



UNIVERSIDAD NACIONAL
AVENIDA DE
MEXICO

227
2ej
Universidad Nacional Autónoma de México

Facultad de Ciencias

Redefinición del dominio Asp-Asp
de reverso transcriptasas

T E S I S

Que para obtener el título de

BIOLOGA

presenta:

Ana María Velasco Velasco

México, D. F. 1992.

TESIS CON
FALLA DE ORIGEN



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE

Resumen.....	1
Introducción.....	2
- Estructura y organización genómica de retrovirus.....	3
- El ciclo biológico de los retrovirus.....	5
- Clasificación de retrovirus.....	8
- La reverse transcriptasa en otras entidades genéticas.....	10
- Diferentes tipos de retroelementos.....	11
- Características moleculares de la reverse transcriptasa.....	12
- El origen y la evolución de la reverse transcriptasa.....	13
- El Dominio DD.....	17
Material y método.....	21
Resultados.....	21
Discusión y conclusiones.....	26
Referencias.....	28

RESUMEN

En la estructura primaria de 15 RNA polimerasas y reverso transcriptasas (RTs) Kamer y Argos (1984), encontraron una región muy conservada, consistente en una secuencia de catorce aminoácidos. Posteriormente, Argos (1988) realizó un estudio más amplio en un conjunto de 57 polimerasas que incluía DNA polimerasas, RNA polimerasas y RTs, localizando en todas el mismo dominio que se reportara con anterioridad. A este motivo se le dió el nombre de "dominio DD" por el hecho de tener en su porción central dos ácidos aspárticos rodeados por aminoácidos en su mayoría hidrofóbicos. Argos (1988) lo caracterizó por un conjunto de reglas que debían cumplir cada una de las posiciones de este dominio. Se cree que el dominio DD juega un papel fundamental en la actividad de las polimerasas, como por ejemplo: es un buen candidato para ser o formar parte del sitio activo de la polimerasa, o ser el sitio de unión a nucleótidos. Al revisar el dominio DD de un conjunto de 53 RTs de origen bacteriano, viral, mitocondrial, de intrones tipo II y de elementos transponibles, encontramos que en especial las RTs encontradas en bacterias, en mitocondrias y en intrones tipo II no cumplían con varias de las reglas propuestas por Argos (1988), lo cual nos llevó a redefinirlas de acuerdo a los nuevos datos. Por otra parte, al calcular los perfiles de hidropatía se encontró que los dominios DD se agrupan en dos conjuntos muy bien definidos y de nuevo se aprecia que en uno de ellos se encuentra la mayoría de las RTs de bacterias, mitocondriales y de intrones tipo II, mientras que en otro se localizan las RTs virales y algunas celulares. Con los resultados obtenidos podemos afirmar que para tener una caracterización más amplia del dominio DD no se deben considerar solamente el conjunto de reglas que debe cumplir cada una de las posiciones de este motivo, sino también tomar en cuenta otros factores como los perfiles de hidropatía.

INTRODUCCION

Hace poco mas de veinte años Baltimore(1970) por una parte, y Temin y Mizutani (1970) por otra, dieron a conocer de manera independiente y simultánea la transferencia de información genética de RNA a DNA como parte del ciclo biológico de ciertos virus de RNA de vertebrados, a los que se les llamó retrovirus. Los retrovirus recibieron este nombre precisamente por el hecho de poder formar una cadena doble de DNA que puede ser incorporada al genoma de la célula infectada, tomando como molde el genoma viral de RNA. A este proceso se le dió el nombre de reverso transcripción y es mediado por una DNA polimerasa RNA dependiente llamada reverso transcriptasa (Baltimore, 1970; Temin y Mizutani, 1970). En los últimos años se han descubierto reverso transcriptasas codificadas por el genoma de eucariontes que se cree juegan un papel relevante en el mantenimiento y evolución del material genético de las células eucariontes. Aunque el interés despertado por la pandemia mundial del SIDA ha aumentado en los últimos años la atención de los científicos en el proceso de reverso transcripción, existen motivos para creer que las RTs han jugado un papel importante en la evolución biológica. Esta posibilidad está apoyada por:

- a) el descubrimiento de la ubicuidad filogenética de RTs incluyendo eubacterias(Lazcano *et al.* 1992)
- b) la posibilidad de que la RT esté involucrada en el transporte intracelular e interespecifico de genes (Weiner *et al.*1986)
- c) las evidencias que apoyan la posibilidad de una época en que las células tuvieron genomas de RNA (Lazcano *et al.* 1992)
- d) el descubrimiento de que la telomerasa eucarionte es una RT (Blackburn, 1990) y
- e) el hecho de que aproximadamente el 10% del material genético de muchas especies eucariontes parece haberse originado por procesos de reverso transcripción (Baltimore, 1985; Temin, 1985).

Años antes que se diera a conocer el fenómeno de reverso transcripción Crick (1958) lo había contemplado ya como uno de los posibles flujos de información genética, proponiendo que estos pueden darse entre los ácidos nucleicos y de ácidos nucleicos a proteínas, pero nunca de proteína a proteína o de proteína hacia ácidos nucleicos. Este descubrimiento era la evidencia que faltaba para ratificar la idea

propuesta por Crick, quien retoma el tema y vuelve a plantear de nuevo los flujos de información en lo que se conoce como el "Dogma Central de la Biología Molecular" (Crick, 1970), (Fig. 1)

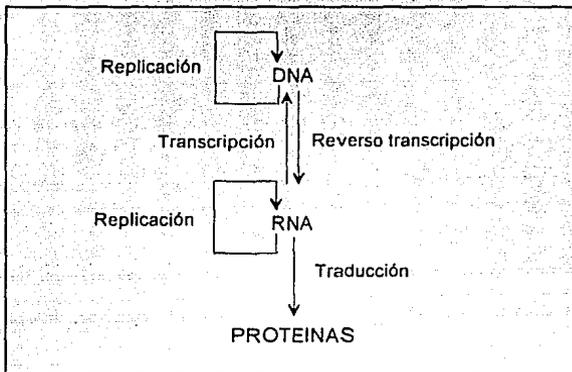


Fig.1 Esquema de los flujos de información genética propuestos por Crick (1970), después del descubrimiento de la reverso transcripción

ESTRUCTURA Y ORGANIZACION GENOMICA DE RETROVIRUS

Los retrovirus son partículas esféricas de 90 a 100 nm de diámetro con una cápside de nucleoproteínas cubierta por una envoltura lipídica. Su material genético lo constituyen 2 hebras de RNA+ iguales, que tienen una longitud de 7 a 9 Kb (Varmus y Brown, 1989). El genoma diploide de todos los retrovirus presenta una estructura consenso formada por tres genes: *gag*, *pol* y *env* (Fig. 2b), aunque pueden existir genes adicionales. Durante la década pasada las proteínas codificadas por estos tres genes se designaron de varias maneras, especialmente con una *p* seguida de un número que indicaba su peso molecular. Ello dió lugar a una confusión considerable, hasta que se propuso una nomenclatura unificada para las proteínas comunes a todos los retrovirus en base a su función biológica, localización en el virión y/o su actividad enzimática (Leis *et al.* 1988). Esta es la nomenclatura que se utiliza en este trabajo y se puede resumir de la siguiente manera.

