J Mol Evol (1992) 35:551–556

Journal of Molecular Evolution

Letter to the Editor

A Redefinition of the Asp-Asp Domain of Reverse Transcriptases

A.M. Velasco,[1] L. Medrano,[1] A. Lazcano,[1] and J. Oró[2]

[1] Departamento de Biología, Facultad de Ciencias, UNAM, Apartado Postal 70-407, Cd. Universitaria, México 04511, D.F. Mexico
[2] Department of Biochemical and Biophysical Sciences, University of Houston, Houston, TX 77204-5500, USA

Summary. The rules defining the Asp-Asp domain of RNA-dependent polymerases deduced by Argos (1988) were tested in a set of 53 putative reverse transcriptases (RTs) sequences. Since it was found that some of these rules are not followed by RTs coded by bacteria, group II introns, and non-LTR retrotransposons, we present here a more strict definition of the Asp-Asp domain.

Key words: DD domain — Reverse transcriptase — Polymerases

Introduction

The existence of a highly conserved 14-amino-acid-residue segment consisting of an Asp-Asp pair flanked by hydrophobic amino acids was first recognized by Kamer and Argos (1984) in a set of 15 reverse transcriptases (RTs) and RNA-dependent RNA polymerases. This motif, also known as the Asp-Asp (DD) domain, has also been described in cellular DNA-dependent RNA polymerases (Lazcano et al. 1988). Sequence analyses of the primary structure of viral and cellular DNA-dependent DNA polymerases have shown the existence of an equivalent domain, formed by hydrophobic amino acids surrounding an Asp-Thr-Asp-Ser tetrad (Argos 1988; Wong et al. 1988; Wang et al. 1989; Bernad et al. 1990).

Although some polymerases lack the DD domain (DTDS, in the case of DNA-dependent DNA polymerases, Argos, 1988), experimental evidence of its functional role includes the inhibition of enzymatic activity by (1) antibodies raised against a small phage protein containing the sequence LIVYSDDYLSLM (Zavriev and Borisova 1987); (2) site-specific mutagenesis of the HIV-1 RT (Larder et al. 1987; Hizi et al. 1988); and (3) amino acid substitutions involving the Qβ replicase Asp-Asp motif (Inokuchi and Hirashima 1987). The existence of a DD domain in the primary structure of the eubacterial polynucleotide phosphorylase (Régnier et al. 1987), a nontemplate-dependent polymerase, suggests that this motif is involved in interaction with the nucleotide substrates and not with DNA or RNA templates. Since the carboxylic side chain of aspartic acid can interact with Mg^{2+} (Argos 1988), Zn^{2+}, and other cations, and can also form hydrogen bonds with nucleic acid bases (Helene and Maurizot 1981), it is likely that the DD domain is part of the metal and substrate binding site of both RTs and RNA polymerases (cf. Lazcano et al. 1992). Similar conclusions have been reached for the ϕ29 DNA polymerase YGDTDS motif (Bernad et al. 1990; Blanco et al. 1991).

Abbreviations: RT, reverse transcriptase; DD, Asp-Asp; ORF, open reading frame; mt, mitochondria; pt, plastid

Offprint requests to: A. Lazcano

Analysis of a databank formed by sequences of 11 DNA-dependent DNA polymerases, 26 RNA-dependent RNA polymerases, and 20 RTs led Argos (1988) to define qualitatively the DD domain by a set of seven rules. As applied to the DD domain of RTs, these rules should be read as follows:

Rule 1. Positions 8 and 9 must be occupied by Asp-Asp.
Rule 2. Position 7 can be occupied by Gly, Met, Cys, Val, or Leu.
Rule 3. Position 6 can be occupied by Tyr, Ala, Phe, Ser, Asn, Cys, Gly, Ile, or Met.
Rule 4. At least two of the amino acid residues in positions 1 to 5, and 10 to 14, must be hydrophobic (Ala, Val, Leu, Ile, Cys, Met, Phe, Tyr, His, Trp or Pro).
Rule 5. If only two amino acid residues in positions 1 to 5 are hydrophobic, then Ser or Gly must also be present in this segment. If only two residues in positions 10 to 14 are hydrophobic, then Ser must be present in this portion.
Rule 6. Charged amino acids (Lys, Arg, Asp, Glu, Gln, and Asn) must be absent from position 4.
Rule 7. Hydrophobic amino acid residues must be present in positions 11 and 12.

