the intensity of the 28*S* ribosomal RNA band in ethidium-stained test gels and methylene blue–stained filters after transfer, and in vitro translation of high molecular weight protein from the RNA in two samples.

Because of difficulty visualizing the 14-kb dystrophin message in disease muscle, even under conditions that produced much nonspecific binding, we used the alternative technique of in situ RNA hybridization, which proved to be both more sensitive and more specific. In situ hybridization studies of sections of normal human skeletal muscle showed binding of the probe to the mRNA sequence within muscle fibers. The signal intensity was consistently greatest in and around muscle fiber nuclei (Fig. 2A). This binding pattern is similar in distribution but less intense than we have found with a probe for creatine kinase. With the DMD gene probe there was no apparent variation in the labeling pattern by muscle fiber type, and sections of cardiac muscle showed a pattern similar to that of skeletal muscle although the intensity of the signal was less. There was very little nonspecific binding by this technique: control sections hybridized with a probe of reverse polarity showed signal intensity about 1/20 of that seen in the experimental sections, without an increase in signal around muscle fiber nuclei (Fig. 2B).

In situ RNA hybridization study was done on sections of biopsied muscle from ten DMD patients. In each case signal was clearly detected (Fig. 2C), although it varied more than in normal muscle. The signal was reduced in fibers with relatively normal appearance, and clusters of regenerating muscle fibers [those with basophilia and small size (see Fig. 2E)] showed particularly high signal intensity. In regenerating fibers the signal was again most intense in the vicinity of the nuclei. The control sections (hybridized with a probe of reverse polarity) showed substantially less signal than the experimental sections, and the characteristic labeling over regenerating fibers and around muscle fiber nuclei was not seen (Fig. 2D).

Our finding of a 14-kb DMD gene transcript of relatively low abundance in RNA from normal muscle is consistent with results reported by others (*3, 4*). The expression found in myogenic cells in culture differs from the initial report (*3*) that the gene is not expressed in cultured myoblasts, but this difference is readily explainable by the state of differentiation of the culture preparations in our experiments. It appears that the gene is normally transcribed in muscle cells only after myoblast fusion (*7*).

Diseased muscle showed evidence of transcription of at least the 5′ end of the gene by in situ hybridization, although full-length mRNA was not seen on Northern blot in most cases. This finding could have been due to greater sensitivity of the in situ hybridization technique. It is also likely that the message seen in diseased muscle by in situ hybridization is not seen on Northern blot because it is truncated or rapidly degraded. This interpretation is consistent with recent findings that 50% or more of DMD patients have deletions of part of the gene, frequently in the central portion (*5, 8*). In situ hybridization studies with another cDNA probe from the central part of the gene (*5*) show a loss of signal (nonspecific background binding only) in muscle from two patients with known deletions in this region. Our results indicate that in some patients transcription of the full-length message does not occur. On the basis of these findings we would predict that the complete gene product is markedly reduced or absent in patients with the disease.

Our finding of signal for the 5′ end of the transcript indicates that none of the DMD patients sampled had a true null mutation, however. Regenerating muscle fibers appeared to have greater signal intensity than normal, perhaps indicating a compensatory increase in transcription. If the reading frame is preserved, these patients might be expected to have protein corresponding to at least this portion of the message. The disease phenotype likely depends on the site of the defect within the gene and whether translation of the message is affected.

