Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 7648–7652, October 1986
Biochemistry

Computer analysis of retroviral *pol* genes: Assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes

(ribonuclease H/reverse transcriptase)

M. S. JOHNSON, M. A. MCCLURE, D.-F. FENG, J. GRAY, AND R. F. DOOLITTLE

Department of Chemistry, University of California at San Diego, La Jolla, CA 92093

Contributed by R. F. Doolittle, June 30, 1986

ABSTRACT **A computer analysis of the amino acid sequences from the putative gene products of retroviral *pol* genes has revealed a 150-residue segment that is homologous with the ribonuclease H of *Escherichia coli*. The segment occurs at the carboxyl terminus of the region assigned to the 90-kDa reverse transcriptase polypeptide. In contrast, a section nearer the amino terminus of this sequence can be aligned with nonretroviral polymerases. The order of activities in the *pol* gene is thus: polymerase–ribonuclease–endonuclease. On another note, all retroviral endonuclease sequences contain a consensus zinc-binding "finger." This should not be confused with the well-known zinc requirement of reverse transcriptases.**

We have been conducting a computer analysis of retroviral protein sequence relationships. During the course of this study, we uncovered a number of unexpected features among the inferred products of the retroviral polymerase gene. In particular, we identified sequences in the *pol* gene products that are clearly related to some nonretroviral enzymes. The results lead to a functional arrangement of activities in the *pol* gene region that differs from that reported by others (1–3). The activities that are under scrutiny here and that are encompassed by the *pol* gene include: the RNA-directed DNA polymerase (reverse transcriptase; EC 2.7.7.49) (4), a ribonuclease H that degrades the viral RNA in the immediate wake of its reverse transcription (5), and an endonuclease ("integrase") that is essential for the integration of the newly synthesized DNA into the host genome (6).

The *pol* gene of retroviruses is expressed initially as a gag–pol precursor that is proteolytically processed to a number of small gag proteins, an approximately 90-kDa protein encompassing both RNA-directed DNA polymerase (reverse transcriptase) and ribonuclease H activities, and, finally, a 40-kDa fragment with endonuclease activity (7). Several reports have presented evidence that the ribonuclease H activity of the 90-kDa reverse transcriptase portion is associated with the amino-terminal end of that protein, and by implication, that the DNA polymerase activity is at the carboxyl-terminal end. These conclusions are based on experiments involving deletion mutants (2), on the one hand, and antibodies to synthetic peptides modeled on the putative sequences, on the other (3).

We now suggest that the opposite must be true: the ribonuclease H activity should be situated at the carboxyl terminus, and the DNA polymerase, at the amino terminus. We draw this conclusion on the basis of comparisons of the retroviral sequences with those of nonviral enzymes of similar function. In this regard, we have uncovered a significant resemblance between a 150-residue segment at the carboxyl-terminal end of the 90-kDa fragment and the reported sequence of a ribonuclease H from *Escherichia coli*. We also provide an alignment of a segment near the amino terminus of the 90-kDa polypeptide with highly conserved sequences from many other polymerases, including the α subunit of *E. coli* DNA-directed RNA polymerase. Finally, there is a distinctive sequence in the endonuclease sequence that is characteristic of a zinc-binding segment.

METHODS

The sequences used were taken from the 1985 version of NEWAT (8) or release 6.0 of the National Biomedical Research Foundation Atlas (9). The particular versions of the retroviral sequences employed are: human T-cell leukemia virus type I (HTLV-I), Seiki *et al.* (10); bovine leukemia virus (BLV), Rice *et al.* (11); Rous sarcoma virus (RSV), Schwartz *et al.* (12); mouse Moloney leukemia virus (Mo-MLV), Shinnick *et al.* (13); human immunodeficiency virus (HIV; formerly HTLV-III/LAV), Ratner *et al.* (14); *E. coli* ribonuclease H, Kanaya and Crouch (15); and the α subunit of *E. coli* DNA-directed RNA polymerase, Ovchinnikov *et al.* (16).