In recent years the interest in the evolution of retroviruses and related genetic entities sparked by the AIDS epidemic has led to an increasingly large set of RT sequences from different sources, including prokaryotes, viruses, transposable elements, mitochondrial (mt) plasmids, and introns. Analysis of the DD domains of these additional RT sequences has shown that many of them do not follow completely the rules described here. In this paper some of Argos's rules are redefined in order to provide a more stringent definition of the DD domain of RTs.

Material and Methods

RT sequences were located and extracted from both the Genbank/EMBL database and the Human Retroviruses and AIDS 1990 and 1991 databases (G.Myers, Los Alamos National Laboratory). Extracted sequences were compiled in a multiple sequence file. A number of RT sequences were not found in the Genbank/EMBL or Human Retroviruses and AIDS databases and were hand entered into the sequence set. The aligned DD domains of the available RT sequences are shown in Table 1. Unless otherwise noted, the amino acid classification developed by Taylor (1986) was used throughout this comparison. The DD domains marked with an asterisk violate at least one of Argos's rules. Hydropathy profiles were plotted (Fig. 1) using the hydrophilicity values of the scale developed by Hopp and Woods (1981). Equivalent groupings (not shown) were observed when other scales were used. The electric charge and hydropathy plots shown in Fig. 1 were smoothed by averaging the values inside a three-amino-acid-residue sliding window.

Discussion and Conclusions

Analysis of Table 1 shows that with the sole exception of the human endogenous retrovirus K (HERV-K) DD domain, Argos's rules (1988) hold for all viral RTs. However, the same is not true for all nonviral RTs. This observation is consistent with the hydropathy profiles shown in Fig. 1, where it can be seen that approximately one-third of the DD domains of cellular, non-LTR, and group II intron RTs form a coherent group (group B), in which a significant shift toward positive hydropathy values can be observed. In our opinion, the most interesting result of Table 1 is the fact that all the DD domains that do not follow Argos's rules (1988) are found in the primary structure of RTs that form part of the cellular, non-LTR retrotransposon, and group II intron branches of the RT evolutionary tree developed by Xiong and Eickbush (1988). The phylogenetic clustering of the DD domains that violate Argos's rules shows that the exceptions marked with an asterisk in Table 1 are not the result of an artifact.

A detailed analysis of Table 1 shows that rule 5 is partially violated by the *Myxococcus* (Mx 162, Mx 65), green alga (Pt1), and *Podospora* mt (IA-PA) RTs, which have only two hydrophobic amino acids in the 1 to 5 positions, but lack Ser and Gly in this segment. Other Argos rules (1988) can be modified to account for the greater flexibility of cellular, group II introns, and non-LTR retrotransposon RTs in accepting different amino acids in their DD domains. Some of these additions are easy to understand. For instance, Ala, which is a hydrophobic amino acid, joins the list of amino acids that can be present in position 7 (rule 2).The same is true of Leu, which can be added to the list of hydrophobic amino acids that can occupy position 6 (rule 3). Threonine has an hydroxyl group, but its aliphatic side chain explains its presence with other hydrophobic amino acids in position 11 (rule 7).

Four anomalous profiles are easily distinguished in Fig. 1, corresponding to the DD domains of the Bs1, RTL-Cr, LaBelle-1b, and T1 putative RTs. In the latter case, i.e., the *Anopheles gambiae* T1 retrotransposable element (Besanky 1990), the hump observed toward the carboxy end is due to the presence of a lysine in position 11, which according to Argos (1988) should be occupied by a hydrophobic residue. An equivalent phenomenon was described by Argos (1988) for four viral RNA-dependent RNA polymerases in his database which did not obey his rule 7; i.e., position 11 and 12 are not occupied by hydrophobic residues.