REFERENCES AND NOTES

1. C. Verellen-Dumoulin *et al.*, *Hum. Genet.* **67**, 115 (1984); U. Francke *et al.*, *Am. J. Hum. Genet.* **37**, 250 (1985); P. N. Goodfellow, K. E. Davies, H. H. Ropers, *Cytogenet. Cell Genet.* **40**, 296 (1985).
2. A. P. Monaco *et al.*, *Nature (London)* **316**, 842 (1985); P. N. Ray *et al.*, *ibid.* **318**, 672 (1985); L. M. Kunkel *et al.*, *ibid.* **322**, 73 (1986).
3. A. P. Monaco *et al.*, *ibid.* **323**, 646 (1986).
4. A. H. M. Burghes *et al.*, *ibid.* **328**, 434 (1987).
5. M. Koenig *et al.*, *Cell* **50**, 509 (1987).
6. E. P. Hoffman, R. H. Brown, L. M. Kunkel, *ibid.* **51**, 919 (1987).
7. A. A. Lev, C. C. Feener, L. M. Kunkel, R. H. Brown, *J. Biol. Chem.* **262**, 15817 (1987).
8. S. M. Forrest *et al.*, *Nature (London)* **329**, 638 (1987); J. T. den Dunnen, E. Bakker, E. G. K. Breteler, P. L. Pearson, G. J. B. van Ommen, *ibid.*, p. 640.
9. T. Maniatis, E. F. Fritsch, J. Sambrook, Eds., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982).
10. J. M. Chirgwin, A. E. Przybyla, R. J. MacDonald, W. J. Rutter, *Biochemistry* **18**, 5294 (1979).
11. D. A. Melton *et al.*, *Nucleic Acids Res.* **12**, 7035 (1984).
12. M. F. Chesselet, L. Weiss, C. Wuenschell, A. J. Tobin, H. U. Affolter, *J. Comp. Neurol.* **262**, 125 (1987).

13. Supported in part by grants from the Muscular Dystrophy Association, the March of Dimes Birth Defects Foundation, and the National Institutes of Health (NS08075 and GM32592).

4 November 1987; accepted 29 January 1988

Human Immunodeficiency Virus May Encode a Novel Protein on the Genomic DNA Plus Strand

ROGER H. MILLER

The genome of the human immunodeficiency virus (HIV) is known to contain eight open reading frames (ORFs) on the minus strand of the double-stranded DNA replicative intermediate. Data presented here indicate that the DNA plus strand of HIV contains a previously unidentified ORF in a region complementary to the envelope gene sequence. This ORF could encode a protein of approximately 190 amino acid residues with a relative molecular mass of 20 kilodaltons if translation began from the first initiation codon. The predicted protein is highly hydrophobic and thus could be membrane associated. It is possible, therefore, that the HIV genome encodes a protein on antisense messenger RNA.

THE HUMAN IMMUNODEFICIENCY virus (HIV) is the primary etiologic agent of the acquired immunodeficiency syndrome (AIDS). Like other retroviruses, HIV replicates by a mechanism involving reverse transcription of RNA and generation of a double-stranded DNA intermediate. Virion RNA of plus strand polarity, transcribed from minus strand DNA, contains the three open reading frames (ORFs) that encode the viral capsid, envelope (*env*), and polymerase proteins, as well as five additional proteins not commonly found in other retroviruses: tat, sor, trs/art, R, and 3′-orf (*1*). Computer analysis of HIV type 1 (HIV-1), presented here, indicates the presence of a ninth ORF with a novel location: the plus strand of the virus DNA genome.

Computer analysis (*2*) of the plus strand DNA sequence of HIV reveals a long ORF located in the region of the genome complementary to the *env* gene sequence of the

Hepatitis Viruses Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD 20892.

minus DNA strand (Fig. 1). The ORF has the capacity to encode a protein of approximately 190 amino acids with a molecular mass of 20 kD, if translation begins at the first initiation codon. Comparison of the predicted amino acid sequences encoded by the DNA plus strand ORF of 12 HIV-1 isolates (*3*) demonstrates a high degree of primary sequence conservation among the proteins (Fig. 2A). The sequence of the amino terminus of the molecule is highly conserved: 83 of the first 100 amino acid residues are identical among the HIV-1 sequences examined. The putative protein, rich in cysteine, leucine, proline, and serine residues, is predicted to be unusually hydrophobic (Fig. 2B). These data are consistent with the hypothesis that the DNA plus strand of the HIV-1 genome encodes a novel protein that may be tightly associated with membrane structures.