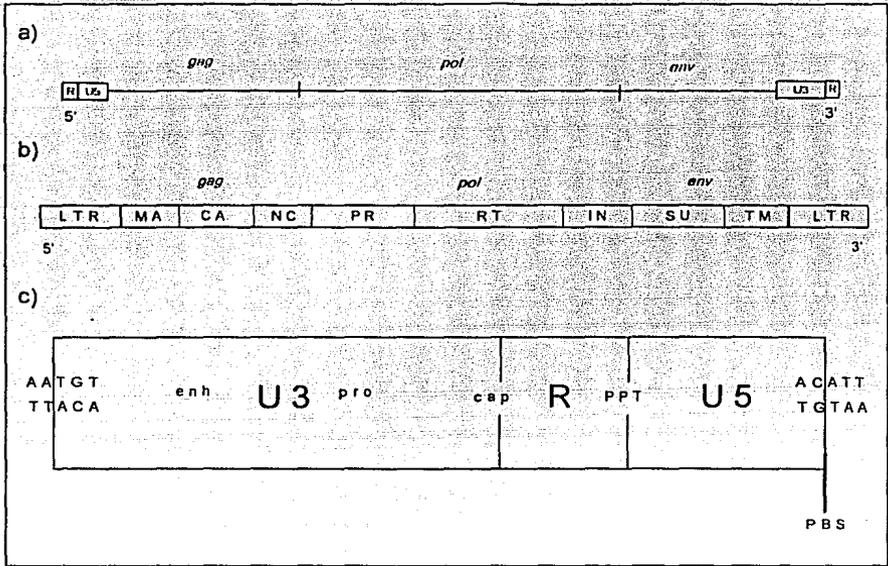


Fig. 2 a) Esquema de la estructura general del genoma de RNA de retrovirus. Tomado de Varmus y Brown (1989); b) forma del provirus resultado de la reversa transcripción del genoma de RNA. Tomado de Doolittle *et al.* (1989); c) estructura y componentes de los LTRs: enhancer (ehn), promotor (pro), sitio de unión al cap (cap), señal de poliadenilación (PPT), sitio de unión al cebador (PBS). Tomado de Temin (1985).

El gen *gag* codifica para el conjunto de proteínas que forman la cápside, que en orden de 5' a 3' son: matriz (MA), una proteína de 15 a 20 kDa, encontrada principalmente en la matriz entre la nucleocápside y la envoltura viral; cápside (CA), es el componente estructural mayor y el antígeno dominante de la cápside viral con un tamaño de 24 a 30 kDa; nucleocápside (NC), una pequeña proteína de 10 a 15 kDa, con una fuerte afinidad por ácidos nucleicos, está fuertemente unida al RNA del virión y genéticamente implicado en el empaquetamiento del RNA y la formación del dímero. El gen *pol* codifica la maquinaria catalítica viral constituida por una proteasa (PR) utilizada en el procesamiento de ciertas poliproteínas producidas después de la transcripción y en la traducción de los genes virales; la reversa transcriptasa (RT), que es la polimerasa viral; al final del gen *pol* se tiene una integrasa o endonucleasa (IN), que juega un papel indispensable en la incorporación del DNA viral en el genoma del hospedero

(Doolittle *et al.* 1989). En *env* se encuentran codificadas las proteínas de la envoltura, que en asociación con componentes de la membrana celular adquiridos por gemación rodean la cápside, lo cual es necesario para la infección (Fuetterer y Hohn, 1987). El producto de este gen es una poliproteína que al ser procesada da como resultado dos proteínas (Fig. 2b). Una de ellas es la de la superficie (SU), con un tamaño que varía de 46 a 120 kDa. Es glicosilada por la maquinaria celular y se encuentra en la superficie del virión interactuando con los receptores del hospedero que median la entrada del virus. Esta proteína es el antígeno reconocido por la mayoría de los anticuerpos y está unida por enlaces disulfuro a la proteína transmembranal (TM), que es el otro producto del gen *env*. El tamaño de esta proteína es de 15 a 37 kDa y no siempre es glicosilada; se cree que media la fusión de la membrana viral con la del hospedero durante la entrada del virus (Varmus y Brown, 1989). El genoma retroviral está flanqueado por dos conjuntos de secuencias repetidas (LTR), en las que se encuentran el promotor, un sitio unión al cebador (PBS) y una señal de poliadenilación (PPT) (Temin, 1985; Cohen y Larsson, 1988) (Fig. 2c).

EL CICLO BIOLÓGICO DE LOS RETROVIRUS

El ciclo biológico de los retrovirus se inicia con la entrada de la partícula viral a una célula susceptible. Se cree que la entrada es por endocitosis, al darse la fusión de la envoltura viral y la membrana celular mediada por el reconocimiento de receptores de la célula (Varmus y Brown, 1989). La mayoría de los elementos necesarios para la reverso transcripción del genoma de retrovirus, como el molde de RNA, el cebador (tRNA celular) y las proteínas catalíticas codificadas por el gen *pol* están presentes en el interior del retrovirus que permanece quiescente hasta que entra a una célula.

Una vez en el interior de la célula, la RT inicia la síntesis de una hebra doble de DNA lineal dentro de la cápside, utilizando como molde el genoma viral de RNA y como cebador un tRNA celular complementario a la región del PBS. El RNA del virus presenta al final de sus dos extremos una secuencia repetida directa (R), la cual en el extremo 5' es seguida por una repetida invertida (U5), mientras que en el 3' es precedida por otra secuencia repetida invertida (U3) (Fig. 2a). Al final de la reverso transcripción las dos secuencias repetidas invertidas se duplican, formando así los LTR (Varmus, 1988). El conjunto de proteínas y DNA migra hacia el núcleo del hospedero, donde el DNA viral se integra al genoma celular. Su

integración presenta varias etapas: los extremos de la molécula se unen formando un círculo covalentemente cerrado que contiene una secuencia de unión de aproximadamente 20pb necesarias para la integración. El círculo se abre cerca de esta secuencia y el par de bases TTAA se pierden en la integración, mientras que unos cuantos nucleótidos del DNA celular son duplicados (Temin, 1985). El genoma viral convertido ahora en provirus se integra como un componente estable del genoma del hospedero y es replicado y transmitido a las células hijas en sincronía con el DNA celular. Eventualmente el provirus es transcrito por la RNA polimerasa II de la célula para ser usado como un mRNA o para ser empaquetado en nuevos virus (Katz y Skalka, 1990). (Fig. 3).

La transcripción y traducción del RNA viral hecha por la maquinaria celular son seguidas por el corte de las poliproteínas producto de los tres genes. De la traducción del gen *env* que es realizada en los ribosomas asociados a retículo endoplásmico (Katz y Skalka, 1990), se obtiene una poliproteína que primero es glicosilada por enzimas celulares y después es procesada proteolíticamente por una enzima del hospedero localizada en el aparato de Golgi durante su transporte a la membrana celular (Orozlan y Luftin, 1990). La traducción de *gag* y *pol* es llevada a cabo en ribosomas libres (Katz y Skalka, 1990; Rhee y Hunter, 1990), siendo su producto poliproteínas que primero se ensamblan con el RNA genómico para formar una estructura en forma de dona que da como resultado partículas inmaduras. Estas son luego encapsuladas por la envoltura viral adquirida mediante un proceso de gemación. Dicha gemación ocurre principalmente en las regiones de la membrana celular que presentan un alto contenido de proteínas SU y TM. Estas partículas formadas por gemación no son infecciosas, para producir virus maduros infecciosos es necesario el corte de las poliproteínas de *gag* y *pol* que es realizado por la PR viral (Orozlan y Luftin, 1990; Katz y Skalka, 1990).

