this region contains a single *Taq*I site at nucleotide 3,415. Note that DNA of mother and maternal grandmother lack the abnormal 5.7-kb and 4.0-kb fragments. In the right-hand panels are *Sst*I and *Kpn*I digests of JH-28 and his mother, probed with the 3′ factor VIII cDNA. In JH-28 the major 3.2-kb *Sst*I fragment of exon 14 is replaced by an abnormal 5.5-kb fragment. The 5.5-kb fragment is not seen in maternal DNA. Also in the patient DNA, the normal 7.3-kb *Kpn*I fragment containing exon 14 is replaced by 5.3-kb and 4.3-kb fragments. These abnormal fragments are not present in maternal DNA. Further analysis showed that JH-28 had insertion into exon 14 of nucleotides 4,020-6,161 of an *L1* element, plus an A-rich tract. This region of the L*1* consensus sequence lacks an *Sst*I site, and the nucleotide sequence of the *L1* element from JH-28 contains a single *Kpn*I site at nucleotide 4,786, thereby explaining the fragment patterns observed.

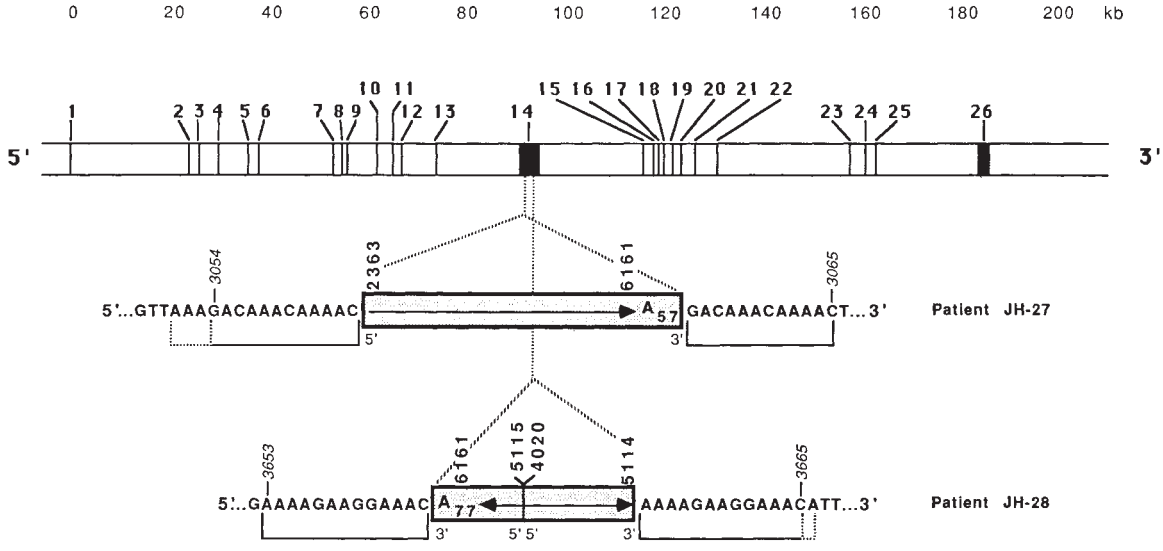

Fig. 2 Diagram of L*1* insertions in exon 14 of the factor VIII gene. The factor VIII gene spans 186 kb and contains 26 exons . In patients JH-27 and JH-28, restriction analysis of leukocyte DNA with a cDNA probe spanning exons 14–26 gave hybridizing fragments of increased size. After cloning the abnormal *Eco*RI fragments derived from exon 14 in λgt10 (ref. 12) and demonstrating that these *Eco*RI fragments contained 3′ L*1* sequences by hybridization to an L*1* probe[20], the cloned fragments were subcloned into M13 and partially sequenced[13]. The 3.8-kb L*1* insertion from patient JH-27 is flanked by a 12-base pair target site duplication of factor VIII cDNA sequence (nucleotides 3,054–3,065, where nucleotide 1 is A of the initiator codon)[8]. Residues 3,051–3,053 of the factor VIII cDNA are adenylic acids and could also be duplicated (shown by the hatched bracket). The rearranged 2.3-kb insertion from patient JH-2 is shown and is flanked by a 13-base pair target site duplication. Residue 3,666 of the factor VIII cDNA is an A and could also be duplicated. Filled boxes represent the L*1* elements, and the arrows within the boxes point towards the 3′ end of the L*1* sequence. The L*1* insertion of patient JH-27 was sequenced between nucleotides 2,363–2,551, 4,964–5,178, 5,977–6,161 and the poly(A) tract, and its restriction map (using *Eco*RI, *Bam*HI, *Sst*I and *Kpn*I) was identical to the restriction map of the L*1* genomic consensus sequence[2]. The L*1* insertion of patient JH-28 was sequenced between nucleotides 4,020–4,237, 4,542–5,269, 5,613–6,161 and the poly(A) tract. Both L*1* insertions showed 98% similarity to the consensus genomic L*1* sequence outside of the 3′-trailer region.

