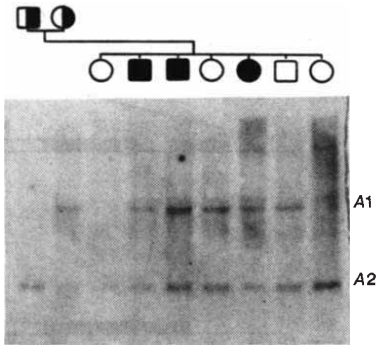

Fig. 2 Segregation of the *Msp*I polymorphism defined by pJ3.11 with the CF locus in a family with three affected (solid symbols) and three unaffected (open symbols) children. The RFLP alleles are 4.2 kilobases (kb) (*A1*) and 1.8 kb (*A2*) long; the frequency of the rare allele *A1* is 0.2 (ref. 26).

Methods. DNA (5 μg) from 10 informative families with two or three affected children was digested with *Msp*I and fractionated by electrphoresis on 0.8% agarose gels by standard methods[9]. Fractionated DNA was transferred to Hybond membranes (Amersham International) and hybridized[9]. The probe used was the 500-bp *Hin*dIII–*Eco*RI fragment from the insert of pJ3.11, labelled to a specific activity of 1×10^9 d.p.m. μg^{-1} by synthesis using random oligonucleotide primers. Blots were washed to a final salt concentration of 0.15 M NaCl in the presence of 0.2% SDS. Autoradiography was for 16 h at −70° using double-intensification screens.

RFLP detected by this sequence do not co-segregate with cystic fibrosis in families with multiple affected sibs[13]. Several groups have reported extensive linkage exclusions between the CF locus and protein and DNA markers[14–16], which in total exclude ~40% of the human genome.

We have tested for linkage between the CF locus and multiple markers on a single human chromosome by multipoint linkage analysis[17]. As part of our attempt to study human chromosomes for which few exclusions had been obtained for cystic fibrosis, we have examined the linkage relationships of the two DNA markers located on chromosome 7 to the CF locus. These are the anonymous DNA segment pJ3.11, which defines an *Msp*I polymorphism and has been mapped to 7cen-q22 by somatic cell hybrid panels[18], and the T-cell receptor β-chain (TCRβ) gene which defines *Bgl*II polymorphism and maps to 7q3 (ref. 19). These localizations are summarized in Fig. 1.

The linkage relationships between the two markers and the CF locus are shown in Table 1. The maximal LOD score between the locus defined by pJ3.11 and the CF locus calculated for combined sexes is 5.24 at a recombination fraction $(\theta) = 0$ (99% confidence interval 0–13 centimorgans). A LOD score of 3.0 is accepted as a sufficient statistic to prove synteny between two loci[20]. No recombinants between the probe and the CF locus have been observed in 27 meioses. A typical pedigree is shown in Fig. 2. All informative pedigrees and typings are available from the authors.

TCRβ (7q3) shows significant linkage (for two syntenic loci[21]) to the CF locus at a genetic distance of 10 centimorgans. We have also demonstrated that the collagen locus *COL1A2* is linked to the CF locus at a genetic distance of ~10 centimorgans[22], suggesting that *COL1A2* (7q22) and *TCRβ* (7q3) may flank the CF locus, as they are not themselves closely linked. We are presently analysing the linkage between pJ3.11, *TCRβ* and *COL1A2* in large reference pedigrees (Centre d'Etude Polymorphisme Humain), as well as in cystic fibrosis families which have been typed for *PON*.

Genetic heterogeneity of the locus mutated in cystic fibrosis (as opposed to multiple mutations causing a defect at a single locus) would be demonstrated by the observation of heterogeneity in the linkage relationship in different families between a given close marker and the CF locus. The informative families in this study originated from Australia, Eire, France, West Germany, United Kingdom and Pakistan. All the families show tight linkage between the pJ3.11 probe and the CF locus. It would be of interest to study these probes in the Amish kindred reported by Klinger[23], as there are data indicating differences in linkage for that family compared with those studied here (K. Klinger, personal communication). A more extensive analysis covering different populations is in progress.

The pJ3.11 probe is presently informative in approximately 40% of matings and, given the tight linkage demonstrated here, would provide accurate heterozygote detection and antenatal diagnosis of cystic fibrosis in some cases. However, such diagnoses will be facilitated markedly when the availability of a set of informative and ordered flanking markers (*PON*, pJ3.11, *COL1A2*, and *TCRβ*) has been confirmed.