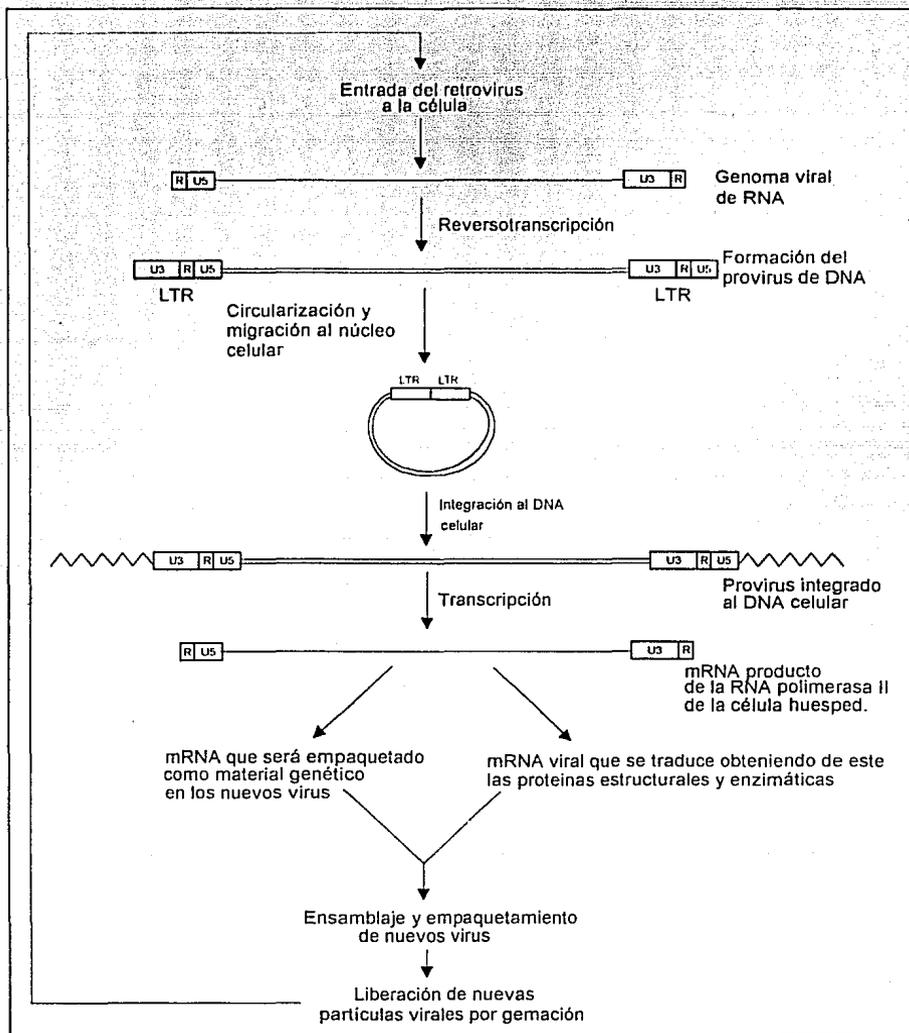


Fig. 3. Ciclo biológico de retrovirus. Modificado de Varmus y Brown (1989); Katz y Skalka (1990).

CLASIFICACION DE RETROVIRUS

Se han hecho varios intentos por clasificar a los retrovirus de acuerdo a distintos intereses. De manera general estos se han dividido en exógenos y endógenos. Los primeros son transmitidos horizontalmente por mecanismos de infección. En cambio, los retrovirus endógenos son llevados de manera benigna en la línea germinal y son heredados a los organismos de la progenie por transmisión vertical mediante mecanismos mendelianos (Temin, 1985). Generalmente los miembros de una especie animal llevan copias múltiples de distintos retrovirus endógenos, de los que muchos son defectivos (Doolittle *et al.* 1990); es decir, carecen de una parte de su genoma, lo cual les impide replicarse. Sin embargo, si en la célula hay otros virus que estén produciendo la parte faltante, los virus defectivos son capaces de utilizar estos elementos y completar su replicación. A los retrovirus que proporcionan los componentes de los que carecen los otros se les conoce como virus auxiliares (Doolittle *et al.* 1989).

Uno de los primeros intentos de clasificación de retrovirus que se hizo está basado en las diferencias morfológicas observadas en microscopía electrónica (Teich, 1982), lo que llevó a los siguientes cuatro grupos:

Partículas tipo A: son formas intracelulares únicamente y no tienen actividad infecciosa. Estas partículas tienen generalmente un diámetro de 60 a 90 nm. Dependiendo de su localización se pueden subdividir en intracisternales o intracitoplásmicas. El papel de las partículas tipo A intracisternales es desconocido, mientras que las intracitoplásmicas aparentemente son formas precursoras del virus del tumor mamario de ratón;

Partículas tipo B: forman la nucleocápside antes de llegar a la membrana celular. Las partículas gemadas presentan una cápside de forma toroidal localizada excentricamente, mientras que la envoltura tiene espículas largas. El prototipo de estos es el virus del tumor de mama de ratón (MMTV);

Partículas tipo C: presentan la estructura más común. Su ensamblaje de la nucleocápside ocurre por debajo de la membrana plasmática en el sitio de la gemación, la forma madura

presenta una cápside central y densa. Las espículas se ven con mucha dificultad. Un ejemplo típico de estos es el virus del sarcoma de Rous (RSV);

Partículas tipo D: están asociadas tanto con formas intracelulares como con extracelulares. Las intracelulares son llamadas partículas tipo D inmaduras, tienen forma de anillo con un diámetro de 60 a 95 nm y generalmente están cerca de la membrana plasmática. La forma extracelular o partícula tipo D maduras tienen un diámetro de 100 a 120 nm y poseen un núcleo denso de electrones localizado excentricamente, las espículas son más pequeñas que las de las partículas tipo B. Uno de los representantes de este tipo de retrovirus y en el cual se basaron para describir las características de este grupo es el virus de mono Mason-Pfizer (MPMV).

Recientemente se propuso un nuevo tipo de retrovirus (Sagata *et al.*1985), designado tipo E, encontrado en el virus de la leucemia de bovinos (BLV) y cuya organización genómica es característica y diferente a la de los otros cuatro grupos. En este se incluirían también a los virus de leucemia de humanos (HTLV). Esta clasificación no es válida por la facilidad con que se cambia de un tipo a otro, así como por no corresponder a las secuencias de RTs. Un ejemplo claro de los problemas que pueden surgir es el caso del virus de mono Mason-Pfizer (MPMV), que pertenece por sus características al grupo de los retrovirus tipo D, pero cuando se da una sustitución de un triptofano por una arginina en la proteína MA del gen *gag*, el virus presenta una morfogénesis como la de los retrovirus tipo C (Rhee y Hunter, 1990).

Los retrovirus se han clasificado también dependiendo del tipo de enfermedad que causan y el hospedero al que afectan con lo que se les divide en:

- 1) virus de leucosis y sarcomas de aves;
- 2) virus de reticuloendoteliosis;
- 3) virus de sarcomas y leucemia de mamíferos;
- 4) virus que causan tumores de mama;
- 5) virus de primates tipo D;
- 6) virus relacionados con leucemias de células T en humanos, y
- 7) virus de inmunodeficiencia y lentivirus (Varmus y Brown, 1989).

Existe otro trabajo de taxonomía que los agrupa en agentes que causan inducción de tumores (*Oncoviridae*), agentes que causan enfermedades lentas y neoplasias (*Lentiviridae*) y aquellos que inducen la vacuolización espumosa de las células infectadas (*Spumiviridae*) (Teich, 1982). Es importante hacer notar que además de todos estos tipos de clasificación actualmente se trata de ordenar a los retrovirus de acuerdo a la estructura y organización de su material genético ya que ha habido problemas con los métodos anteriores.

LAS REVERSO TRANSCRIPTASAS EN OTRAS ENTIDADES GENÉTICAS

Como se mencionó anteriormente, el proceso de la reversotranscripción fue originalmente descubierto como una fase esencial en el ciclo biológico de retrovirus y durante más de 10 años se pensó que tal fenómeno era privativo de este tipo de replicación (Baltimore, 1985). Investigaciones posteriores del genoma de diversos organismos permitieron el descubrimiento de otros elementos genéticos tanto celulares como virales que codifican para RTs y que además tenían otras características presentes en retrovirus (Varmus, 1983).