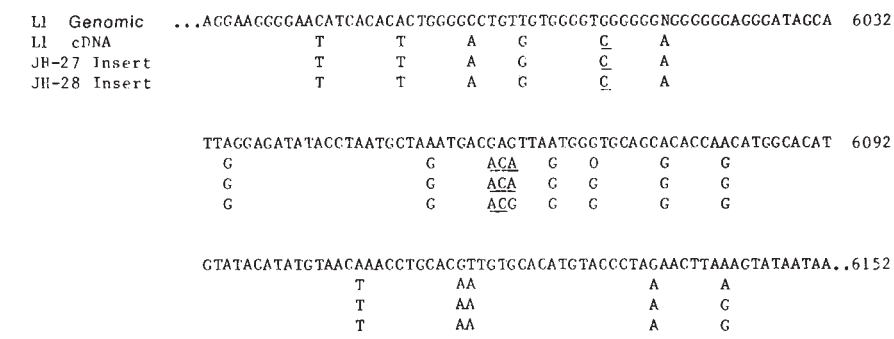

Fig. 3 Comparison of 3′ trailer region sequence (186 nucleotides) of the L*1* element from genomic consensus[2] (top line), cDNA consensus[6] (middle line) and factor VIII L*1* insertion sequences of JH-27 and JH-28. The cDNA consensus shown is that of subset Ta (ref. 6), and the additional four nucleotides which differ from the L*1* cDNA and genomic consensus sequences are underlined.

spontaneous insertion of an F element in *Drosophila*, which has a similar organization and sequence to the L*1* elements of mammals, produces a reversion of the white-ivory mutation at the white locus[15]. F-element insertion sites contain 8–13 nucleotide duplications which are not A-rich, but are random in sequence[16]. The presence of a canine L*1* homologue 5′ to a *myc* gene has been implicated in the aetiology of canine venereal tumour[17].

The consensus sequences of mammalian L*1* elements and *Drosophila* F elements contain two open-reading frames (*orf*-1 and *orf*-2) (refs 2–4, 18, 19). *Orf*-2 has blocks of sequence similarity to the polymerase domain of various reverse transcriptases[3,4]. Because full-length L*1* RNAs have been detected in the poly(A)$^+$ cytoplasmic RNA of human teratocarcinoma cells, it has been postulated that in germ-line cells a small number of full-length L*1* sequences are transcribed and translated[1]. The polypeptide of *orf*-2 has been implicated as a reverse transcriptase of the L*1* RNA itself in producing cDNA copies, many of which are truncated. These cDNAs could then re-enter genomic DNA, perhaps at A-rich sequences, as implied by the sequences of the target sites in the factor VIII gene. We postulate that the poly(T) tail of L*1* cDNA is involved in base-pairing with the A-rich factor VIII sequences after staggered single-strand breaks in the factor VIII gene. Rejoining of the L*1* cDNA to the factor VIII sequence and filling-in of the complementary strand of the cDNA would complete the event.

Insertion of L*1* elements, involving retrotransposition of DNA sequences through an RNA intermediate into a new and distant location in the genome, represents a mechanism for mutation to produce human disease fundamentally different from those previously described. Because we do not know when these L*1* insertion events occur, whether in the sperm or ovum, after fertilization, or during early stages of embryogenesis, the proportion of such insertions that are heritable is unknown. Yet finding two L*1* insertions among 240 patients with haemophilia A suggests that this mechanism of mutation is not uncommon.

We thank Drs T. Fanning, M. Singer and M. Edgell for helpful discussions, E. Pasterfield for assistance in preparation of the manuscript and J. Strayer for the artwork. H.H.K., A.F.S. and S.E.A. are supported by grants from the NIH.

Received 11 November 1987; accepted 22 January 1988.