We thank the Cystic Fibrosis Research Trust, and especially Ron Tucker, the MRC and the Foulkes Foundation for consistent and generous financial support, and our many clinical colleagues and the CF families for their help in obtaining and giving samples. We particularly thank Kay Davies, Lap-Chee Tsui, Kathy Klinger, Veronica van Heyningen and Nigel Spurr for helpful discussions, and for providing data in advance of publication.

Received 4 November; accepted 7 November 1985.

1. Williamson, R. *et al.* in *Cystic Fibrosis—Horizons* (ed. Lawson, D.) 139–153 (Wiley, Chichester, 1984).
2. Quinton, P. M. & Bijman, J. *New Engl. J. Med.* **308,** 1185 (1983).
3. Romeo, G. *et al. Am. J. hum. Genet.* **37,** 334–338 (1985).
4. Eiberg, H., Mohr, J., Schmiegelow, K., Nielsen, L. S. & Williamson, R. *Clin. Genet.* **28,** 265–271 (1985).
5. Davies, K. E. *et al. Nucleic. Acids Res.* **11,** 2303–2312 (1983).
6. Gusella, J. F. *et al. Nature* **306,** 234–238 (1983).
7. Reeders, S. T. *et al. Nature* **317,** 542–544 (1985).
8. Mayo, B. J., Klebe, R. J., Barnett, D. R., Lankford, B. J. & Bowman, B. H. *Clin. Genet.* **18,** 329–386 (1980).
9. Scambler, P., Gilliam, C., Tippet, P., Boylston, A. & Williamson, R. *Hum. Genet.* **82,** 6167–6171 (1985).
10. Edwards, J. H., Jonasson, J. A. & Blackwell, N. L. *Lancet* **i,** 1020 (1984).
11. Scambler, P. J. & Williamson, R. *Cytogenet. Cell Genet.* **39,** 231–233 (1985).
12. Scambler, P. J., Wainwright, B. J., MacGillivray, R., Fung, M. R. & Williamson, R. *Am. J. hum. Genet.* (in the press).
13. Davies, K. E., Gilliam, T. C. & Williamson, R. *Molec. Biol. Med.* **1,** 185–190 (1983).
14. Scambler, P. J. *et al. Cytogenet. Cell Genet.* (in the press).
15. Tsui, L.-C., Kennedy, D., Zsiga, M. & Buchwald, M. *Am. J. hum. Genet.* **36,** 156 (1984).
16. Bowcock, A. M., Lee, G. M., Sanders, M. F. & Cavalli-Sforza, L. *8th int. Workshop Human Gene Mapping,* Helsinki, abstr. 37 (Karger, New York, in the press).
17. Farrall, M., Scambler, P., North, P. & Williamson, R. *Am. J. hum. Genet.* (in the press).
18. Bartels, I., Grzeschik, K.-H., Cooper, D. & Schmidtke, J. *Am. J. hum. Genet.* (in the press).
19. Morton, C. C. *et al. 8th int. Workshop Human Gene Mapping,* Helsinki, abstr. 256 (Karger, New York, in the press).
20. Morton, N. *Am. J. hum. Genet.* **7,** 277–318 (1955).
21. Chotai, J. *Ann. hum. Genet.* **48,** 359–378 (1984).
22. Scambler, P. J. *et al. Lancet* (submitted).
23. Klinger, K. *Hum. Genet.* **65,** 94–98 (1983).
24. Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. *Proc. natn. Acad. Sci. U.S.A.* **81,** 3443–3446 (1984).
25. Tsipouras, P. *et al. Am. J. hum. Genet.* **36,** 1172–1179 (1984).
26. Cooper, D. N., Smith, B. A., Cooke, H. J., Niemann, S. & Schmidtke, J. *Hum. Genet.* **69,** 201–205 (1985).