Table 1. Aligned DD domains of available RT sequences[a]

Source		DD domain	Reference
* *Escherichia coli*	Ec67	C T Y S R Y A D D I T I S T	Lampson et al. 1989
* *E. coli*	EcB86	L I Y T R Y A D D L T L S A	Lim & Maas 1989
* *Myxococcus xanthus*	Mx162	F T Y T R Y A D D L T F S W	Inouye et al. 1989
* *M. xanthus*	Mx65	Y T Y T R Y A D D L T F S G	Inouye et al. 1990
* *Trypanosoma brucei*	INGI-3	L Q H G F F A D D L T L L A	Kimmel et al. 1987
* *T. brucei*	TRS	L Q H G F F A D D L T L F S	Murphy et al. 1987
* *Chlamydomonas reinhardtii*	RTL-Cr	M D F T I Y A D N F A G V F	Boer & Gray 1988
* Green alga (pt)	Ptl	N T Y C R Y A D D M V I L T	Kück 1989
Dictyostelium discoideum	DIRS-1	V S V I A Y L D D L L I V G	Cappello et al. 1985
Yeast	Ty1	V T I C L F V D D M V L F S	Clare & Farabaugh 1985
* *Saccharomyces cerevisiae* (mt)	a1-Sc	I K Y V R Y A D D I L I G V	Bonitz et al. 1980
* *S. cerevisiae* (mt)	a2-Sc	A Y F V R Y A D D I I I G V	
* *Schizosaccharomyces pombe* (mt)	B1-Sp	L M Y V R Y A D D W I V A V	Lang & Ahne 1985
* *Podospora anserina* (mt)	a1-Pa	I Y Y V R Y A D D W L I G V	Cummings et al. 1989
* *P. anserina* (mt)	IA-Pa	V R Y T R Y A D D W V I G I	
* *Neurospora intermedia* (mt)	LaBelle-1b	A V G S I Y A D E G Y K K V	Pande et al. 1989
* *Bombyx mori*	R1Bm	T E M V A Y A D D V T V L V	Xiong & Eickbush 1988
* *B. mori*	R2Bm	V S A L A Y A D D L V L L A	Burke et al. 1987
* *Anopheles gambiae*	T1	D G H L L Y A D D I K I F L	Besansky 1990
Drosophila melanogaster	Copia	I Y V L L Y V D D V V I A T	Mount & Rubin 1985
* *D. melanogaster*	jockey	V L I A T Y A D D T A V L T	Priimägi et al. 1988
* *D. melanogaster*	F Factor	L T V S T F A D D T A I L S	Di Nocera & Casari 1987
* *D. melanogaster*	I Factor	I K F N A Y A D D F F L I I	Fawcett et al. 1986
D. melanogaster	17.6	K H C L V Y L D D I I V F S	Saigo et al. 1984
D. melanogaster	297	K H C L V Y L D D I I I F S	Inouye et al. 1986
D. melanogaster	412	S Q A F L Y M D D L I V I G	Yuki et al. 1986
D. melanogaster	GYPSY	K I C Y V Y V D D V I I F S	
D. melanogaster	1731	M L I L V Y V D D L I L A C	Fourcade-Peronnet et al. 1988
* Mouse	L1Md	V K I S L L A D D M I V Y I	Loeb et al. 1986
Oenothera (mt)	mtOe	H F V V V Y L D D L V V Y T	Schuster & Brennicke 1987
* *Zea mays*	Bs1	V R V R F R V D D G C V V E	Jin & Bennetzen 1989
Syrian hamster intracisternal A-Particle	IAP-H18	L I V I H Y M D D I L I C H	Ono et al. 1985
Rous sarcoma virus	RSV	L C M L H Y M D D L L L A A	Schwartz et al. 1983
Equine infectious anemia virus	EIAV	V Q L Y Q Y M D D L F V G S	Kawakami et al. 1987
Caprine arthritis encephalitis virus	CAEV	I Q F G I Y M D D I Y I G S	Saltarelli et al. 1990
Feline immunodeficiency virus	FIV	L D I Y Q Y M D D I Y I G S	Olmsted et al. 1989
Moloney murine leukemia virus	MoMLV	L I L L Q Y V D D L L L A A	Shinnick et al. 1981
Bovine immunodeficiency-like virus	BIV	V M L Y Q Y M D D L L I G S	Garvey et al. 1990
Bovine leukemia virus	BLV	S L L V S Y M D D I L Y A S	Sagata et al. 1985
Visna	VISNA	I Q F G I Y M D D I Y I G S	Sonigo et al. 1985
Visna-related ovine lentivirus	SA-OMVV	I Q F G I Y M D D I Y I G S	Querat et al. 1990
Simian retrovirus	SRV-1	M Y I I H Y M D D I L I A G	Power et al. 1986
Simian T-cell lymphotropic virus type III	SIVK6W	V T L V Q Y M D D I L I A S	Franchini et al. 1987
Mazon-Pfizer monkey virus	MPMV	M T I I H Y M D D I L I A G	Sonigo et al. 1986
Baboon endogenous retrovirus	BaEV	V T L L Q Y V D D L L L A A	Kato et al. 1987
Human T-cell leukemia virus type 1	HTLV-I	C T I L Q Y M D D I L L A S	Seiki et al. 1983
Human T-cell leukemia virus type 2	HTLV-II	S T I V Q Y M D D I L L A S	Shimotohno et al. 1985
Human immunodeficiency virus type 1	HIV-1	I V I Y Q Y M D D L Y V G S	Wain-Hobson et al. 1985
Human immunodeficiency virus type 2	HIV-2	V I I I Q Y M D D I I L I A S	Guyader et al. 1987
* Human endogenous retrovirus K	HERV-K	C Y I I H Y I D D I L C A A	Ono et al. 1986
Human endogenous retrovirus C	HERV-C	C V L L Q Y V D D L L L G H	Repaske et al. 1985
Cauliflower mosaic virus	CaMV	K F C C V Y V D D I L V F S	Franck et al. 1980
Carnation etched ring virus	CERV	K Y C C V Y V D D I L V F S	Hull et al. 1986