Several lines of evidence suggest that the plus strand ORF represents a genuine gene sequence. First, it is feasible that the DNA plus strand of HIV-1 is transcribed because the region 5′ to the ORF contains conserved elements similar to eukaryotic promoter elements (*4*), and the region 3′ to the ORF contains the signal sequences necessary for polyadenylation [poly(A)] of messenger RNA (mRNA) (Fig. 2C). The latter sequences consist of a poly(A) addition signal, a poly(A) addition site, and a downstream guanine (G) and thymidine (T) domain (*5*). Thus, both the 5′ and 3′ signal sequences necessary for production of an mRNA transcript of plus strand genomic DNA are present in the regions adjacent to the ORF. Second, it is possible that this transcript is translated because the sequence of the plus strand ORF possesses the codon periodicity found in protein encoding sequences. Such sequences have been shown to have a nonuniform distribution of G residues where G is preferred in the first (G1) position of triplets and avoided in the second (G2) position (*6*). Thus, a codon periodicity of G-nonG-any nucleotide is maintained in protein encoding sequences, but not in noncoding sequences, presumably to ensure correct translation of the mRNA on the ribosome. The ratio of G1 to G2 in the plus DNA strand ORF of HIV-1 is 1.5, in good agreement with the previously published values of 1.7 and 1.6 for simian virus 40 and Rous sarcoma virus genes, respectively (*6*). Third, the plus strand ORF is much longer than would be expected to occur by chance because ORFs longer than 300 nucleotides, or 100 codons, are rare in the DNA strands complementary to known genes (*7*). Also, the ORF is present in the *env* sequences of 12 of 13 HIV-1 isolates examined. The solitary isolate lacking this ORF, isolate Z3 (*3*), is a noninfectious proviral genome with a mutation that introduces a premature termination codon in both the *env* and plus strand ORF gene sequences. Fourth, the plus strand ORF spans the region of the HIV *env* gene corresponding to the junction of the gp120 and gp41 proteins (Fig. 1). This region is unusually well conserved among HIV *env* sequences (*8*). The high degree of conservation of the amino acid sequence of the *env* proteins may result from the constraints imposed upon the nucleotide sequence by encoding proteins from both strands. Fifth, an analogous ORF is present in the genome of hepadnaviruses and retroviruses that encode proteins involved in transactivation of virus gene expression (*9*). Therefore, there is substantial theoretical evidence that the DNA plus strand ORF of HIV-1 is a genuine gene sequence.

It is possible, however, that the ORF described here does not represent a true gene sequence. For example, the reading frame may be open because of a random distribution of termination codons. This explanation is unlikely because the plus strand ORF is present in *env* sequences of 12 independent HIV-1 isolates and in the sequences of HIV-related viruses. Alternatively, the *env* gene of HIV-1 might exhibit a bias against codons that are read as termination codons in the complementary translation frame (for example, UCA for serine, and UUA or CUA for leucine). However, this is not the case because the plus strand ORF is not in phase with the *env* gene (that is, the initial AUG of the plus strand ORF is not complementary to a CAU codon in phase with the AUG of the *env* gene). Thus, it is difficult to explain the presence of the plus strand ORF as an artifact of analysis.

It seems plausible, therefore, that the DNA plus strand of HIV-1 encodes a protein during the course of virus infection. Bidirectional transcription of overlapping genes, although rare, is known to occur.

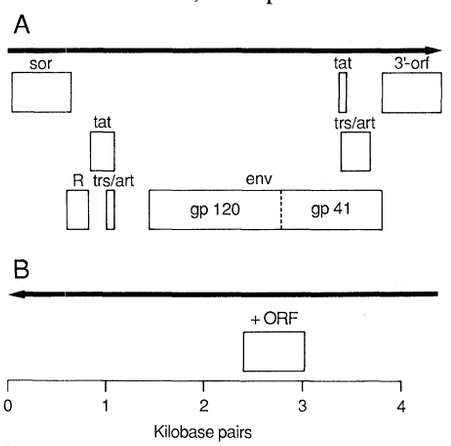

Fig. 1. HIV-1 DNA plus strand ORF. The previously identified ORFs (see text) of the DNA minus strand (**A**) and the newly identified ORF of the DNA plus strand (**B**) at the 3′ end of the double-stranded, linear DNA genome of HIV isolate H3 are shown (*3*). The DNA plus strand ORF (+ ORF) is complementary to nucleotides 7403–7972 of the *env* gene and spans the junction of glycoproteins 120 (gp120) and 41 (gp41). The arrows represent the direction of transcription.