Todos los elementos genéticos que codifican para RT son llamados retroides (Fuetterer y Hohn, 1987) y la comparación de secuencias de aminoácidos ha demostrado que están relacionados con virus tanto exógenos como endógenos (Doolittle *et al.* 1989; Xiong y Eickbush, 1990). Recientemente Temin (1989) propuso el término "retroelemento" para designar a todas las entidades genéticas que codifican para una RT o que se originaron por reverso transcripción. Entre los primeros elementos genéticos que se supo codificaban para RT se encuentran los virus de la hepatitis B de humano y el de mosaico de la coliflor (siendo ambos virus de DNA), así como también los elementos encontrados en levaduras, insectos y vertebrados. Los retroides están ampliamente distribuidos entre los seres vivos ya que se han encontrado en organismos representantes de todos los reinos biológicos (Tabla 1).

Tabla 1. Ejemplos de retroides encontrados en organismos de los 5 reinos biológicos

REINO	RETROIDE	ORGANISMO
Monera	Ec67	<i>Escherichia coli</i> (Lampson et al. 1989)
	Mx162	<i>Mixococcus xanthus</i> (Inouye et al. 1989)
Protista	INGI-3	<i>Trypanosoma brucei</i> (Kimmel et al. 1987)
	RTL-Cr Pt1	<i>Chlamydomonas reinhardtii</i> (Boer y Gray, 1988)A Alga verde (Kück, 1989)
Fungi	IA-Pa	<i>Podospora anserina</i> (Cummings et al. 1989)
	Ty1	<i>Saccharomyces cerevisiae</i> (Clare y Farabaugh, 1985)
	ai-Nc	<i>Neurospora crassa</i> (Nargang et al. 1984)
Animal	T1	<i>Anopheles gambiae</i> (Besansky, 1990)
	L1Md	Ratón (Loeb et al. 1986)
	Tx1	<i>Xenopus laevis</i> (Garret et al. 1989)
	microplia	<i>Drosophila hydei</i> (Huisjer et al. 1988)
Vegetal	mOe	<i>Oenothera</i> (Schuster y Brennicke, 1987)
	Bs1	<i>Ea mayz</i> (Jin y Bennetzen, 1989)
	Tai	<i>Arabidopsis thaliana</i> (Voytas y Ausubel, 1988)

DIFERENTES TIPOS DE RETROELEMENTOS

La estructura y distribución de secuencias codificadoras del genoma de los retroelementos no es constante, ya que mientras unos presentan la misma organización que la de los retrovirus, otros solo poseen un marco único de lectura para una RT. Por ello Temin (1989), propuso ordenar a los retroelementos con relación a su organización genómica (Tabla 2) en los siguientes grupos:

- a) Retrones: son los retroelementos más sencillos, su único marco de lectura presenta homología con la RT, no tiene LTR, ni la capacidad de transponerse ni de formar partículas virales
- .b) Retroposones: tienen una secuencia que codifica para RT, pueden transponerse pero carecen de LTR y no forman víriones.
- c) Retrotransposones: presentan marco de lectura con homología a la RT, tienen LTRs, pueden transponerse, pero no forman víriones .
- d) Retrovirus: son virus de RNA, codifican para una RT, presentan LTRs, son capaces de transponerse y forman víriones.

- e) Pararetrovirus: son virus de DNA de plantas o animales que codifican para una RT, tienen LTR y forman virus, pero no se transponen .
- f) Retrosecuencias: son secuencias de DNA que no tienen RT, ni LTRs, no se transponen ni forman partículas virales. Se cree que son el resultado de la reverso transcripción de un mRNA en base a su similitud con secuencias de genes que codifican para proteínas. En el extremo 3' tienen una cola de poli (A), carecen de intrones y algunos de ellos presentan secuencias repetidas directas bordeandolas, que son características de los eventos de transposición del DNA.

Tabla 2. Características de los diferentes tipos de retroelementos y ejemplos de estos*

RETROELEMENTO	RT	TRANSPOSICION/ INTEGRACION	LTR	FORMACION DE VIRIONES	EJEMPLO
Retron	+	0	0	0	EcE86 de <i>E. coli</i>
Retroposon	+	+	0	0	RIBm de <i>Bombyx mori</i>
Retrotransposon	+	+	+	0	412 de <i>Drosophila melanogaster</i>
Retrovirus	+	+	+	+	BLV virus de la leucemia de bovino
Pararetrovirus	+	0	+	+	CaMV virus del mosaico de la coliflor

*El signo + Indica que presenta esa característica, el 0 que carece de ella. Modificado de Temin (1989).

CARACTERISTICAS MOLECULARES DE LA REVERSO TRANSCRIPTASA

La RT es una proteína de 50 a 95 kDa (Varmus y Brown, 1989) con actividad de DNA polimerasa RNA o DNA dependiente en el extremo NH₃, mientras que en el COOH⁻ presenta actividad de ribonucleasa H (Johnson *et al.* 1986; Tanese y Goff, 1988). La RNasa H de retrovirus y retroides en general parece actuar como una exonucleasa con especificidad, ya sea de 3' a 5' o de 5' a 3' (Varmus, 1983) y es la encargada de degradar el RNA cuando se forma el híbrido de RNA y DNA en la primera etapa de la reverso transcripción (Bishop, 1978). Dependiendo del organismo o virus donde se encuentre puede presentarse en forma monomérica o dimérica. Así, la RT del virus del sarcoma de Rous (RSV-RT) es un heterodímero formado por una subunidad a de 63 kDa y una b de 95 kDa (Schwartz *et al.* 1983), mientras que el virus de la leucemia de murino (Mo-MLV) tiene una RT monomérica de 80 kDa

(Shinnick *et al.* 1981). Al igual otras DNA polimerasas DNA dependientes necesita de un cebador cuando se sintetiza la hebra de DNA complementaria al RNA+ (Varmus, 1983), que en este caso es un tRNA de la célula del hospedero. Para iniciar la síntesis de la segunda hebra la polimerasa utiliza un oligorribonucleotido rico en purinas generado por la RNasa H al degradar el RNA del heteroduplex (Varmus y Brown, 1989). A diferencia de las DNA polimerasas no presenta actividad correctora, por lo que su tasa de error durante la replicación es más alta que la de las DNA polymerases, las que tienen una tasa de error de 10^6 a 10^{11} por nucleótido incorporado, mientras que las RNA polimerasas presentan una tasa de error de 10^3 a 10^4 (Holland *et al.* 1982).

EL ORIGEN Y LA EVOLUCION DE LA REVERSO TRANSCRIPTASA

Para poder explicar el origen de la reverso transcripción es necesario revisar la teoría que se tiene actualmente para entender las características de las primeras células. Varios autores (Alberts, 1986; Gilbert, 1986; Lazcano, 1986), han propuesto que los primeros organismos eran sistemas que dependían del RNA, a partir de los cuales evolucionaron células con RNA y proteínas (Lazcano *et al.* 1992a). Es decir, hubo una etapa temprana en la evolución biológica durante la cual los genomas celulares eran de RNA y no de DNA. La idea de que el RNA apareció antes que el DNA como molécula almacenadora de información genética está basada en las siguientes evidencias circunstanciales que se tienen hasta el momento (Lazcano *et al.* 1988a) y que se mencionan a continuación:

- 1) La biosíntesis de proteínas puede llevarse a cabo en ausencia de DNA pero no de RNA.
- 2) La existencia de viroides y virus de RNA indican que los sistemas replicativos pueden usar ya sea polirribonucleótidos de hebra doble o sencilla para almacenar información genética.
- 3) Las moléculas de RNA tienen propiedades catalíticas que incluyen la de endonucleasa, RNA ligasa, exonucleasa y RNA polimerasa, la de esterasa y la de formación del enlace peptídico.
- 4) Los cofactores ribonucleotídicos se encuentran ampliamente distribuidos en una gran cantidad de rutas metabólicas.
- 5) El grupo 2'-OH de la ribosa puede estar directamente involucrado en un gran número de reacciones de fosforilación y condensación de aminoácidos.