[a] Asterisks indicate domains that violate at least one of Argos's rules

The three other anomalous DD domains, i.e., those corresponding to the *Bs1*, LaBelle, and RTL-Cr retroelements, may have their peculiarities due to the fact that they form part of nonfunctional RTs. As noted by Jin and Bennetzen (1989), the relatively small *Bs1* element may have inserted itself in the null allele of the maize *Adh1* locus in which it was discovered, with the aid of a RT function provided in *trans*, since the open reading frame (ORF) encoded by *Bs1* lacks convincing sequence similarity to other RTs. The two other profiles correspond to the "DD domains" in which the Asp-Asp pair is in

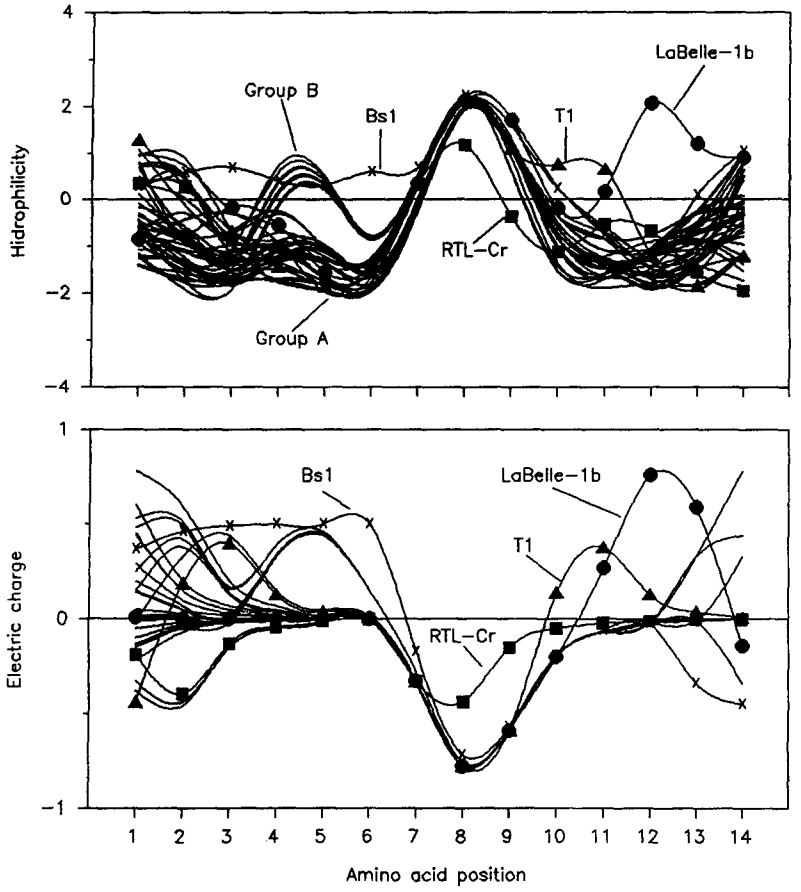

Fig. 1. Smoothed hydropathy and electric charge plots of the RT DD domains of Table 1.