The human *met* oncogene is related to the tyrosine kinase oncogenes

Michael Dean*, Morag Park†, Michelle M. Le Beau‡, Terry S. Robins*, Manuel O. Diaz‡, Janet D. Rowley‡, Donald G. Blair† & George F. Vande Woude*

* LBI-Basic Research Program, NCI-Frederick Cancer Research Facility and † Laboratory of Molecular Oncology, National Cancer Institute, PO Box B, Frederick, Maryland 21701, USA
‡ Joint Section of Hematology/Oncology, University of Chicago, Chicago, Illinois 60637, USA

The *met* oncogene was previously isolated from a chemically transformed human cell line, MNNG-HOS (refs 1, 2). Recent evidence has demonstrated that two classes of transcripts are expressed from the *met* proto-oncogene locus. The *met* oncogene, however, expresses an aberrant RNA which has sequences in common with both transcripts. We now report partial nucleotide sequencing of the human *met* oncogene and show that *met* is

related to the protein kinase oncogenes and growth factor receptors. The *met* nucleotide sequence is not identical to that of any published gene, and it is more closely homologous to the tyrosine kinases than to the serine/threonine kinases. Within the tyrosine kinase family, the sequenced *met* domains are most closely related to the human insulin receptor and the viral *abl* gene. *In situ* chromosome hybridization has mapped *met* to human chromosome 7 band 7q21–q31, a location distinct from that of other kinases. This is also a region associated with nonrandom chromosomal deletions observed in a portion of patients with acute nonlymphocytic leukaemia. The accompanying paper[3] shows that this chromosomal locus is also tightly linked with the human heredity disease cystic fibrosis.

To characterize the *met* oncogene and its protein product(s), we have sought to determine the nucleotide sequence of coding portions of the locus. The sequences required for transformation by *met* are contained within 35 kilobases (kb) of human genomic DNA and NIH 3T3 cells transformed with *met* express a 6.5-kb messenger RNA[1,4]. Initial attempts to isolate a complementary DNA copy of this RNA generated a 2.0-kb clone. However, the partial nucleotide sequence indicated that this clone contains only untranslated sequences (M.P., data not shown). As an alternative approach, a 9-kb fragment was cloned into a retrovirus vector derived from Moloney sarcoma virus (Fig. 1).

Retroviral vectors have been shown to process segments of genomic DNA, generating cDNA-like copies[5]. We hoped that by cloning a segment of *met* genomic DNA into such a vector we could recover a clone devoid of intervening sequences. An S_1 nuclease experiment revealed that passage of this segment of *met* genomic DNA through a retroviral intermediate resulted in the excision of at least one intervening sequence; the recovered clone protected a fragment 1–200 bases longer than the corresponding segment of genomic DNA (data not shown). Nucleotide sequencing of the recovered clone and the corresponding portions of genomic DNA revealed an open reading frame 126 amino acids long (Fig. 2*A*). When this putative *met* amino-acid sequence was used to search an amino-acid sequence database, homology with several tyrosine kinase oncogenes and growth factor receptors was revealed (Fig. 2*B*).

In addition, we have also determined the nucleotide sequence of a 1.1-kb fragment of genomic DNA from the middle of the *met* oncogene (fragment D) which is known to hybridize with *met* mRNA transcripts[2]. This sequence contained an open reading frame bounded by splice acceptor and donor signals which shows dramatic homology to the tyrosine kinases (Fig. 2). Figure 2*B* compares the amino-acid sequence of the two open reading frames with several members of the tyrosine kinase family. The putative exon is identical in 18/23 residues with the human insulin receptor[6,7] and in 15/23 amino acids with the virus *abl*

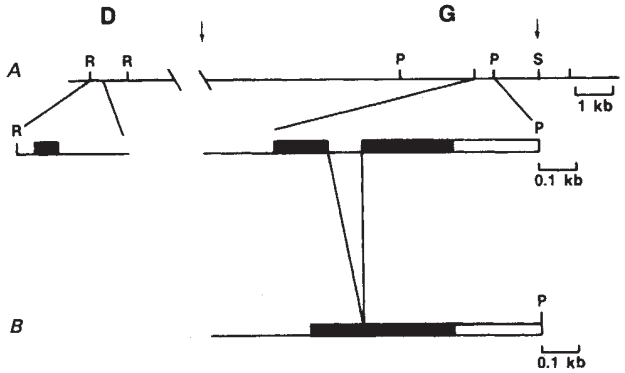

Fig. 1 Sequencing strategy of the *met* oncogene. *A*, Genomic DNA located at the 3′ end. Filled areas depict coding regions or putative exons. D and G refer to fragments described previously[1,2]. Both strands of the expanded regions were sequenced by the chain-termination method[26]. The 9-kb *Sal*I fragment inserted into the retrovirus vector is depicted by the arrows. *B*, *met* sequences contained within the spliced clone recovered. R, *Eco*RI; P, *Pst*I; S, *Sal*I.