6) La biosíntesis de desoxirribonucleótidos siempre ocurre mediante la reducción enzimática de ribonucleótidos.

Se ha postulado que éstas células con genomas de RNA se vieron sujetas a fuertes presiones de selección, siendo el resultado de ellas la aparición del DNA, probablemente debido a que este tipo de polinucleótidos incrementa la estabilidad del material genético (Lazcano, 1986; Lazcano *et al.* 1988 a; Wintersberger y Wintersberger, 1987). De esta manera, la aparición del DNA es el resultado de la evolución biológica de las células del Arqueano, en las que la información genética almacenada en un principio en RNA cambió por un proceso de selección natural a DNA (Lazcano *et al.* 1992a). Las ventajas que llevaron al DNA a tomar el lugar del RNA han sido resumidas (Lazcano *et al.* 1992a) en :

- a) el incremento en la estabilidad ya que el enlace fosfodiéster es más resistente a la hidrólisis que su equivalente ribo, pues carece del OH-2' de la ribosa;
- b) la aparición de mecanismos de reparación, pues la DNA polimerasa presenta actividad correctora así como de exonucleasa 3'-5', características de las que carecen las RNA polimerasas, por lo que su replicación es menos fiel que la del DNA;
- c) la deaminación de citosina en uracilo tanto en RNA como en DNA es considerablemente más rápida que otras reacciones de deaminación sufridas por purinas y pirimidinas y como los sistemas basados en RNA no tienen la maquinaria necesaria para detectar estos cambios de citosina a uracilo la información del RNA tiende a degradarse; y por último
- d) debido a la estructura de doble hebra del DNA, este es menos susceptible al daño por radiación UV, mientras que los genomas de RNA probablemente solo tenían pequeñas regiones complementarias entre sí que formaban estructuras de hebra doble, mientras que la gran mayoría se presentaba en forma de una hebra, que estaban expuestas a un alto índice de lesiones por acción fotoquímica de la radiación UV (Lazcano *et al.* 1988a).

Como se mencionó arriba, los primeros sistemas vivos estaban basados en RNA, por lo que se cree que una de las primeras proteínas que debió surgir fue una RNA polimerasa que leía genomas de RNA y los replicaba (Lazcano *et al.* 1988b). Por lo tanto, durante la transición hacia genomas celulares de DNA debe haber habido una enzima con actividad de RT.

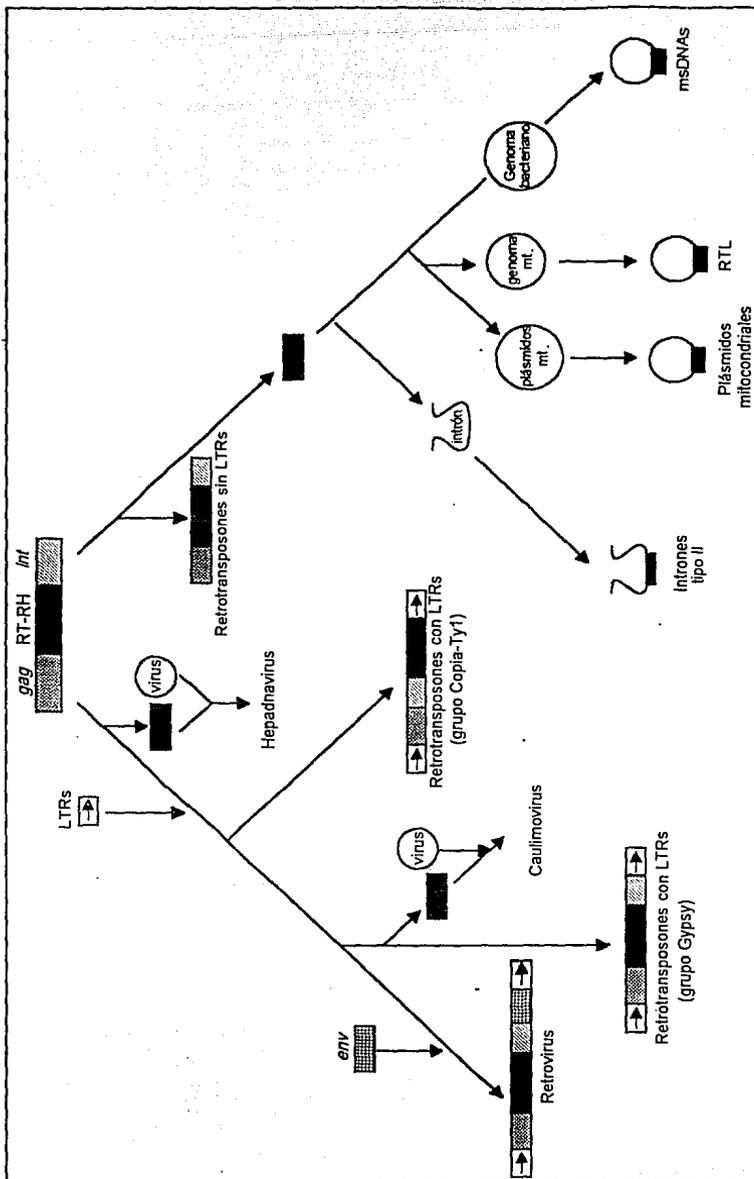


Fig. 4 Representación esquemática del posible origen de los diferentes tipos de retroideos que se conocen. Tomado de Xiong y Eickbush, 1990.

¿ Fue esta la enzima ancestral a la que presentan los retroides (Weiner y Maizels, 1991), o las primeras polimerasas eran poco específicas como sugieren Lazcano *et al.* (1992b, c)? La RNA replicasa ancestral pudo haber tomado desoxirribonucleótidos e incorporarlos durante la replicación del RNA debido a su baja especificidad (Lazcano *et al.* 1992c) y evolucionar luego cambiando su preferencia por el molde y entonces leer DNA. Ello es fácil de aceptar, ya que las polimerasas actuales requieren de modificaciones menores en el medio en el que se encuentra para cambiar su especificidad por el molde. Así, la actividad que presentaría esta polimerasa ancestral sería la de RT, leyendo moldes de RNA y formando la hebra complementaria de DNA.

Debido a su amplia distribución en una gran variedad de organismos, se cree que la aparición de la RT es mucho más antigua que la división entre procariontes y eucariontes (Doolittle *et al.* 1989; Temin, 1989) y que a lo largo de su historia evolutiva se ha visto relacionada con otras estructuras genéticas, dando origen a los diferentes tipos de retroides (Temin, 1989; Xiong y Eickbush, 1990). Siguiendo el esquema de aparición de los retroides propuesto por Temin (1989), aunque la RT se conservó evolutivamente, primero fue un retron, del cual surgió un retroposon con la habilidad de transponerse. Posteriormente aparecieron los retrotransposones que ya presentaban LTRs y a partir de estos se formarían los retrovirus que portaban el gen *env*, mientras que por otro lado surgían los pararetrovirus que perdieron su capacidad de integrarse y transponerse (Fig. 4).