fact substituted by Asp-Glu or Asp-Asn; they are part of the ORF encoded by the *Neurospora intermedia* LaBelle-1b mt plasmid (Pande et al. 1989), and the RTL-Cr from *Chlamydomonas reinhardtii* mt (Boer and Gray 1988), respectively. Under other circumstances, both Glu and Asn would be considered equivalent to Asp, rendering a DE or DN pair equivalent to a Asp-Asp. As argued below, it is likely that the LaBelle and RTL-Cr anomalous DD domains are part of nonfunctional RTs.

In fact, the RTL-Cr is distantly related to fungal group II intron mt RTs, and appears to be a recently acquired gene (Boer and Gray 1988). This could explain the presence of its Asp-Asn pair instead of the typical DD. The Asp-Glu pair of the *N. intermedia* LaBelle-1b plasmid may reflect a different phenomenon. Although Pande et al. (1989) had suggested that the long ORF of 1151 amino acids encoded by the LaBelle plasmid exhibited sequence similarity to different RTs, recent experiments by Schulte and Lambowitz (1991) have shown that this polypeptide is in fact an mt DNA-dependent DNA polymerase in which no reverse transcriptase activity could be detected. These results have led Schulte and Lambowitz (1991) to suggest that the LaBelle protein is actually a DNA polymerase which may be derived from an RT. This conclusion is consistent with the LaBelle profile shown in Fig. 1.

The Asp-Asp domain is so short that the possibility that its almost universal distribution among RTs and other polymerases is due to convergent evolution cannot be dismissed altogether (Lazcano et al. 1992). However, both its presence in functional polymerases and its practical significance in the theoretical (Argos 1988) and experimental determination of polymerases (Yuki et al. 1986) suggest that it should not be substituted for equivalent pairs (i.e., Asp-Glu, Asp-Asn, etc.) in the definition of DD domains. Thus, we suggest that in spite of the exceptions discussed here, i.e., the RTL-Cr and LaBelle elements, Argos (1988) rule 1 should be kept unchanged. We also propose the modification of rules 2, 3, 5, and 6, which should now read as follows:

Rule 2. Position 7 can be occupied by Gly, Ala, Met, Cys, Val, or Leu.
Rule 3. Position 6 can be occupied by Tyr, Ala, Phe, Ser, Asn, Cys, Gly, Leu, Ile, or Met.
Rule 5. If only two amino acids in positions 10 to 14 are hydrophobic, then Ser must be present in this segment.
Rule 6. Position 4 can be occupied by a charged amino acid residue.

The results discussed in this paper suggest that a

detailed characterization of a DD domain should take into account not only Argos's (1988) rules with the modifications discussed here, but also additional information like the hydrophatic profiles and secondary structure determinations, which help us to understand the physicochemical properties of this motif. Further experimental studies on the functional and/or structural role on the DD domain of RTs could benefit from the more stringent definition of this motif provided in this paper.

Note added in proof. The recently published RT sequence from *Stigmatella aurantiaca* (Hsu et al. (1992) J Bacteriol 174:2384–2387) has a DD domain whose composition is consistent with the new rules suggested here.

Acknowledgments. We are indebted to Dr. Gerald Myers (Los Alamos National Laboratory, New Mexico) for kindly providing us with a copy of the Human Retroviruses and AIDS 1990 and 1991 databases. We thank Alejandro Martínez Mena for his help with Fig. 1. This work has been supported in part by CONASIDA-México and Proyecto UNAM.IN 105289 (A.L.) and NASA grant 44-005-002 (J.O.).