Methods. The 9-kb *Sal*I fragment was cloned into pGV16 (ref. 27), which was modified to accept *Sal*I fragments by inserting a *Sal*I linker. Virus was rescued from this plasmid by co-transfection with an equal amount of pMOV-3 plasmid[28] into NIH 3T3 cells. Infected NIH 3T3 cells were subsequently selected for resistance to G418 using 500 μg ml^{-1} Geneticin (Gibco). The viral DNA from the transfected cells was recloned by fusing G418-resistant NIH 3T3 cells with COS cells using polyethylene glycol as described previously[29]. Hirt supernatants[30] were prepared after 72 h and used to transform *Escherichia coli* strain HB101 (BRL) using the procedure described by Hanahan[31] and selected on kanamycin (35 μg ml^{-1}). Only colonies which hybridized to the G fragment probe (see text) were analysed.

gene[8]. This region is highly conserved among all members of the tyrosine kinase family[9], and moderately conserved between other kinases such as *mos* (ref. 10) (Fig. 2*B*), *raf* (ref. 11), the yeast CDC28 protein[12] and cyclic AMP-dependent kinase[13]. Although the two exons at the C-terminus of *met* contain a region which is less well conserved among kinases[9], certain portions of this domain are conserved among all members of the family,including *met* (Fig. 2*B*).

We have previously demonstrated that although *met* maps to chromosome 7, it is not the *erb-B*/epidermal growth factor (EGF) receptor gene, which is located at chromosome 7 p12–p14 (refs 1, 14). To map the location of *met* on chromosome 7 more precisely, we performed *in situ* chromosomal hybridization. A ^{3}H-radiolabelled *met* probe was hybridized to metaphase cells

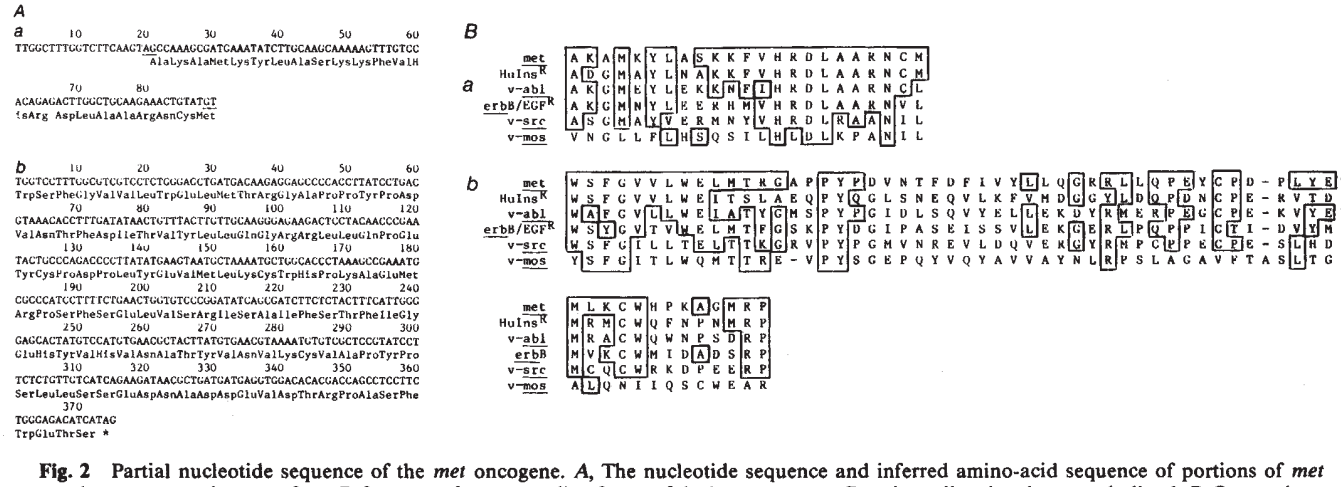

Fig. 2 Partial nucleotide sequence of the *met* oncogene. *A*, The nucleotide sequence and inferred amino-acid sequence of portions of *met* are shown: *a*, putative exon from D fragment; *b*, open reading frame of the last two exons. Putative splice signals are underlined. *B*, Comparison of the *met* amino-acid sequence with corresponding regions of the human insulin receptor (Hu Ins^R)[6,7], viral *abl* oncogene[8], human EGF receptor[32], and viral *src* and human *mos* proto-oncogenes[10,33].