The search program used a moving window of 40 residues and a table of weighted values taken from the mutation matrix of Dayhoff *et al.* (17). Alignments were performed with programs based upon the original algorithm of Needleman and Wunsch (18) as described by Feng *et al.* (19).

RESULTS

Ribonuclease H and Polymerase Sequences. The sequence of ribonuclease H from *E. coli* resembles the carboxyl-terminal portion of retroviral reverse transcriptases. In the case of the Mo-MLV comparison, the two segments are 30% identical (Fig. 1). Binary comparison of each of the retroviral sequences with the *E. coli* ribonuclease H sequence, followed by statistical evaluation by a randomization method, gave authentic alignment scores from 4 to 10 standard deviations above the means of the jumbled comparisons. The cumulative weight of the multiple alignment (Fig. 2) further bears out the significance of the overall relationship.

That the polymerase portion of the viral reverse transcriptase system must encompass the amino-terminal portion of the 90-kDa fragment is established by the alignment shown in Fig. 3. The key region here involves a sector previously shown by Kamer and Argos (20) to be present in a number of nonretroviral polymerases; these consistently have two aspartic acid residues surrounded by a set of nonpolar amino acids. To make the point further, we added the sequence of the α subunit of *E. coli* DNA-directed RNA polymerase to the alignment (Fig. 3).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Mo-MLV, mouse Moloney leukemia virus; HIV, human immunodeficiency virus; HTLV-I, human T-cell leukemia virus, type I; BLV, bovine leukemia virus; RSV, Rous sarcoma virus.

Downloaded by guest on November 26, 2020

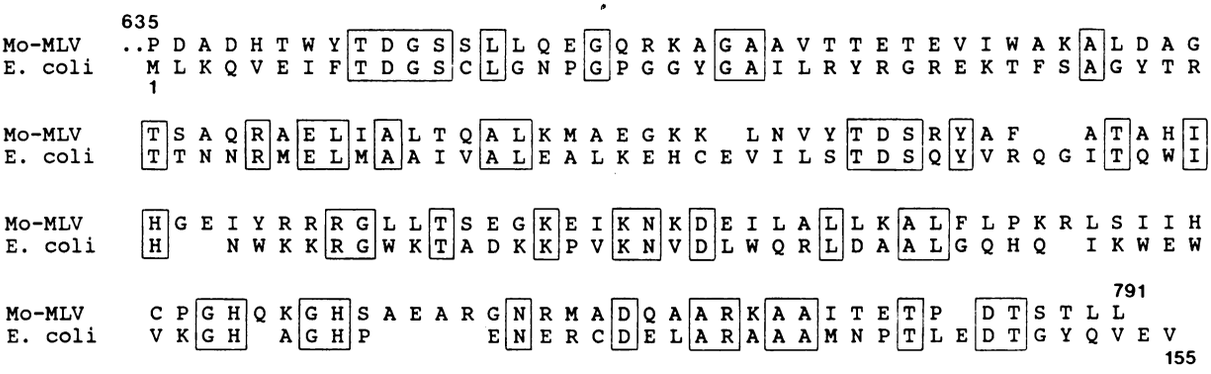

FIG. 1. Alignment of Mo-MLV and *E. coli* ribonuclease H (E. coli). Residues 635–791 of Mo-MLV correspond to the carboxyl-terminal portion of the reverse transcriptase/RNase H portion of the *pol* gene as diagrammed in Fig. 6. The aligned sequences are identical in 30% of the positions as indicated by the boxed residues.

It is known that limited proteolytic digestion of the 90-kDa fragment can give rise to smaller polypeptides with only ribonuclease H activity (21), implying that the two functions exist in two quite different settings. In this regard, the retroviral sequences of the *pol* system are highly conserved over the course of their first 250 residues, a section of about the same dimensions as the α subunit of *E. coli* polymerase, and over the course of their carboxyl-terminal 150 residues, a segment approximating the length of the *E. coli* ribonuclease H protein. Between these two conserved regions, however, is a 200-residue sector that is considerably more variable from one retrovirus to another. As such, it seems to us a good candidate for a connecting tether between the regions embodying the two enzymatic activities.