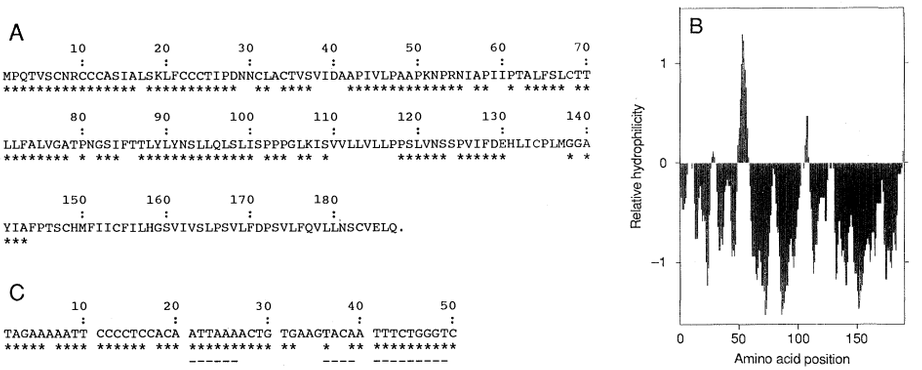

Fig. 2. (**A**) Amino acid sequence of the DNA plus strand ORF of HIV-1. The consensus sequence of the predicted amino acids, beginning with the first Met residue, of 12 HIV-1 isolates is shown (*3*). Asterisks (*) represent identical amino acids in the sequences of all 12 HIV-1 isolates. D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. (**B**) Hydrophilicity profile of the predicted protein encoded by the HIV-1 DNA plus strand ORF. The relative hydrophilicity of the amino acid residues is plotted as a function of amino acid position of the residue on the protein molecule. Values were calculated by the method of Hopp and Woods (*15*). The hydrophilicity profile is strikingly similar to that of MAL, a hydrophobic protein associated with human T cell differentiation (*16*). (**C**) Polyadenylation signal sequences on HIV-1 plus strand DNA. The consensus sequence of the 50 nucleotides, beginning with the termination signal sequence TAG (nucleotides 1 to 3), immediately 3′ of the DNA plus strand ORF are shown. Asterisks (*) identify positions containing identical nucleotides in 12 HIV-1 sequences. The 3′ signal sequences known to be necessary for polyadenylation [poly(A)] of mRNA (*5*) are underlined: ATTAAA (nucleotides 21–26), poly(A) addition signal; TA or CA (nucleotides 36–39), poly(A) addition site; and TTTCTGGGT (nucleotides 41–49), downstream guanine (G) and thymidine (T) domain necessary for efficient poly(A) addition.

There is evidence that both DNA strands of the herpes simplex virus ICP-0 gene (*10*), the murine c-*myc* gene (*11*), the rat insulin II (*12*), a pupal cuticle gene of *Drosophila* (*13*), and the human gonadotropin-releasing hormone gene (*14*) are transcribed. In addition, the complementary DNA strand mRNAs of the last two genes are known to be translated into protein. Thus, there are precedents for bidirectional transcription and concomitant protein synthesis of eukaryotic genes. Therefore, the ORF on the DNA plus strand of HIV may represent a genuine gene sequence that ultimately encodes a protein with novel features.