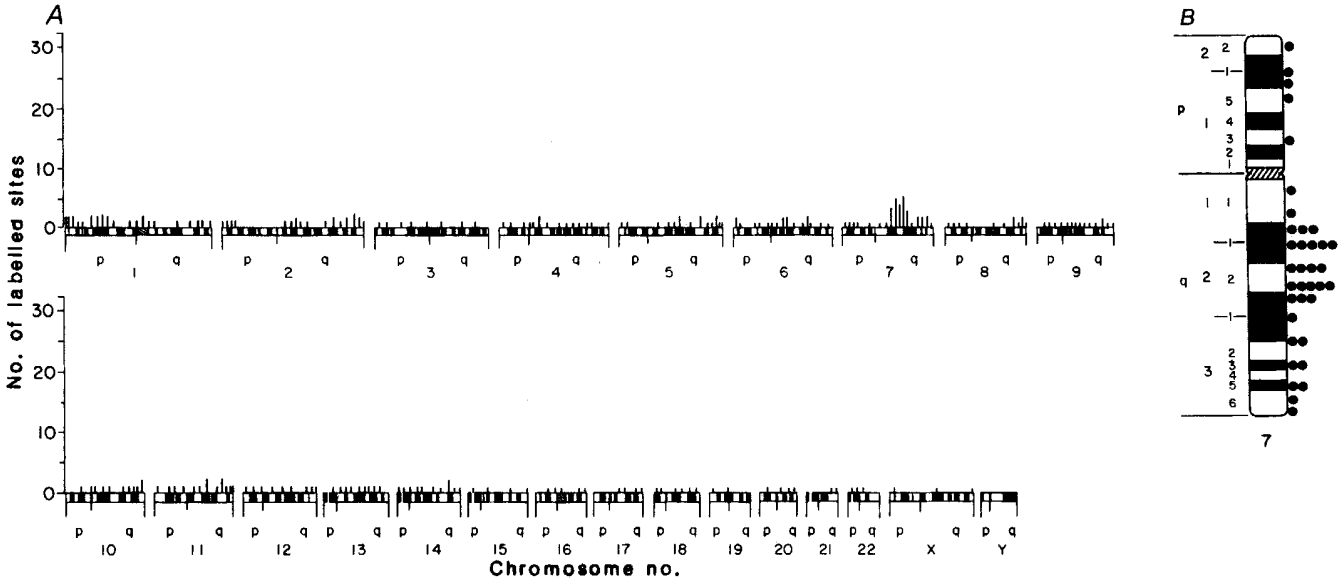

Fig. 3 *In situ* chromosomal hybridization of the *met* oncogene. *A*, Distribution of labelled sites in 100 normal metaphase cells. Human metaphase cells prepared from phytohaemagglutinin-stimulated peripheral blood lymphocytes were hybridized with the ^{3}H-labelled E probe[1]. Specific labelling was observed on chromosome 7 at bands q21–q31. *B*, Distribution of labelled sites on chromosome 7. Of 100 metaphase cells examined, 19 (19%) were labelled on bands q21–q31 of chromosome 7. Of a total of 36 grains on chromosome 7, 21 or 58% were located on q21–q31; these sites represented 8.2% (21/256) of all labelled sites ($P > 0.005$). A modification of the method of Harper and Saunders was used[34,35].

from phytohaemagglutinin-stimulated peripheral blood lymphocytes. Figure 3*A* shows the distribution of grains observed in 100 metaphase cells. A significant clustering of grains was observed on the long arm of chromosome 7 at bands q21–q31 ($P < 0.005$); no other chromosome showed specific labelling. Of the 100 metaphase cells examined, 19 were labelled on bands q21–q31 of one or both of the chromosome 7 homologues (Fig. 3*B*); these sites represented 8.2% (21/256) of all labelled sites. The largest cluster of grains was observed at band 7q22. Specific labelling was also observed at this chromosomal region when *met* was hybridized to metaphase cells from peripheral blood lymphocytes obtained from a second individual. Of 50 metaphase cells examined, 10 showed label on one or both chromosome 7 homologues, representing 7.1% (12/169) of all labelled sites. These results confirm the location of this gene on the long arm of human chromosome 7 and refine the location to band q21–q31.