The Endonuclease Fragment. A thorough search of our data bases did not reveal any obvious relationships between retroviral endonucleases and other proteins, although an intriguing, albeit marginal, resemblance is discernible with a portion of an *E. coli* "transposase" (22). On the other hand, analysis of the retroviral endonuclease sequences did reveal the presence of a constellation of amino acids recently reported to be diagnostic for a zinc-binding site of the sort that can interact with DNA (23–25). In this case, the consensus, which is rigorously conserved in all of the retroviral endonucleases we have examined (upwards of a dozen), involves two histidines separated by 20–30 residues from a brace of closely spaced cysteines (Fig. 4). It has been postulated that the zinc is tetrahedrally coordinated by the histidyl and cysteinyl sidechains and that the residues between the two sets of ligands exist as a ribbon that can wrap around the DNA strand (23). A depiction of the endonuclease segment from HIV is presented in such a form in Fig. 5. The zinc predicted on the basis of this sequence should not be confused with the demonstrated zinc of reverse transcriptases (26, 27); the latter presumably acts in a catalytic fashion in all polymerases.

A number of other residues are highly conserved in the retroviral endonucleases, although the degree of conservation falls off markedly near the carboxyl terminus. This is taken to an extreme in the Mo-MLV sequence, in which case a 36-residue intrusion occurs; so far this segment has not been seen in any of the other retroviral sequences.

DISCUSSION

Several laboratories have reported data that have been interpreted as indicating that the ribonuclease H activity is near the amino-terminal end of the 90-kDa reverse-transcriptase fragment (1–3). As a result, some investigators

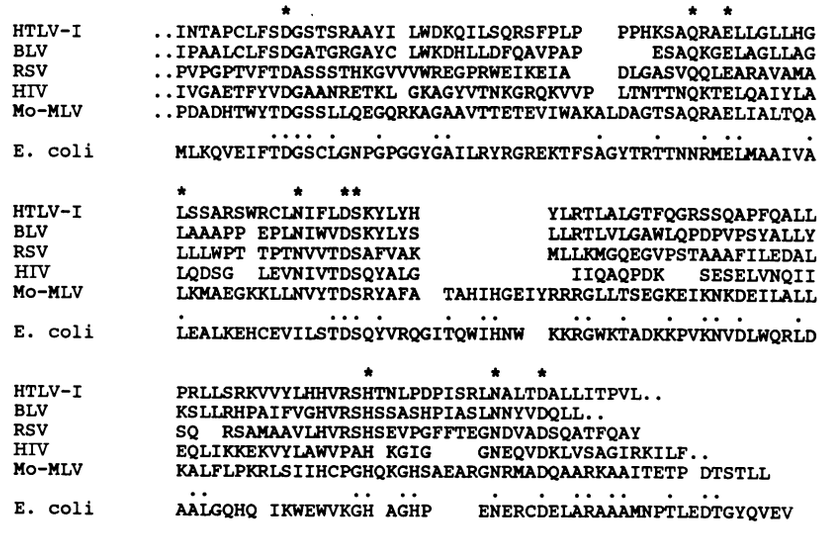

FIG. 2. Alignment of the full-length 155-residue *E. coli* ribonuclease H sequence (E. coli) with five retroviral protein sequences derived from the *pol* genes: HTLV-I, residues 465–599; BLV, 432–557; RSV, 441–572; HIV, 601–728; and Mo-MLV, 635–791. For the HIV and the Mo-MLV sequences, the long open reading frame encodes the protease, which is followed by the *pol*; residue numbering is based on the first position of this long open reading frame. These sequences correspond to the carboxyl-terminal region of the reverse transcriptase/RNase H sequences as depicted in Fig. 6; the experimentally determined carboxyl-terminal residues for RSV, Mo-MLV (7), and *E. coli* RNase H are in fact the last residues shown in the alignment. Asterisks above residues indicate identities in all five of the *pol* sequences; dots between the *E. coli* and the Mo-MLV sequence denote identical residues.