References

Argos A (1988) A sequence motif in many polymerases. Nucleic Acid Res 16:9909–9916

Bernad A, Lázaro JM, Salas M, Blanco L (1990) The highly conserved amino acid sequence motif Tyr-Gly-Asp-Thr-Asp-Ser in alpha-like DNA polymerases is required by phage ϕ29 DNA polymerase for protein-primed initiation and polymerization. Proc Natl Acad Sci USA 87:4610–4614

Besanky NJ (1990) A retrotransposable element from the mosquito *Anopheles gambiae.* Mol Cell Biol 10:863–871

Blanco L, Bernad A, Blasco MA, Salas M (1991) A general structure for DNA-dependent DNA polymerases. Gene 100:27–38

Boer PH, Gray MW (1988) Genes encoding a subunit of respiratory NADH dehydrogenase (ND1) and a reverse transcriptase-like protein (RTL) are linked to ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. The EMBO Jour 7:3501–3508

Bonitz SG, Coruzzi G, Thalenfeld BE, Tzagoloff A (1980) Assembly of the mitochondrial membrane system. J Biol Chem 255:11927–11941

Burke WD, Calalang CC, Eickbush TH (1987) The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzymne. Mol Cell Biol 7:2221–2230

Cappello J, Handelsman K, Lodish HF (1985) Sequence of *Dictyostelium* DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. Cell 43:105–115

Clare F, Farabaugh P (1985) Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. Proc Natl Acad Sci USA 82:2829–2833

Cummings DJ, Michel F, McNally KL (1989) DNA sequence analysis of the 24.5 kilobase pair cytochrome oxidase subunit I mitochondrial gene from *Podospora anserina:* a gene with sixteen introns. Curr Genet 16:381–406

Di Nocera PP, Casari G (1987) Related polypeptides are encoded by *Drosophila melanogaster* F elements, I factos, and mammalian L1 sequences. Proc Natl Acad Sci USA 84:5843–5847

Fawcett DH, Lister CK, Kellett E, Finnegan DJ (1986) Transposable elements controlling 1-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINEs. Cell 47:1007–1015

Fourcade-Peronnet F, d'Auriol L, Becker J, Galibert F, Best-Belpomme M (1988) Primary structure and functional organization of *Drosophila* 1731 retrotransposon. Nucleic Acid Res 16:6113–6125

Franchini G, Guo HG, Gallo RC, Collalti E, Fargnoli KA, Hall LF, Wong-Staal F, Reitz MS Jr (1987) Sequence of simian immunodeficiency virus and its relationship to the human immunodeficiency viruses. Nature 328:539–543

Franck A, Guilley H, Jonard G, Richards K, Hirth L (1980) Nucleotide sequence of cauliflower mosaic virus DNA. Cell 21:285–294

Garvey KJ, Oberste MS, Elser JE, Braun MJ, Gonda MA (1990) Nucleotide sequence and genome organization of biologically active proviruses of the bovine immunodeficiency-like virus. Virology 175:391–409

Guyader M, Emerman M, Sonigo P, Clavel F, Montagnier L, Alizon M (1987) Genome organization and transactivation of the human immunodeficiency virus type 2. Nature 326:662–669

Helene C, Maurizot JC (1981) Interactions of oligopeptides with nucleic acids. CRC Critical Rev Biochem 10:213–258

Hizi A, McGill C, Hughes SH (1988) Expression of soluble, enzymatically active, human immunodeficiency virus reverse transcriptase in *Escherichia coli* and analysis of mutants. Proc Natl Acad Sci USA 85:1218–1222

Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci USA 78:3824–3828

Hull R, Sadler J, Longstaff M (1986) The sequence of carnation etched ring virus DNA: comparison with cauliflower mosaic virus and retrovirus. The EMBO J 5:3083–3090

Inokuchi Y, Hirashima A (1987) Interference with viral infection by defective RNA replicase. J Virol 61:3946–3949

Inouye S, Yuki S, Saigo K (1986) Complete nucleotide sequence and genome organization of a *Drosophila* transposable genetic element, 297. Eur J Biochem 154:417–425

Inouye S, Hsu MY, Eagle S, Inouye M (1989) Reverse transcriptase associated with the biosynthesis of the branched RNA-linked msDNA in *Myxococcus xanthus.* Cell 56:709–717