A pesar de las altas tasas de mutación (Holland *et al.* 1982), el análisis de la estructura primaria de varias RTs permitió concluir que todas compartían un ancestro común, pues en la comparación de secuencias se obtuvieron elevados índices de similitud (Fuetterer y Hohn, 1987; Doolittle *et al.* 1989; Xiong y Eickbush, 1988a, 1990). Debido a su conservación las RTs han sido utilizadas por varios grupos para dilucidar las relaciones evolutivas existentes entre los retroides que se conocen en base a la comparación de sus secuencias (Chiu *et al.* 1985; Garvey *et al.* 1990; Gojobori *et al.* 1990; Ono *et al.* 1986; Sagata *et al.* 1985) y a partir de estos datos se obtienen árboles filogenéticos. Un buen ejemplo es el árbol obtenido por Xiong y Eickbush (1988a, 1990) en el que se aprecia claramente que los retroides están agrupados también de acuerdo a su estructura genómica. En este árbol existe una rama en la que se encuentran todos los retones, en otra los intrones tipo II y los retrotransposones con LTRs y sin LTRs

también están separados en diferentes ramas al igual que ocurre con los retrovirus, los hepadnavirus y los caulimovirus (Fig. 5).

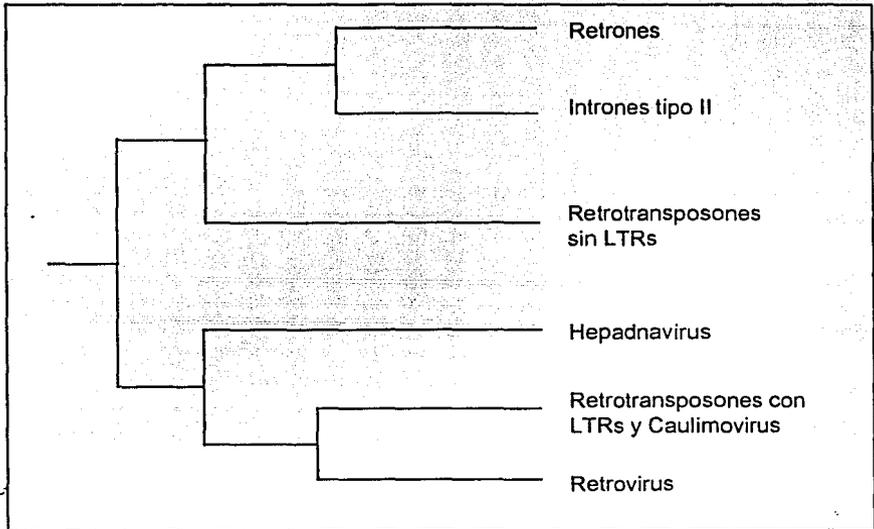


Fig. 5. Relaciones filogenéticas propuestas para los diferentes grupos de retroideos. Modificado de Xiong y Eickbush (1990).

EL DOMINIO DD

A raíz de la comparación de un conjunto de secuencias formado por 10 RNA polimerasa celulares y virales y 5 RTs, Kamer y Argos (1984) detectaron una región altamente conservada. Un estudio posterior mucho más amplio llevado a cabo por Argos (1988), en un conjunto de polimerasas que incluían a 11 DNA polimerasa DNA dependientes, 20 RTs y 26 RNA polimerasa RNA dependientes, reveló que consistía de 15 aminoácidos en las DNA polimerasas y de 14 en las RNA polimerasas y RTs, permitiéndole confirmar los resultados anteriores ya que encontró en todas las polimerasas la secuencia que se había reportado anteriormente. A este motivo se le llamó "Dominio DD" por el hecho invariable de

que en RNA polimerasas y RTs se presentan dos ácidos aspárticos (Asp-Asp) en la parte central de esta secuencia rodeados por 12 aminoácidos hidrofóbicos en su mayoría, para las RTs y las RNA polimerasas. En cambio, en las DNA polimerasas se encontró una región semejante en la que entre los dos aspárticos se encontraba una treonina y después del segundo ácido aspártico había una serina (D-T-D-S) (Argos, 1988; Bernad *et al.* 1980b).

Debido a que la distribución de los diferentes aminoácidos en este motivo presentaba características muy definidas Argos (1988), propuso una serie de reglas para caracterizar este dominio :

- Regla 1 : las posiciones 8 y 9 deben estar ocupadas por Asp.
- Regla 2 : en la posición 7 se puede encontrar Gly, Met, Cys, Val o Leu
- Regla 3 : la posición 6 puede estar ocupada por Tyr, Ala, Phe, Ser, Asn, Cys, Gly, Ile o Met.
- Regla 4 : por lo menos dos residuos de la posición 1 a 5 y de la 10 a 14 deben ser hidrofóbicos (Ala, Val, Leu, Ile, Cys, Met, Phe, Tyr, His, Trp, Pro)
- Regla 5 : si solamente dos residuos de las posiciones 1 a 5 son hidrofóbicos, entonces, debe haber una Ser o una Gly. Si solamente dos residuos de la posición 10 a 14 son hidrofóbicos, entonces debe haber una Ser
- Regla 6 : en la posición 4 no debe haber Lys, Arg, Asp, Glu, Gln o Asn
- Regla 7 : las posiciones 11 y 12 deben estar ocupadas por aminoácidos hidrofóbicos.

Ya que este motivo está altamente conservado en todas las polimerasas, se cree que juega un papel importante en la actividad de la enzima permaneciendo a lo largo de su trayectoria evolutiva (Kamer y Argos, 1984; Argos, 1988; Lazcano *et al.* 1988b). Sin embargo debido a su tamaño tan reducido, no se puede descartar que tenga un origen polifilético. No se conoce su función precisa y se le han atribuido diferentes funciones, como por ejemplo la de unión al molde (Larder *et al.* 1987), de formar parte del sitio activo de la polimerasa (Kamer y Argos, 1984), o de unión al sustrato (Lazcano *et al.* 1992c)

Aunque no se conoce el papel que juega el dominio DD, su importancia funcional y/o estructural ha quedado demostrada por experimentos de mutagénesis dirigida, en los que se encontró que al cambiar ciertos aminoácidos de este motivo la polimerasa perdía su actividad. En uno de estos experimentos Larder *et al.* (1987), sustituyeron un total de 8 aminoácidos a lo largo de la RT de HIV, encontrando que

solo dos mutaciones daban lugar a la perdida total de actividad de la enzima. Las mutaciones que inactivan a la polimerasa eran las que cambiaban el aspártico totalmente conservado de la secuencia que Delarue *et al.* (1990), denominan Motivo A, así como el primer ácido aspártico del doblete invariamente conservado en el dominio DD. En cambio al cambiar la tirosina que ocupa la posición seis y la glicina en la posición trece del dominio DD, solo se presentó una actividad reducida pero sin llegar a perderla (Larder *et al.* 1987).

Además del dominio DD se han encontrado otras secuencias altamente conservadas en las RTs, que parecen estar relacionadas con la actividad de la polimerasas y con el dominio DD, que forman parte del llamado "módulo ancestral de las polimerasas" (Poch *et al.* 1989). Por ejemplo, Poch *et al.* (1989) detectaron la existencia de otros cuatro motivos, que junto con el dominio DD dan un total de 5 motivos conservados, a los que se designó en orden de NH₂ a COOH como Motivo A, B, C (Dominio DD), D y E. Los primeros cuatro poseen equivalentes en RNA polimerasas RNA dependientes y según Delarue *et al.* (1990) forman un dominio mucho mayor o "módulo" de gran importancia, ya que corresponden al sitio de reconocimiento del molde y/o el sitio activo de la polimerasa. Por otra parte, Delarue *et al.* (1990), encontraron tres motivos conservados en DNA polimerasas DNA dependientes, que del extremo amino al carboxilo son el A, B y C (Dominio D-T-D-S), observando similitud de A y C con dos regiones de RTs. El motivo A tiene un ácido aspártico conservado en todas las secuencias analizadas al igual que el segundo aspártico del doblete del dominio DD. Es posible que el ácido aspártico del motivo A coopere con el segundo del dominio DD para unirse a iones Mg que pueden ser parte del sitio de union de dNTP (Poch *et al.* 1989; Delarue *et al.* 1990).