Downloaded by guest on November 26, 2020

```
                                       *     * *     *      *   * *  *
HTLV-I      ..KPERLQALQHL   VRKALEAGHIEPYTGPGNNPVFPVKKA NGTWRFIHDLRATNSLTIDL S
BLV         ..KLERLQALQDL   VHRSLEAGYISPWDGPGNNPVFPVRKP NGTWRFVHDLRATNALTKPI P
RSV         ..PEGKLVALTQL   VEKELQLGHIEPSLSCWNTPVFVIRKA SGSYRLLHDLRAVNAKLVPF G
HIV         ..TEEKIKALVEICTEMEKEGKISKIGPE NPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDF W
Mo-MLV      ..SQEARLGIKPH   IQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP
                      ..    .  . .                 .  .         . .       .
E. coli     ..HDGDVEIVKPQH VICHLTDENASISMRIKVQRGRGYV  PASTRIHSEEDERPIGRLLVDA C

                                    *  *  *            **                  ***
HTLV-I        SSSPGPPDLSSLPTTLAHLQTIDLRDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQ
BLV           ALSPGPPDLTAIPTHLPHIICLDLKDAFFQIPVEDRFRSYFAFTLPTPGGLQPHRRFAWRVLPQ
RSV           AVQQGAPVLSALPRGWPL MVLDLKDCFFSIPLAEQDREAFAFTLPSVNNQAPARRFQWKVLPQ
HIV           EVQLGIPHPAGLKKKKSV TVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQ
Mo-MLV        TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP EMGISGQLTWTRLPQ
                .                   ..                   .               . .
E. coli       YSPVERIAYNV  EAARVEQRTDLDKLVIE            METNGTIDPEEAIRRAATILAE

              *   **          *              * **                      *   *
HTLV-I        GFKNSPTLFEMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDLLLLSEATM ASLISHGLP
BLV           GFINSPALFERALQEPLRQVSAAFSQSLLVSYMDDILIASPTEEQRSQ CYQALAARLRDLGFQ
RSV           GMTCSPTICQLVVGQVLEPLRLKHPSLCMLHYMDDLLLAASSHDGLEA AGEEVISTLERAGFT
HIV           GWKGSPAIFQSSMTKILEPFKKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLT
Mo-MLV        GFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALL QTLGNLGYR
                     . .               . ...  .... .   .   .           .
E. coli       QLEAFVDLRD  VRQPEVKEEKPEFDPILLRPVDDLELTVRS ANCLKAEAIHYIGDLVQRTEV
                                       -----------------

                  *           *                                  *
HTLV-I        VSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSR   WALPELQALLGEIQWVSKGTPTL..
BLV           VASEKTRQTPSPVPFLGQMVHNQIVTYQSLPTLQISSP   ISLHQLQAVLGDLQWVSRGTPTT..
RSV           ISPDKVQREPG VQYLGYKL   GSTYVAPVGL VAEP RIATLWDVQKLVGSLQWLRPALGIP..
HIV           TPDKKHQKEPP FLWMGYEL HPDKWTVQPIVLP  EK DSWTVNDIQKLVGKLNW  ASQIYP..
Mo-MLV        ASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGF..
                 .                  .   ...                 .             .
E. coli       ELLK              TPNLGKKSLTEIKD        VLASRGLSLGMRLENWPPASIADE
```

FIG. 3. A portion (residues 117–329) of the 329-residue *E. coli* α subunit of DNA-directed RNA polymerase (E. coli) is aligned with amino-terminal segments of retroviral pols. The match centers around the polymerase consensus "Asp-Asp" (19) sequence (dashed underline). The reverse transcriptase from HTLV-I (residues 33–279), BLV (7–253), RSV (27–269), HIV (194–439), and Mo-MLV (187–436) are derived from the reverse transcriptase/RNase H sequence as shown in Fig. 6. Asterisks indicate positions where each of the five retroviral sequences have identical residues; dots denote residues in common to both Mo-MLV and the *E. coli* sequence. See Fig. 2 for the identification of sequence codes.