REFERENCES AND NOTES

1. S. Arya and R. Gallo, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2209 (1986); J. Sodroski *et al.*, *Science* **231**, 1549 (1986); J. Sodroski *et al.*, *Nature (London)* **321**, 412 (1986); N. C. Kan *et al.*, *Science* **231**, 1553 (1986); T.-H. Lee *et al.*, *ibid.*, p. 1546; A. B. Rabson *et al.*, *ibid.* **229**, 1388 (1985); F. Wong-Staal, P. K. Chanda, J. Ghrayeb, *AIDS Res. Hum. Retro.* **3**, 33 (1987).
2. R. Miller and W. Robinson, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2531 (1986); R. Miller, *Science* **236**, 722 (1987).
3. The isolates were as follows: (CDC) S. M. Desai, V. S. Kalyanaraman, J. M. Casey, A. Srinivasan, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8380 (1986); (NY5, Z3) R. Willey, R. Ruthledge, S. Dias, T. Folks, T. Theodore, *ibid.*, p. 5038; (RF, WM) B. R. Starcich, B. H. Hahn, G. M. Shaw, P. D. McNeely, S. Modrow, *Cell* **45**, 637 (1986); (Z6) A. Srinivasan, R. Anand, D. York, J. Curran, GenBank; (ARV2) R. Sanchez-Pescador *et al.*, *Science* **227**, 484 (1985); (BRU) S. Wain-Hobson, P. Sonigo, O. Danos, S. Cole, M. Alizon, *Cell* **40**, 9 (1985); (ELI, MAL) M. Alizon, S. Wain-Hobson, L. Montagnier, P. Sonigo, *Cell* **46**, 63 (1986); (H3) L. Ratner *et al.*, *Nature (London)* **313**, 277 (1985); (H3D) M. A. Muesing *et al.*, *ibid.*, p. 450; (HXB) A. G. Fisher *et al.*, *Science* **233**, 655 (1986).
4. R. M. Myers, K. Tilly, T. Maniatis, *Science* **232**, 613 (1986); W. S. Dynan and R. Tjian, *Nature (London)* **316**, 774 (1985).
5. M. L. Birnstiel, M. Busslinger, K. Strub, *Cell* **41**, 349 (1985); C. Swimmer and T. Shenk, *Nucleic Acids Res.* **13**, 8053 (1985).
6. E. N. Trifonov, *J. Mol. Biol.* **194**, 643 (1987).
7. A. Casino *et al.*, *Nucleic Acids Res.* **9**, 1499 (1981); P. Sharp, *ibid.*, p. 1389.
8. F. Wong-Staal and R. C. Gallo, *Nature (London)* **317**, 395 (1985).
9. A plus strand DNA ORF of approximately 200 amino acids is also present in the X gene region (the 3′ end of the linear genome) of 18 out of 18 isolates of hepatitis B virus (HBV) and HBV-like viruses of animals. Other, shorter ORFs are found in the DNA plus strand complementary to: (i) the gene sequence encoding the transactivating protein of human T cell leukemia virus types 1 and 2, bovine leukemia virus, and simian T cell leukemia virus; and (ii) the *env* region of HIV type 2, visna virus of sheep, equine infectious anemia virus, and the simian immunodeficiency virus. The plus strand DNA ORFs of these viruses are heterogeneous in size and many do not possess initiation codons within the 5′ region of the ORF. Computer analysis suggests that the plus strand ORF of the ancestral virus of HIV-1, and the other viruses, was once much longer, but is gradually shrinking in size because of the accumulation of termination codons. It is possible that the plus DNA strand of many of these viruses has lost the ability to encode protein during the course of evolution.
10. J. G. Stevens, E. K. Wagner, G. B. Devi-Rao, M. L. Cook, L. T. Feldman, *Science* **235**, 1056 (1987); K. D. Croen *et al. N. Engl. J. Med.* **317**, 1427 (1987); D. L. Rock *et al.*, *J. Virol.* **61**, 3820 (1987); D. L. Rock. S. L. Beam, J. E. Mayfield, *ibid.*, p. 3827; J. G. Spivack and N. W. Fraser, *ibid.*, p. 3841.
11. M. Kindy *et al.*, *Mol. Cell. Biol.* **7**, 2857 (1987).
12. S. Efrat and D. Hanahan, *Mol. Cell. Biol.* **7**, 192 (1987).
13. S. Henikoff, M. Keene, K. Fechtel, J. Fristrom, *Cell* **44**, 33 (1986).
14. J. Adelman, C. Bond, J. Douglass, E. Herbert, *Science* **235**, 1514 (1987).
15. T. P. Hopp and K. R. Woods, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3824 (1981).
16. M. Alonso and S. Weissman, *ibid.* **84**, 1997 (1987).