Inouye S, Herzer PJ, Inouye M (1990) Two independent retrons with highly diverse reverse transcriptase in *Myxococcus xanthus.* Proc Natl Acad Sci USA 87:942–945

Jin YK, Bennetzen JL (1989) Structure and coding properties of *Bs1,* a maize retrovirus-like transposon. Proc Natl Acad Sci USA 86:6235–6239

Kamer G, Argos A (1984) Primary structural comparisons of RNA dependent polymerases from plant, animal and bacterial viruses. Nucleic Acid Res 12:7269–7282

Kato S, Matsuo M, Nishimura N, Takahashi N, Takano T (1987) The entire nucleotide sequence of baboon endogenous virus DNA: a chimeric genome structure of murine type C and simian type D retrovirus. Jpn J Genet 62:127–137

Kawakami T, Sherman L, Dahlberg J, Gazit A, Yaniv A, Tronick SR, Aaronson SA (1987) Nucleotide sequence analysis of equine infectious anemia virus proviral DNA. Virology 158: 300–312

Kimmel BE, Ole-Moiyoi OK, Young JR (1987) Ingi, a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. Mol Cell Biol 7:1465–1475

Kück U (1989) The intron of a plastid gene from a green algae contains an open reading frame for a reverse transcriptase-like enzyme. Mol Gen Genet 218:257–265

Lampson BC, Sun J, Hsu MY, Vallejo-Ramírez J, Inouye S, Inouye M (1989) Reverse transcriptase in a clinical strain of Escherichia coli: production of branched RNA-linked msDNA. Science 243:1033–1038

Lang BF, Ahne F (1985) The mitochondrial genome of the fission yeast *Schizosaccharomyces pombe*. J Mol Biol 184:353–366

Larder BA, Purifoy DJM, Powell KL, Darby G (1987) Site specific mutagenesis of AIDS virus reverse transcriptase. Nature 327:716–717

Lazcano A, Fastag J, Gariglio P, Ramírez C, Oró, J (1988) On the early evolution of RNA polymerase. J Mol Evol 27:365–376

Lazcano A, Valverde V, Hernández G, Gariglio P, Fox GE, Oró J (1992) On the early emergence of reverse transcription: theoretical basis and experimental evidence. J Mol Evol 35:524–536

Lim D, Maas WK (1989) Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. Cell 56:891–904

Loeb DD, Padgett RW, Hardies SC, Shehee WR, Comer MB, Edgell MH, Hutchinson CA (1986) The sequence of a large L1Md element reveals a tandemly repeated 5′ end and several features found in retrotransposons. Mol Cell Biol 6:168–182

Mount SM, Rubin GM (1985) Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. Mol Cell Biol 5:1630–1638

Murphy NB, Pays A, Tebabi P, Coquelet H, Guyaux M, Steinert M, Pays E (1987) *Trypanosoma brucei* repeated elements with unusual structural and transcriptional properties. J Mol Biol 195:855–871

Olmsted RA, Hirsch VM, Purcell RH, Johnson PR (1989) Nucleotide sequence analysis of feline immunodeficiency virus: genome organization and relationship to other lentiviruses. Proc Natl Acad Sci USA 86:8088–8092

Ono M, Toh H, Miyata T, Awaya T (1985) Nucleotide sequence of the syrian hamster intracisternal A-particle gene: closing evolutionary relationship of type A particle gene to types B and D oncovirus genes. J Virol 55:387–394

Ono M, Yanusaga T, Miyata T, Ushikubo H (1986) Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. J Virol 60:589–598

Pande S, Lemire EG, Nargang FE (1989) The mitochondrial plasmid from *Neurospora intermedia* strain LaBelle-1b contains a long open reading frame with blocks of amino acids characteristic of reverse transcriptases and related proteins. Nucleic Acid Res 17:2023–2042

Power MD, Marx PA, Bryant ML, Gardner MB, Barr PJ, Luciw PA (1986) Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. Science 231:1567–1572

Priimägi AF, Mizrokhi LJ, Ilyin YV (1988) The *Drosophila* mobile element *jockey* belongs to LINEs and contains coding sequence homologous to some retroviral proteins. Gene 70: 253–262