report their results in the context of a "ribonuclease–polymerase–endonuclease" arrangement (28, 29). At the same time, other workers, perhaps unaware of these assignments, have clearly shown that segments of less than 200 residues from the amino terminus can be aligned with portions of nonretroviral polymerases, including those from hepatitis B virus and cauliflower mosaic virus (30), as well as from tobacco mosaic and brome mosaic viruses and several picornaviruses (20). As far as is known, the latter do not exhibit ribonuclease H activity.

The question arises: what could have misled some workers into thinking that the ribonuclease activity is near the amino terminus? The problem seems to have two roots. In the one case, experiments involving a murine leukemia virus mutant with a frameshift in the *pol* gene region revealed that premature chain termination gave rise to a truncated poly-

```
                      *   *                                *  *
HTLV-I       ..QLSPAELhSFThCGQTALT LQGATTTE      ASNILRS   cHAcRGGNPQHQMP RGHI
BLV         ..PLETPEQWhKLThCNSRALSRWPNPRIS       AWDPRSPATLcETcQRLNPTGGGKMRT I
RSV           PLPEAKDLhTALhIGPRALSKACNISMQQ      AREVVQT   cPHc   NSAPALE AGVN
HIV          ..IDKAQDEhEKYhSNWRAMASDFNLPPVV      AKEIVAS   cDKc   QLKG    EAMH
Mo-MLV       ..FELDEFLhQLThLSFSKMKALLERSHSPYYMLNRDRTLKNITETcKAcAQVNASKSAVKQGTR
                      -----------------------------------------

                       *  * *             *     *           *
HTLV-I        RRGLLPNHIWQGDITHFKYKNTLYR LHVWVDTFSGAISATQKRKETSSEAISSLLQAIAHLGK
BLV           QRGWAPNHIWQADITHYKYKQFTYA LHVFVDTYSGATHASAKRGLTTQMTIEGLLEAIVHLGR
RSV           PRGLGPLQIWQTDFT LEPRMAPRSWLAVTVDTASSAIVVTQHGRVTSVAVQHHWATAIAVLGR
HIV           GQVDCSPGIWQLDCTHLEGKVIL     VAVHVASGYIEAEVIPAETGQETA YFLLKL  AGR
Mo-MLV        VRGHRPGTHWEIDFTEIKPGLYGYKYLLVFIDTFSGWIEAFPTKKETAKVVTKKLLEEIFPRFG

                     ** *    *            *               **  *   *
HTLV-I          PSYINTDNGPAYISQDFLNMCTSL AIRHTTHVPYNPTSSGLVERSNGILKTLLYKY
BLV             PKKLNTDQGANYTSKTFVRFCQQF GISLSHHVPYNPTSSGLVERTNGLLKLLLSKY
RSV             PKAIKTDNGSCFTSKSTREWLARW GIAHTTGIPGNSQGQAMVERANRLLKDRIRVLAEGDG
HIV           WPVKTIHTDNGSNFTSATVKA ACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQV
Mo-MLV        M PQVLGTDNGPAFVSKVTQT VADLLGIDWKLHCAYRPQSSGQVERMNRTIKETLTKL  TLA

                              *
HTLV-I         FTDKPDLPMDNALSIALWTINHL    NVLTNCHKTRWQLHHS  PRLQPIP ETRSLSNKQT
BLV            HLDEPHLPMTQALSRALWTHNQI    NLL PILKTRWELHHS  PPLAVIS EGGETPKGSD
RSV           FMKRIPTSKQGELLAKAMYALNHFERGENTKTPIQK HWRPTVLTEGPPVKIRIETGEWEKGWN
HIV              RDQAEHLKTAVQMAVFIHN            FKRKGGIGGYSAGERIVDIIATDIQTKELQ
Mo-MLV        TGSRDWVLLLPLALYRAR          NTPGPHGLTPYEILYGAPPPLVNFPDPDMTRVTNSP

                                                                      *
HTLV-I        HWYYFKLPGLNSRQ                                          WKGPQEAL
BLV           KLFLYKLPGQNNRR                                          WLGPLPAL
RSV           VLVWGRGYAAVKNRDTDKVI                                    WVPSRKVK
HIV           KQITKIQNFRVYYRDSRNPL                                    WKGPAKLL
Mo-MLV        SLQAHLQALYLVQHEVWRPLAAAYQEQLDRPVVPHPYRVGDTVWVRRHQTKNLEPRWKGPYTVL

HTLV-I        QEAAGAALI       PVSASSAQWIPWRLLKRAACPRPVGGP ADPKEKDLQHHG
BLV           VEASGGALL         ATNPPVWVPWRLLKAFKCPKN DGP EDAHNRSSDG
RSV           PDITQKDEVTKKDEASPLFAGISDWIPWEDEQEGLQGETASNKQERPGEDTLAANES
HIV           WKGEGAVVIQDNSDIK  VVPRRKAKIIRDYGKQMAGDDCVASRQDED
Mo-MLV        LTTPTALKV          DGIAAWIHAAHVKAADPGGGPSSRLTWRVQRSQNPLKIRLTREAP
```