We conclude that the *met* oncogene is a member of the tyrosine kinase family, although analysis of the *met* protein product will be required to establish whether it possesses this activity. Analysis of *met*-related mRNAs has revealed that the transforming gene expresses a hybrid truncated mRNA[4]. Thus, we speculate that the *met* oncogene transforms cells by expressing a truncated form of the kinase encoded by the proto-oncogene. A similar mechanism appears to be responsible for the activation of both the v-*erb-B* and c-*erb-B*/EGF receptor genes[15,16], and the *bcr*/*abl* gene in chronic myelogenous leukaemia[17].

The homology between *met* and the kinase domain of the insulin receptor is intriguing. Insulin is a member of a gene family which includes insulin-like growth factors I and II and nerve growth factor[18]. Similarly, there are several distinct receptors for insulin-related mitogens[19]. It has not escaped our attention that the *met* class I gene product may encode a receptor for an insulin-like peptide. However, *met* is also closely related to *abl*. Identification and characterization of the *met* protein(s) should help address these interesting questions.

Nonrandom deletions of the long arm of chromosome 7 are frequently observed in the leukaemic cells of patients with acute non-lymphocytic leukaemia (ANLL). Interestingly, *met* was isolated from a human cell line (HOS) transformed with the chemical carcinogen *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine, and increasing evidence now relates the development of ANLL to mutagen exposure both in patients subjected to cytotoxic therapy for a primary malignancy and in patients exposed to environmental carcinogens[20,21].

It is notable that recurring abnormalities involving chromosome 7 are common in such patients. These abnormalities include loss of a whole chromosome 7 or of part of the long arm of this chromosome, del(7q) (refs 20, 22). It may be significant that *met* is located in the region of the proximal breakpoint of the interstitial deletions of 7q observed in patients with ANLL[23].

In the accompanying paper, White *et al.*[3] describe restriction fragment length polymorphisms (RFLPs) from the *met* locus which were developed for examining ANLL patients. They find that these RFLPs are tightly linked to the hereditary disease cystic fibrosis (CF). The *in situ* chromosomal mapping of *met* thus localizes the CF locus in man. At the very least then, *met* is a marker for CF, a disorder that is speculated to involve an ion-channel imbalance[24]. However, we show here by sequence homology that *met* is a member of the tyrosine kinase family and we note that tyrosine kinase receptor activation has been linked to the regulation of the sodium ion-proton antiport[25]. We also note that the *met* proto-oncogene locus expresses two transcripts which appear to be noncoordinately expressed overlapping transcripts[3]. The genomic sequences encoding these two transcripts could cover greater than several hundred kilobases. Present genetic data[3] suggest that the distance between *met* and the CF locus is less than 5,000 kb.

We thank Doug Ludtke and Martin Schmidt for assistance with DNA sequencing, Chris Westphal and Linda Miller for technical assistance, and Friedrich Propst for help with the computer analysis. Research was sponsored by the National Cancer Institute, DHHS, under contract NO1-CO-23909 with Litton Bionetics, Inc. Additional support was provided by contract DE-ACO2-80EV 10360 from the US Department of Energy, PHS grant CA 23954 from the NCI and grant 84-20 from the Illinois Division of the American Cancer Society. M.D. is a Fellow of the ACS and M.M.Le B. is a Special Fellow of the Leukemia Society of America. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US government.

Received 31 October; accepted 6 November 1985.

