

Nucleotide sequence of the integrase (IN) gene of an endogenous murine leukemia retroviral DNA

Padmini Kedar and Arifa S.Khan*

Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

Submitted April 26, 1990

EMBL accession no. X52622

We report the nucleotide (1) and predicted amino acid sequence of a complete IN gene of an endogenous mink cell focus-forming (MCF)-related murine leukemia virus (MuLV) DNA, B-34, which was cloned from a BALB/c mouse embryo DNA library (2). The B-34 IN gene had 82% nucleotide sequence homology and 86% amino acid sequence homology with the IN gene of a known infectious AKR ecotropic MuLV DNA (3). The majority of differences were due to base substitutions except for a 12 bp insertion (nucleotides 1159–1170 in the figure) resulting in 4 additional amino acids in B-34 DNA. The presence of the latter sequence is unique to endogenous MCF-related MuLV DNAs (4). B-34 IN gene had 96% nucleotide and amino acid sequence

homology to the IN gene of another endogenous MuLV DNA cloned from a RFM/Un mouse; the latter however had a premature termination codon located 19 amino acids from the 3' end of IN (5). We are currently investigating the biological properties of B-34 IN.

REFERENCES

1. Sanger, F. *et al.* (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
2. Khan, A.S. *et al.* (1982) *J. Virol.* **44**, 625–636.
3. Herr, W. (1984) *J. Virol.* **49**, 471–478.
4. Khan, A.S. (1984) *J. Virol.* **50**, 864–871.
5. Nikbakht, K.N. *et al.* (1987) *J. Gen. Virol.* **68**, 683–693.

```

I E A S T P Y T P D F F H Y T E T D I K N L R E L G A T Y D
1 ATAGAGGCTCAACCCGTACACTCCAGACTTCTTCCATTATACTGAAACAGATATAAGAACCTACGAGAGTTGGGAGCCACATATGAT
  R E K K Y W V L Q G K P V M P D Q F T F E L L D F L H Q L T
91 AGAGAGAAAAAATATTGGGTCTGCAAGTAAACCTGTGATGCTGACCAGTTCACCTTTGAATTATTAGACTTCCTTCACCAGCTCACC
  H L S Y Q K M R A L L D R K E S P Y Y M L N K D K I L H E V
181 CACCTTAGCTATCAGAAGATGAGGGCACTTCTAGACAGAAAAGAAAGCCCTATTACATGCTAAATAAAGATAAGATCCTCCACGAGGTG
  A E S C Q A C V Q V N A S K A K V G P G V R V R G H R P G T
217 GCGAATCATGCCAAGCTGTGTCCAAGTAAATGCCAGTAAAGCTAAGTCCGGTCCCGGGTCCGAGTAAGAGGACATCGACCAGGCACC
  H W E I D F T E V R P G L Y G H K Y L L V F V D T F S G W V
361 CATTGGGAAATTGACTTTACTGAAGTAAGGCCCGGACTGTATGGGCATAAGTATCTTCTGGTGTGTTGTGGACACGTTCTCTGGCTGGGTG
  E A F P T K H E T A K V V T K K L L E E I F P R F G M P Q V
451 GAAGCCTTCCCAACCAAGCATGAGACTGCCAAAGTTGTGACCAAAAAGCTTCTAGAAGAAATATTTCCAAGGTTTGAATGCCCAAGTA
  L G T D N G P A F V S Q V S Q S V A K L L G I D W K L H C A
541 TTGGGACTGATAATGGGCTGCCTTCGTCTCCAGGTAAGTCAGTCCGGTGGCCAAGCTACTGGGATTGATTGGAACTACATTGTGCT
  Y R P Q S S G Q V E R M N R T I K E T L T K L T L A T G T R
631 TACAGACCCAGAGTTCAGGTCAGGTAGAAAAGTGAATAGGACAATCAAGGAGACTTTGACCAATTAACGCTTGAACCTGGCACTAGA
  D W V L L L P L A L Y R A R N T P G P H G L T P Y E I L Y G
721 GACTGGGTACTCCTACTTCCCTTGGCCCTCTACCGAGCCGCAACACTCCGGGCCCCCATGGACTCACTCCGTATGAAATCCTGTATGGG
  A P P P L V N F H D P E M S K F T N S P S L Q A H L Q A L Q
811 GCGCCCCGCCCCCTTGTAAATTTCCATGATCCTGAAATGTCAAAGTTTACTAATAGCCCCCTCTCCAAGCTCACTTACAGGCCCTCCAA
  D T V W V R R H Q T K N L E P R W K G P Y T V L L T T P T A
901 GCGGTACAACGAGAGGTCTGGAAGCCTCTGGCCGCTGCCTATCAGGACCAACAAGACCAGCCTGTGATACCACACCCCTTCTGGTCCGGC
  A V Q R E V W K P L A A A Y Q D Q D Q P V I P H P F L V G
991 GACACCGTGTGGGTACGCCGGCACCAGACTAAGAAGTGAACCTCGCTGGAAAGGACCTACACCGTCTGTGACCACCCCAACCGCT
  L K V D G I A A W I H A A H V K A A T T P P A G T A S G P T
1081 CTCAAAGTAGACGGCATCGCTGCGTGGATCCACGCCGCTCACGTAAGCGGCGACAACCCCTCCGGCCGGAACAGCATCAGGACCGACA
  W K V Q R S Q N P L K I R L T R G P P *
1171 TGAAGGTCCAGCGTTCTCAAACCCCTTAAAGATAAGATTAACCCGTGGGCCCCCTGA

```

* To whom correspondence should be addressed