En años recientes el interes en la evolución de retrovirus y entidades genéticas relacionadas se despertó por la epidemia causada por el SIDA, lo que llevó a un incremento en las secuencias disponibles de RTs de diferentes fuentes, incluyendo procariontes, virus, elementos transponibles, plásmidos mitocondriales (mt) e intrones. El análisis del dominio DD de estas secuencias adicionales de RTs muestran que muchas de ellas no siguen completamente las reglas mencionadas anteriormente. En este trabajo algunas reglas de Argos son redefinidas, para tener así una caracterización mucho más precisa del dominio DD presente en RTs.

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. Oro²

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, Texas 77204-5500

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 LIVYSD-DYLSLM (Zavriev and Borisova 1987); (2) site-specific mutagenesis of the HIV-1 RT (Lärder et al. 1987; Hizi et al. 1988); and (3) amino acid substitutions involving the Q β 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 Mg²⁺ (Argos 1988), Zn²⁺, 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. 1991). Similar conclusions have been reached for the ϕ 29 DNA polymerase YGDTDS motif (Bernard et al. 1990; Blanco et al. 1991).

Offprint requests to: A. Lazcano

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

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 defined 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 indicate at least one of Argos's rules. Hydrophathy profiles were plotted (Fig. 1) using the hydrophobicity 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 hydrophathy 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 hydrophathy 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 hydrophathy 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 *Podospira* 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 Bsl, 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*

Source	DD domain	Reference	
<i>Escherichia coli</i>	Ec67	CTYSRYADDITIST	Lampson et al. 1989
<i>E. coli</i>	EcB86	LIYTRYADDLTLSA	Lim & Mnas 1989
<i>Myxococcus xanthus</i>	Mx162	FTYTRYADDLTFSW	Inouye et al. 1989
<i>M. xanthus</i>	Mx65	YTYTRYADDLTFSG	Inouye et al. 1990
<i>Trypanosoma brucei</i>	ING1-3	LQHGFFADDLTLA	Kimmel et al. 1987
<i>T. brucei</i>	TRS	LQHGFFADDLTLFS	Murphy et al. 1987
<i>Chlamydomonas reinhardtii</i>	RTL-Cr	MDFTIYADNFAGVF	Boer & Gray, 1988
<i>Green alga (pt)</i>	Pl1	NTYCRYADDMVILT	Köck 1989
<i>Dietystellium discoideum</i>	DIRS-1	VSVIAYLDDLLIVG	Cappello et al. 1985
Yeast	Ty1	VTICLFVDDMVLFS	Clare & Faraugh 1985
<i>Saccharomyces cerevisiae</i> (mt)	a1-Sc	IKYVRYADDILIOV	Bonitz et al. 1980
<i>S. cerevisiae</i> (mt)	a2-Sc	AYFVRYADDILIVG	
<i>Schizosaccharomyces pombe</i> (mt)	B1-Sp	LMYVRYADDWIVAV	Lang & Ahne 1985
<i>Podospira anserina</i> (mt)	a1-Pa	IYYVRYADDWLVIG	Cummings et al. 1989
<i>P. anserina</i> (mt)	IA-Pa	VRYTRYADDWVIGI	
<i>Neurospora intermedia</i> (mt)	LaBelle-1b	AVGSIYADEGKVKV	Pande et al. 1989
<i>Bombyx mori</i>	R1Bm	TEMVAYADDVTVLV	Xiong & Eickbush 1988
<i>B. mori</i>	R2Bm	VSALAYADDLVLLA	Burke et al. 1987
<i>Anopheles gambiae</i>	T1	DGHLLYADDIKIFL	Besansky 1990
<i>Drosophila melanogaster</i>	Copia	IYVLLYVDDVVIAT	Mount & Rubin 1985
<i>D. melanogaster</i>	jockey	VL IATYADDTAVLT	Primagi et al. 1988
<i>D. melanogaster</i>	F Factor	LTVSTFADDTAILS	Di Nocera & Casari 1987
<i>D. melanogaster</i>	I Factor	IKFNAYDDDFLLI	Fawcett et al. 1986
<i>D. melanogaster</i>	17.6	KHCLVYDDIIVFS	Salgo et al. 1984
<i>D. melanogaster</i>	297	KHCLVYDDIIVFS	Inouye et al. 1986
<i>D. melanogaster</i>	412	SQAFLYDDLVIVIG	Yuki et al. 1986
<i>D. melanogaster</i>	GYPSY	KICYVYVDDVIVFS	
<i>D. melanogaster</i>	1731	MLILVYVDDLVLLAC	Fourcade-Peronnet et al. 1988
Mouse	LIMd	VKISLLADDIVYI	Loeb et al. 1986
<i>Oenothera</i> (mt)	mtOe	HFVVYVDDLVVYT	Fenster & Brennicke 1987
<i>Zea mays</i>	Bs1	VRVRFVDDGGVVE	Jin & Bennetzen 1989
Syrian hamster intracisternal A-Particle	IAP-H18	LIVIHVMDDILICH	Ono et al. 1985
Rous sarcoma virus	RSV	LCMLHYMDDL LLA	Schwartz et al. 1983
Equine infectious anemia virus	EIAV	VQLYQYMMDLFVGS	Kawakami et al. 1987
Caprine arthritis encephalitis virus	CAEV	IQFGIYMDDIYIGS	Saltarelli et al. 1990
Feline immunodeficiency virus	FIV	LDIYQYMMDDIYIGS	Olmsted et al. 1989
Molbney murine leukemia virus	MoMLV	LILLQYVDDLL LAA	Shinnick et al. 1981
Bovine immunodeficiency-like virus	BIV	YMLYQYMMDDL LIGS	Garvey et al. 1990
Bovine leukemia virus	BLV	SLLVSYMDDI L YAS	Sagata et al. 1985
Visna	VISNA	IQFGIYMDDIYIGS	Sonigo et al. 1985
Visna-related ovine lentivirus	SA-OMVV	IQFGIYMDDIYIGS	Querat et al. 1990
Simian retrovirus	SRV-1	MYI IHVMDDI L IAG	Power et al. 1986
Simian T-cell lymphotropic virus type III	SIVK6W	VTLVQYMMDDI L IAS	Franchini et al. 1987
Mazon-Pfizer monkey virus	MPMV	MTI IHVMDDI L IAG	Sonigo et al. 1986
Baboon endogenous retrovirus	BaEV	VTLLQYVDDLL LAA	Kato et al. 1987
Human T-cell leukemia virus type 1	HTLV-I	CTILQYMMDDI L LAA	Seiki et al. 1983
Human T-cell leukemia virus type 2	HTLV-II	STIVQYMMDDI L LAA	Shimotohno et al. 1985
Human immunodeficiency virus type 1	HIV-1	IVIYQYMMDDLVYGS	Wain-Hobson et al. 1985
Human immunodeficiency virus type 2	HIV-2	VIIYQYMMDDI L IAS	Guyader et al. 1987
Human endogenous retrovirus K	HERV-K	CYI IHYIDDL LCA	Ono et al. 1986
Human endogenous retrovirus C	HERV-C	CVLLQYVDDLL L LGG	Repaske et al. 1985
Cauliflower mosaic virus	CaMV	KFCVYVDDI L VFS	Franck et al. 1980
Carnation etched ring virus	CERV	KYCCVYVDDI L VFS	Hull et al. 1986