1. Cooper, J. A. *et al. Nature* **311**, 29-33 (1984).
2. Cooper, C. S. *et al. Cancer Res.* **44**, 1-10 (1984).
3. White, R. *et al. Nature* **318**, 382-384 (1985).
4. Park, M. *et al.* (in preparation).
5. Robins, T., Jhappan, C., Chirikjian, J. & Vande Woude, G. F. *Gene analysis Techq.* (in the press).
6. Ullrich, A. *et al. Nature* **313**, 756-761 (1985).
7. Ebina, Y. *et al. Cell* **40**, 747-758 (1985).
8. Reddy, E. P., Smith, M. J. & Srinivasan, A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3623-3627 (1983).
9. Hunter, T. & Cooper, J. A. in *The Enzymes* (eds Boyer, P. D. & Krebs, E.) (Academic, New York, in the press).
10. van Beveran, D., van Straaten, F., Galleshaw, J. A. & Verma, I. M. *Cell* **27**, 97-108 (1981).
11. Bonner, T. I. *et al. Molec. cell. Biol.* **5**, 1400-1407 (1985).
12. Lorinez, A. T. & Reed, S. I. *Nature* **307**, 183-185 (1984).
13. Shoji, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 848-851 (1981).
14. Merlino, G. T. *et al. Molec. cell. Biol.* **5**, 1722-1734 (1985).
15. Yamamoto, T. *et al. Cell* **35**, 71-78 (1983).
16. Nilsen, T. W. *et al. Cell* **41**, 719-726 (1985).
17. Shtivelman, E., Lifshitz, B., Gale, R. P. & Canaani, E. *Nature* **315**, 550-553 (1985).
18. Froesch, E. R., Schmid, C., Shwander, J. & Zapf, J. *A. Rev. Physiol.* **47**, 443-467 (1985).
19. Rechler, M. M. & Nissley, S. P. *A. Rev. Physiol.* **47**, 426-442 (1985).
20. Rowley, J. D., Golomb, H. M. & Vardiman, J. W. *Blood* **58**, 759-767 (1981).
21. Mitelman, F. *et al. Cancer Genet. Cytogenet.* **9**, 197-214 (1981).
22. Yunis, J. J. *Cancer Genet. Cytogenet.* **11**, 125-137 (1984).
23. Rowley, J. D. & Testa, J. R. *Adv. Cancer Res.* **36**, 103-149 (1982).
24. van Heyningen, V. *Nature* **311**, 104-105 (1984).
25. l'Allemain, G., Paris, S. & Pouyssegur, J. *J. biol. Chem.* **259**, 5809-5815 (1984).
26. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
27. Jhappan, C., Vande Woude, G. & Robins, T. *J. Virol.* (submitted).
28. Harbers, K., Schnieke, A., Stuhlman, H., Jahner, D. & Jaenishch, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7609-7613 (1981).
29. Pontecorvo, G. *Somat. Cell Genet.* **1**, 397-400 (1975).
30. Hirt, B. *J. molec. Biol.* **26**, 365-369 (1967).
31. Hanahan, D. *J. molec. Biol.* **166**, 557-586 (1983).
32. Downward, J. *et al. Nature* **307**, 521-527 (1984).
33. Czernilofsky, A. P. *et al. Nature* **287**, 198-203 (1980).
34. Harper, M. E. & Saunders, G. F. *Chromosoma* **83**, 431-439 (1981).
35. LeBeau, M. M., Westbrook, C., Diaz, M. O. & Rowley, J. D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6692-6696 (1985).

Retroviral *gag* and DNA endonuclease coding sequences in IgE-binding factor gene

Hiroyuki Toh*, Masao Ono† & Takashi Miyata*‡

* Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan
† Department of Molecular Biology, School of Medicine, Kitasato University, Sagamihara-shi, Kanagawa 228, Japan
‡ To whom correspondence should be addressed.

Immunoglobulin-binding factors are known to regulate the synthesis of B-cell-derived immunoglobulin heavy-chain isotypes[1]. Cloning and nucleotide sequence determination of complementary DNA encoding rodent IgE-binding factors (IgE-BF) revealed that messenger RNA encodes a glycoprotein of 557 amino acids which is expressed as a precursor of relative molecular mass (M_r) 60,000 (60K) in COS7 monkey cells[2]. We report here that the 3′ two-thirds of the IgE-BF coding sequence shows a surprising homology (72%) at the DNA level with coding sequences of the *gag* and *pol* (DNA endonuclease) genes of the Syrian hamster intracisternal A particle (IAP H18)[3], an endogenous retrovirus. This marked homology demonstrates that the rodent gene encoding IgE-BF is a hybrid gene which evolved very recently by integrating genes of viral origin, and that the encoded polypeptide comprises three separate domains: an IgE-BF domain and retrovirus-derived *gag* and DNA endonuclease-like domains. This may represent the first report of a cellular gene containing a virus-derived coding sequence.