FIG. 4. Alignment of the endonuclease sequences of retroviral *pol* sequences: HTLV-I (residues 600–896), BLV (558–852), RSV (573–895), HIV (732–1015), and Mo-MLV (841–1199). The sequences contain a pair of histidines (lower-case letters) and cysteines (lower-case letters) in the amino-terminal portion of the sequences that may coordinate a zinc metal ion (dashed underline) and form a nucleotide binding finger (see Fig. 5). The amino-terminal residue of RSV shown in the alignment is known to be the amino-terminal residue of that endonuclease (7), as is depicted in Fig. 6; in each case, the putative carboxyl-terminal residues that complete the sequence are based on the stop codon that terminates each pol polyprotein. Asterisks have been placed above residues that are identical in each of the five sequences. See Fig. 2 for the identification of sequence codes.

Downloaded by guest on November 26, 2020

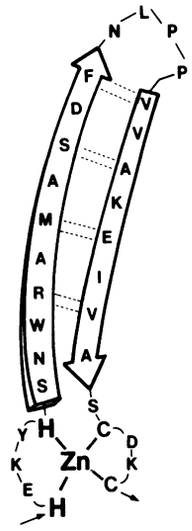

FIG. 5. Ribbon drawing of the proposed zinc metal-binding finger for the retroviral endonuclease of HIV virus that is essential for integration of newly synthesized DNA into the host genome. Dotted lines indicate potential hydrogen bonds oriented along a presumed β structure. Arrows indicate the direction of the chain from the amino to the carboxyl terminus.

merase about two-thirds of normal size. In line with their main goal, these workers accurately demonstrated that the retroviral protease activity must be upstream of the mutated region. They extended their interpretation, however, to include a presumption about the location of the ribonuclease H. They noted that a small amount of transcription took place in their system, and this led them to believe that both ribonuclease H and polymerase activities must lie within the amino-terminal portion of the polymerase protein. It should be pointed out, however, that second-strand synthesis was not reported by those authors, a situation that would be consistent with the absence of ribonuclease H activity.

Another set of experiments to the contrary involved a protein fragmentation study in conjunction with antibodies to synthetic peptides. Grandgenett *et al.* (3) synthesized a series of peptides corresponding to various parts of the *pol* gene product sequence and raised antibodies to them. They then fragmented the equivalent of the 90-kDa fragment from RSV by a route previously reported to yield ribonuclease H activity associated with a 24-kDa fragment (22). In fact, their antibodies to peptides based on the amino-terminal region reacted with a fragment that they presumed to be 24 kDa. What was not commented on, however, but is clearly shown in the published photograph (3) is that antibodies to peptides from the carboxyl-terminal region of the 90-kDa fragment reacted with a somewhat smaller component, consistent with what might be expected for the approximately 150-residue sequence we have assigned to the ribonuclease H.