Querat G, Audoly G, Sonigo P, Vigne R (1990) Nucleotide sequence analysis of SA-OMVV, a visna-related ovine lentivirus: phylogenetic history of lentivirus. Virology 175:434–447

Régnier P, Grunberg-Manago M, Portier V (1987) Nucleotide sequence of the *pnp* gene of *Escherichia coli* encoding polynucleotide phosphorylase: homology of the primary structure of the protein with the RNA-binding domain of ribosomal protein S1. J Biol Chem 262:63–68

Repaske R, Steele PE, O'Neill RR, Rabson AB, Martin MA (1985) Nucleotide sequence of a full-length human endogenous retroviral segment. J Virol 54:764–772

Sagata N, Yasunaga T, Tsuzuku-Kawamura J, Ohishi K, Ogawa Y, Ikawa Y (1985) Complete nucleotide sequence of the genome of bovine leukemia virus: its evolutionary relationship to other retroviruses. Proc Natl Acad Sci USA 82:677–681

Saigo K, Kugimiya W, Matsuo Y, Inouye S, Yoshioka K, Yuki S (1984) Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. Nature 312:659–661

Saltarelli M, Querat G, Konings DAM, Vigne R, Clements JE (1990) Nucleotide sequence and transcriptional analysis of molecular clones of CAEV which generate infectious virus. Virology 179:347–364

Schulte U, Lambowitz AM (1991) The LaBelle mitochondrial plasmid of *Neurospora intermedia* encodes a novel DNA that may be derived from a reverse transcriptase. Mol Cell Biol 11:1696–1706

Schuster W, Brennicke A (1987) Plastid, nuclear and reverse transcriptase sequences in the mitochondrial genome of *Oenothera:* is genetic information transferred between organelles via RNA? The EMBO J 6:2857–2863

Schwartz DE, Tizard R, Gilbert G (1983) Nucleotide sequence of Rous sarcoma virus. Cell 32:853–869

Seiki M, Hattori S, Hirayama Y, Yoshida M (1983) Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc Natl Acad Sci USA 80:3618–3622

Shimotohno K, Takakashi Y, Shimizu N, Gojobori T, Golde DW, Chen ISY, Miwa M, Sugimura T (1985) Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: an open reading frame for the protease gene. Proc Natl Acad Sci USA 82:3101–3105

Shinnick TM, Lerner RA, Sutcliffe JG (1981) Nucleotide sequence of the Moloney murine leukaemia virus. Nature 293: 543–548

Sonigo P, Alizon M, Staskus K, Klatzmann D, Cole S, Danos O, Retzel E, Tiollais P, Haase A, Wain-Hobson S (1985) Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. Cell 42:369–382

Sonigo P, Barker C, Hunter E, Wain-Hobson S (1986) Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D type retrovirus. Cell 45:375–385

Taylor WR (1986) The classification of amino acid conservation. J Theor Biol 119:205–218

Wain-Hobson S, Sonigo P, Danos O, Cole S, Alizon M (1985) Nucleotide sequence of the AIDS virus, LAV. Cell 40:9–17

Wang TSF, Wong SW, Korn D (1989) Human DNA polymerase alpha: predicted functional domains and relationships with viral DNA polymerases. The FASEB Jour 3:14–21

Wong SW, Wahl AF, Yuan MP, Arai N, Pearson BE, Arai K, Korn D, Hunkapiller MW, Wang TSF (1988) Human DNA polymerase alpha gene expression is cell proliferation dependent and its primary structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases. The EMBO Jour 7:37–47

Xiong Y, Eickbush TH (1988) The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. Mol Cell Biol 8:114–123

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. The EMBO Jour 9:3353–3362

Yuki S, Ishimaru S, Inouye S, Saiogo K (1986) Identification of genes for reverse transcriptase-like enzymes in two *Drosophila* retrotransposons *412* and *gypsy;* a rapid detection method of reverse transcriptase genes using YXDD box probes. Nucleic Acid Res 14:3017–3030

Zavriev SK, Borisova OV (1987) Possible functional role of the "DD domain" of RNA-dependent polymerases. Molekuyarnaya Biologiya 21:229–241 (English translation)

Received December 1, 1991/Accepted May 25,1992