Figure 1 shows a homology matrix comparison between the nucleotide sequence coding for the rodent IgE-BF and that of the Syrian hamster IAP genome[3]. A striking homology with IAP is found for the 3′ two-thirds (nucleotides 700-1,770) of the IgE-BF coding sequence, whereas the remaining region (1-699) shares no obvious homology with IAP. In IAP, the homology extends over two regions (nucleotides 1,483-2,140 and 5,364-5,772) which are interrupted by a non-homologous sequence of ~3,200 nucleotides bounded by direct repeats of non-identical but homologous sequences (2,142-2,185 and 5,364-5,406). This non-homologous sequence is completely absent in the IgE-BF coding sequence (Fig. 1*b*). The direct repeats may result in the deletion of a large region of DNA in the IgE-BF coding sequence. The 5′ homologous region of IAP exists within the *gag* coding region, while the 3′ block contains part of the DNA endonuclease coding sequence[3,4]. Alignment (Fig. 2) of the nucleotide sequences reveals 72% homology at the DNA level. Martens *et al.*[2] noted that the deduced amino-acid sequence of the IgE-BF precursor polypeptide shows local homology with retroviral *pol* gene products in the C-terminal region of 50 amino acids. This high degree of sequence homology indicates unequivocally that the two sequences are evolutionarily related.

These results strongly suggest that the rodent IgE-BF precursor comprises two components of distinct origin: the N-terminal one-third (amino-acid positions 1-202) carries a function unique to IgE-BF while the remaining C-terminal portion consisting of *gag* (203-422) and DNA endonuclease (423-557) related domains was derived from a retrovirus or endogenous retrovirus very recently during evolution. That the IgE-BF coding sequence shows a high degree of nucleotide sequence homology with the rodent IAP, together with the existence of *pol*-related sequences in various retroviruses and transposons across a wide taxonomic distance[4-9], strongly suggests that a certain type of retrovirus or endogenous retrovirus was inserted into a cellular DNA region proximal to the primordial IgE-BF gene, and these were later

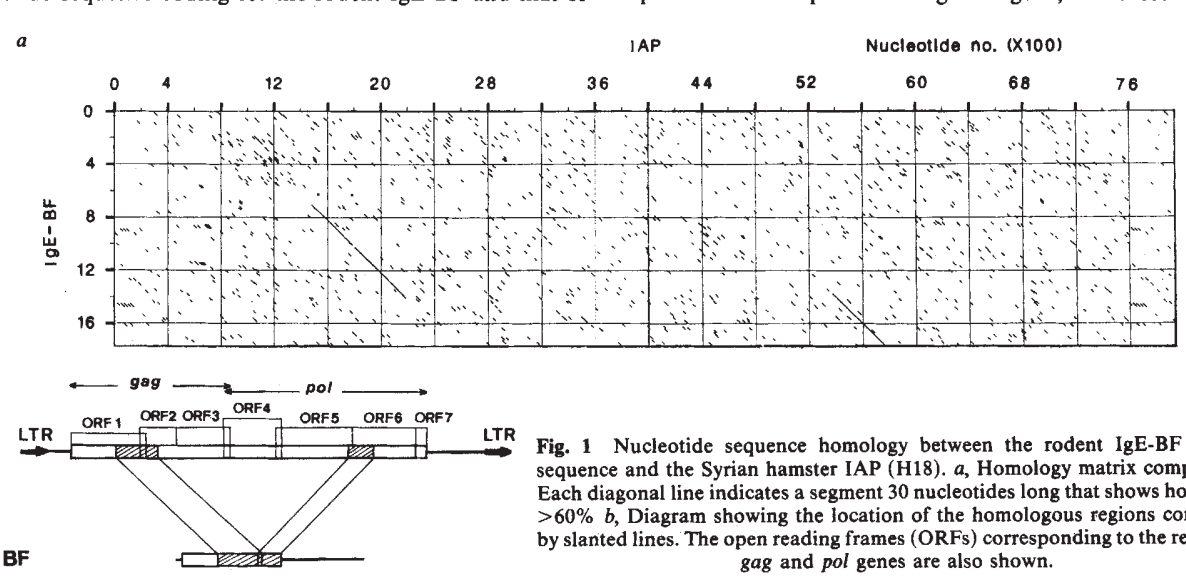

Fig. 1 Nucleotide sequence homology between the rodent IgE-BF coding sequence and the Syrian hamster IAP (H18). *a*, Homology matrix comparison. Each diagonal line indicates a segment 30 nucleotides long that shows homology >60% *b*, Diagram showing the location of the homologous regions connected by slanted lines. The open reading frames (ORFs) corresponding to the retroviral *gag* and *pol* genes are also shown.