1. Fanning, T. & Singer, M. F. *Biochem. biophys. Acta* (in the press).
2. Scott, A. F. *et al. Genomics* **1**, 113–125 (1987).
3. Hattori, M. *et al. Nature* **321**, 625–628 (1986).
4. Sakaki, Y. *et al. Cold Spring Harb. Symp. quant. Biol.* **51**, 465–469 (1986).
5. Skowronski, J. & Singer, M. F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6050–6054 (1985).
6. Skowronski, J. & Singer, M. F. *Cold Spring Harb. Symp. quant. Biol.* **51**, 457–464 (1986).
7. Gitschier, J. *et al. Nature* **312**, 326–330 (1984).
8. Toole, J. J. *et al. Nature* **312**, 342–347 (1984).
9. Gitschier, J. *et al. Nature* **315**, 427–430 (1985).
10. Youssoufian, H. *et al. Nature* **324**, 380–382 (1986).
11. Antonarakis, S. E., Youssoufian, H. & Kazazian, H. H. Jr *Molec. biol. Med.* **4**, 81–94 (1987).
12. Maniatis, T., Fritsch, E. G. & Sambrook, J. *Molecular Cloning* (Cold Spring Harbor, 1982).
13. Sanger, K., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
14. Mager, D. L., Henthorn, P. S. & Smithies, O. *Nucleic Acids Res.* **13**, 6559–6575 (1985).
15. DiNocera, P. P. & Casari, G. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5843–5847 (1987).
16. DiNocera, P. P., Digan, M. E. & Dawid, I. B. *J. molec. Biol.* **168**, 715–727 (1983).
17. Katzir, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 1054–1058 (1985).
18. Loeb, D. D. *et al. Molec. cell. Biol.* **6**, 168–182 (1986).
19. Soares, M. B., Schon, E. & Efstratiadis, A. *J. molec. Evol.* **22**, 117–133 (1985).
20. Adams, J. W. *et al. Nucleic Acids Res.* **8**, 6113–6128 (1980).

Oncogene *jun* encodes a sequence-specific *trans*-activator similar to AP-1

Peter Angel, Elizabeth A. Allegretto, Steve T. Okino, Kazue Hattori, William J. Boyle*, Tony Hunter* & Michael Karin

Department of Pharmacology M-036, Center for Molecular Genetics, School of Medicine, University of California, San Diego, La Jolla, California 92093, USA
* The Salk Institute for Biological Sciences, San Diego, California 02138, USA

Proto-oncogenes encode proteins with three main sites of action: the cell-surface membrane, the cytoplasm and the nucleus[1,2]. Although the exact biochemical function of most proto-oncogene products is not understood, several of them are known to be involved in signal transduction[3–7]. A role in gene regulation through DNA binding has been suggested for a recently isolated member of the group of oncogenes acting at the nucleus, v-*jun*. The C-terminus of the putative v-*jun*-encoded protein is similar in sequence to the C-terminus of the yeast transcriptional activator GCN4 (refs 8, 9), which forms its minimal DNA-binding domain[10]. GCN4 binds to specific sites whose consensus sequence[11] is highly similar to the recognition sequence of the mammalian transcriptional activator AP-1 (refs 12, 13). Like GCN4, AP-1 binds to promoter elements of specific genes and activates their transcription[12–15]. Because of the similarity between the recognition sites for GCN4 and AP-1, we examined the possibility that AP-1 could be the product of the c-*jun* proto-oncogene. The experimental results reported here indicate that the JUN oncoprotein is a sequence-specific transcriptional activator similar to AP-1.

Recently AP-1 has been purified from HeLa cells as a protein of relative molecular mass (M_r) 44,000–45,000 (44–45K)[12,13]. Mapping of tryptic peptides indicates that two bands of M_r 44–45K and 40K present in highly purified preparations of AP-1 (see Fig. 1, lane 4) are structurally related (F. Mercurio and W.J.B., unpublished results). To examine the relationship between AP-1 and v-JUN we tested whether antisera raised against two specific peptides derived from the v-JUN sequence[16] can interact with AP-1. Different AP-1 preparations were transferred to nitrocellulose membranes and incubated with the anti-peptide antisera (kindly provided by T. Bos and P. Vogt). As shown in Fig. 1, antiserum directed against a synthetic peptide (USC 4) derived from the putative DNA-binding domain of v-JUN (amino acids 199–215; see Fig. 2) reacted with polypeptides having the same mobility as AP-1 (see lanes 2–5). AP-1 reacted in the same way with another antiserum raised against an unrelated peptide (USC 2) whose sequence is located within the N-terminal portion of v-JUN (amino acids 75–87; see lane