17. Computer-assisted nucleic acid and protein analysis was through the BIONET National Computer Resource for Molecular Biology, which is supported by NIH grant U41-01685-05. I thank R. Chanock, J. Ecker, P. Johnson, M. Martin, R. Purcell, A. Rabson, and J. Woodcock for helpful comments.

18 September 1987; accepted 2 February 1988

Localized Dispersal and Recruitment in Great Barrier Reef Corals: The Helix Experiment

PAUL W. SAMMARCO AND JOHN C. ANDREWS

To examine the problem of how far coral larvae disperse from their natal reef, coral recruitment densities were experimentally determined at distances up to 5 kilometers from a small, relatively isolated platform reef, Helix Reef, on the central Great Barrier Reef for 7 months. High concentrations of recruits, accounting for up to 40 percent of all recruitment, were found downstream of the reef in areas of high water residence time, suggesting that near-field (proximal) circulation has a profound influence on dispersal and recruitment of coral larvae. Coral recruitment declined logarithmically with distance from the reef, decreasing by an order of magnitude at radial distances of only 600 to 1200 meters. On an ecological time scale, advective dispersal of semipassive marine larvae with relatively short planktonic lives (minimally days) may be extensive, but success of recruitment is highly limited. Through evolutionary time, sufficient dispersal occurs to ensure gene flow to reef tracts hundreds or possibly thousands of kilometers apart. In the short term, however, coral reefs appear to be primarily self-seeded with respect to coral larvae.

MANY MARINE ORGANISMS POSsessing larvae with potentially high longevity are assumed to be long-range dispersers (*1*). Short-range dispersal in reproductive propagules with high dispersal capabilities has long been known to occur in the terrestrial environment (*2*), but evidence for such in the marine environment has only recently emerged (*3, 4*), prompting us to surmise that localized dispersal and recruitment may be important in reef corals. Thus we examined coral recruitment around a relatively isolated reef.

Two major modes of coral reproduction are release of fully developed larvae (*5*) and external fertilization (*6, 7*). Estimates of larval longevity vary widely. The minimum post-release period for settlement in brooded planulae is 4 hours (*8*); externally fertilized eggs have an obligate planktonic development period of 24 to 72 hours (*7, 9*). Upper time limits for dispersal and settlement of coral planulae approach 3 months (*10*). Until now, the patterns of effective larval dispersal in corals and the resultant distribution of their settlement have not been known.

Twenty-four oceanographic moorings were deployed around Helix Reef (central Great Barrier Reef: 147°18′E, 18°38′S), which is 800 m in diameter, 10 km from its nearest neighbor, and rises from a continental shelf floor of 55 m depth. Moorings were placed at 0, 0.3, 0.6, 1.2, 2.5, and 5.0 km

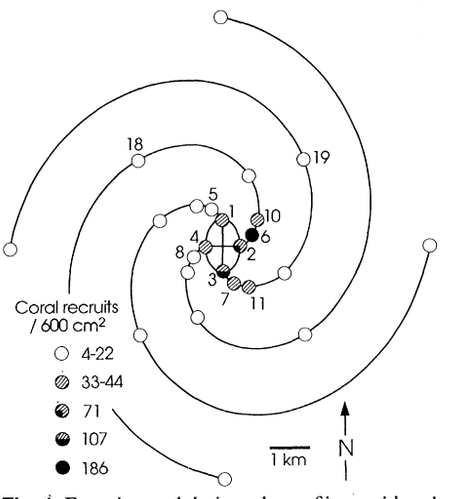

Fig. 1. Experimental design: the reef is considered to be an ellipse; circles represent recruitment sample points where arms of 270° spiral intersect concentric ellipses at 0, 0.3, 0.6, 1.2, 2.5, and 5.0 km from perimeter. Selected station numbers shown. Coral recruitment density is designated by shaded circles; each group differs significantly from next [$P < 0.001$, one-way analysis of variance; $P < 0.01$, sum-of-squares simultaneous test procedure; $(Y + 0.5)^{1/2}$ transformation used for analysis].

Australian Institute of Marine Science, P.M.B. No. 3, Townsville M.C., Queensland, 4810, Australia.