It can also be asked how it was that workers who reported the *E. coli* ribonuclease H sequence did not notice the resemblance to retroviral sequences. In fact, Kanaya and Crouch (15) compared the *E. coli* sequence to that of the RSV *pol* gene product, but the computer dot-matrix method they used (31) was apparently not sensitive enough to bring out the similarity.

It should be noted that the ribonuclease H from *E. coli* is an endonuclease, whereas the ribonucleases H from retroviruses are exonucleases (1). Similarly, the *E. coli* polymerase that we have aligned with the amino-terminal region of the retrovirus *pol* gene product is a DNA-directed RNA polymerase. These differences notwithstanding, the similarities in sequence are compelling (Figs. 1 and 2), and we contend that the arrangement of activities in the retroviral *pol* gene is "polymerase–ribonuclease–endonuclease." Moreover, the polymerase and ribonuclease functions are separated by a poorly conserved region that may be a tether between two better-defined structures (Fig. 6). The basis for these functional assignments is the similarity in sequence to nonviral enzymes of similar function.

We thank Inder Verma for helpful comments on the manuscript. This work was supported by a grant from the American Cancer Society and National Institutes of Health Grant GM-34434.

1. Crouch, R. J. & Dirksen, M.-L. (1982) in *Nucleases*, eds. Linn, S. M. & Roberts, R. J. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 211–241.
2. Levin, J. G., Hu, S. C., Rein, A., Messer, L. I. & Gerwin, B. I. (1984) *J. Virol.* **51**, 470–478.
3. Grandgenett, D., Quinn, T., Hippenmeyer, P. J. & Oroszlan, S. (1985) *J. Biol. Chem.* **260**, 8243–8249.
4. Temin, H. M. & Mizutani, S. (1970) *Nature (London)* **226**, 1211–1213.
5. Verma, I. M. (1975) *J. Virol.* **15**, 121–126.
6. Schwartzberg, P. J., Colicelli, J. & Goff, S. P. (1984) *Cell* **37**, 1043–1052.

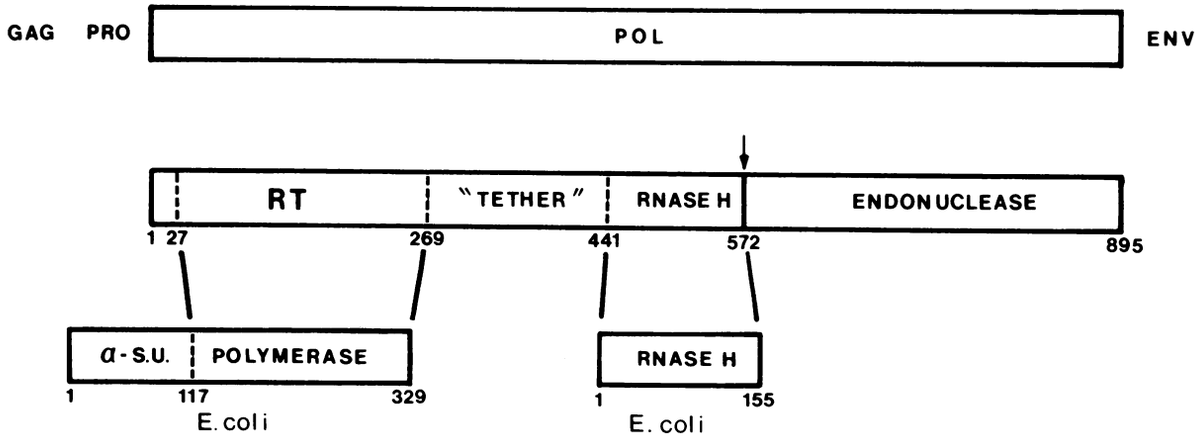

FIG. 6. Postulated map of the retroviral *pol* gene. The *pol* gene is between the *gag* (GAG)-protease (PRO) genes and the *env* (ENV) gene. The protein product, shown in this case for RSV, is depicted with functional assignments assigned to specific regions of the 895-residue product. The RSV pol polyprotein is known to be proteolytically cleaved at a Tyr–Pro bond (arrow), giving rise to the reverse transcriptase (RT)/RNase H polypeptide and the endonuclease (7). The RNase H was localized to the carboxyl terminus of the RT/RNase H protein by virtue of its homology with the RNase H of *E. coli*; similarly, the *E. coli* α subunit of RNA-directed DNA polymerase can be aligned with residues at the amino terminus of the polypeptide.

Downloaded by guest on November 26, 2020

7. Van Beveren, C., Coffin, J. & Hughes, S. (1985) in *RNA Tumor Viruses/Supplements and Appendixes*, eds. Weiss, R., Teich, N., Varmus, H. & Coffin, J. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), 2nd Ed., pp. 589–594; pp. 773–779.
8. Doolittle, R. F. (1981) *Science* **214,** 149–159.
9. George, D. G., Barker, W. C. & Hunt, L. T. (1986) *Nucleic Acids Res.* **14,** 11–15.
10. Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3618–3622.
11. Rice, N. R., Stephens, R. M., Burny, A. & Gilden, R. V. (1985) *Virology* **142,** 357–377.
12. Schwartz, D. E., Tizard, R. & Gilbert, W. (1983) *Cell* **32,** 853–869.
13. Shinnick, T. M., Lerner, R. A. & Sutcliff, J. G. (1981) *Nature (London)* **293,** 543–548.
14. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Pettewat, S. R., Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C. & Wong-Staal, F. (1985) *Nature (London)* **313,** 277–284.
15. Kanaya, S. & Crouch, R. J. (1983) *J. Biol. Chem.* **258,** 1276–1281.
16. Ovchinnikov, Yu. A., Lipkin, V. M., Modyanov, N. N., Chertov, O. Yu. & Smirnov, Yu. V. (1977) *FEBS Lett.* **76,** 108–111.
17. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Nat. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 345–358.
18. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol.* **48,** 443–453.
19. Feng, D.-F., Johnson, M. S. & Doolittle, R. F. (1985) *J. Mol. Evol.* **21,** 112–125.
20. Kamer, G. & Argos, P. (1984) *Nucleic Acids Res.* **12,** 7269–7282.
21. Lai, M.-H. T. & Verma, I. M. (1978) *J. Virol.* **25,** 652–663.
22. Heffron, F., McCarthy, B. J., Ohtsubo, H. & Ohtsubo, E. (1979) *Cell* **18,** 1153–1163.
23. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4,** 1609–1614.
24. Hartshorne, T. A., Blumberg, H. & Young, E. T. (1986) *Nature (London)* **320,** 283–287.
25. Berg, J. M. (1986) *Science* **232,** 485–487.
26. Poiesz, B. J., Battula, N. & Loeb, L. A. (1974) *Biochem. Biophys. Res. Commun.* **56,** 959–964.
27. Auld, D. S., Kawaguchi, H., Livingston, D. M. & Vallee, B. L. (1975) *Biochem. Biophys. Res. Commun.* **62,** 296–302.
28. Dunwiddie, C. & Faras, A. J. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 5097–5101.
29. Freeman-Wittig, M.-J., Vinocour, M. & Lewis, R. A. (1986) *Biochemistry* **25,** 3050–3055.
30. Toh, H., Hayashida, H. & Miyata, T. (1983) *Nature (London)* **305,** 827–829.
31. Maizel, J. V., Jr., & Lenk, R. P. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 7665–7669.

Downloaded by guest on November 26, 2020