* 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

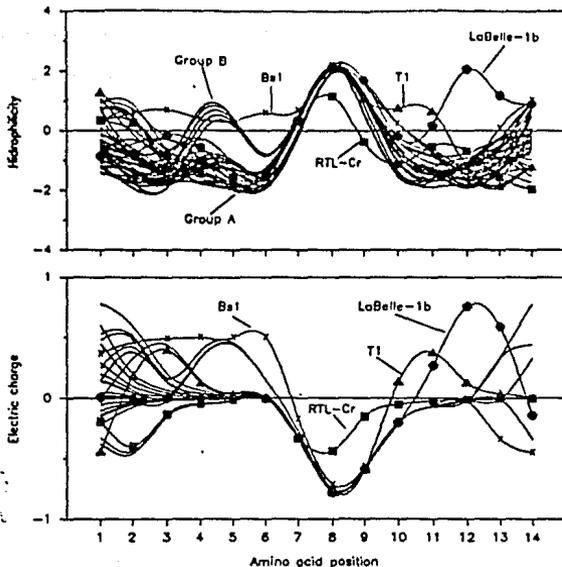


Fig. 1. Smoothed hydropathy and electric charge plots of the RT DD domains of Table I

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. 1991). However, both its presence in functional polymerases and its practical significance in the theoretical (Argos 1988) and experimental determination of polymerases: (Yuk: 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 hydrophobic 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.

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.I.) and NASA grant 44-005-002 (J.O.).

References

- Argos A (1988) A sequence motif in many polymerases. *Nucleic Acid Res* 16:9909-9916
- Bernard 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 ϕ 29 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 enzyme. *Mol Cell Biol* 7:2221-2230
- Cappello J, Handelsman K, Lodish HF (1985) Sequence of *Dicyostelium* 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 *Fodospora 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 factors, and mammalian LI sequences. *Proc Natl Acad Sci USA* 84:5843-5847
- Fawcett DH, Lister CK, Kelleit E, Finnegan DJ (1986) Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINEs. *Cell* 47:1007-1015
- Fourcade-Peronnet F, d'Auriol B, Becker J, Galibert F, Best-Belpomme M (1988) Primary structure and functional organization of *Drosophila* 1731 retrotransposon. *Nucleic Acid Res* 16:6113-6123
- 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, Guillely 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, Hirasima 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 *Mycococcus xanthus*. *Cell* 56:709-717
- Inouye S, Herzer PJ, Inouye M (1990) Two independent retrons with highly diverse reverse transcriptase in *Mycococcus xanthus*. *Proc Natl Acad Sci USA* 87:942-945
- Jin YK, Benezien JL (1989) Structure and coding properties of *Bt1*, 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 retroviruses. *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
- Köck U (1989) The intron of a plastid gene from a green alga 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, Ramirez C, Ord J (1988) On the early evolution of RNA polymerase. *J Mol Evol* 27:365-376
- Lazcano A, Valverde V, Greco-Hernández, Gariglio P, Fox GE, Ord J (1991) On the early emergence of reverse transcription: theoretical and experimental evidence. *J Mol Evol* (accepted for publication)
- 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, Sheehy WR, Comer MB, Edgell MH, Hutchinson CA (1986) The sequence of a large L1/Mid 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, Coquelle 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, Awata 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 *jackey* 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
- Sugata 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
- Sallarelli 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 (1971) 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, Brunnicke 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* 6:2837-2863
- Schwartz DE, Tizard R, Gilbert G (1983) Nucleotide sequence of Kous 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
- Shintohno K, Takahashi Y, Shimizu N, Gobjobori 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 leukemia 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-Dixon 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 J* 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 J* 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 J* 9:3353-3362
- Yuki S, Ishimaru S, Inouye S, Saigo 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 bnx probes. *Nucleic Acid Res* 14:3017-3030
- Zavriev SK, Borisova OV (1987) Possible functional role of the "DD domain" of RNA-dependent polymerases. *Molekulyarnaya Biologiya* 21:229-241 (English translation)

Received December 1, 1991/Accepted May 25, 1992

DISCUSION Y CONCLUSIONES

Analizando los resultados que se presentan en la tabla 1 podemos ver que la mayoría de las RTs encontradas en bacterias, intrones tipo II, mitocondrias y retrotransposones, no cumplen por lo menos con alguna de las reglas de Argos (1988). Ello se podría explicar suponiendo que estas enzimas no juegan un papel fundamental en el ciclo biológico de dichos elementos genéticos, sino que mas bien dependen de la polimerasa celular para replicarse, pues lo hacen en sincronía con la célula. Es decir, se puede pensar que se encontrarían bajo presiones de selección más relajadas que las RTs virales, en donde su mantenimiento es necesario para completar el ciclo biológico de los retrovirus.

Es importante hacer notar que el aminoácido que se encuentra completamente conservado en este conjunto de secuencias es el primer ácido aspártico del doblete Asp-Asp, y en cambio el segundo se presenta sustituido en dos ocasiones, en RTL-Cr que es parte del DNA mitocondrial de *Chlamydomonas reinhardtii* se encuentra una asparagina y en el plásmido mitocondrial de *Neurospora intermedia* denominado Labelle-1b se tiene un ácido glutámico. Ello viene a reforzar la idea de que el primer residuo de ácido aspártico es probablemente uno de los más importantes para mantener la actividad de la RT, como lo sugieren los experimentos hechos por Larder *et al.* (1987). Surge así la pregunta de cual es realmente el papel funcional de este dominio y cual sería su relación con el dominio DD de RNA y DNA polimerasas, pues presenta características que comparte con ambas a la vez que existen también diferencias entre el dominio de RTs y RNA polimerasas con el de DNA polimerasas. Por ejemplo mientras que en RTs y RNA polimerasas el dominio DD se encuentra en la región amino terminal de la proteína, en las DNA polimerasas está localizado hacia el extremo carboxilo (Bernad *et al.* 1990a, b). Otra diferencia se observa al cambiar el primer aspártico de las RTs, estas pierden su actividad por completo. Ello no ocurre sin embargo al sustituir el primer residuo de ácido aspártico en DNA polimerasas (DTDS), pero si se sustituye la treonina o el segundo aspártico, entonces la polimerasa se inactiva (Bernad *et al.* 1990a). Es importante hacer un estudio en el cual se puedan comparar los dominios de las diferentes polimerasas y tratar entender las diferencias que presentan, así como encontrar cual es su relación entre sí ya que existen casos notables como el de LaBelle que presentando un dominio que podría decirse es de una RT actúa como una DNA polimerasa (Schulte y Lambowitz, 1991), pero carece del dominio DD encontrado en DNA polimerasas.

Hasta el momento se le han atribuido diferentes funciones a este dominio, pero queda todavía la duda de cual es realmente su papel en el desempeño de la actividad enzimática. Una cosa es clara de todo esto y es que cuando se han realizado experimentos de mutagénesis dirigida que afectan a este sitio la actividad de la proteína se ve disminuida o inhibida por completo, lo que indica la importancia de este dominio dentro de la polimerasa. Recientemente se encontró que la posible RNA polimerasa del virus de necrosis pancreática infecciosa de salmones no tiene el dominio DD (Duncan *et al.* 1991), pero no se sabe aún si son proteínas activas o no.

Por último es probable que no sea necesario que este motivo esté ocupado en cada una de sus posiciones por aminiácidos específicos, sino que mas bien el conjunto de aminiácidos presentes en él le confieran una estructura secundaria y terciaria que formen parte de la proteína activa. Para tratar de ayudarnos en esto son útiles los perfiles de hidropatía y carga que olvidamos, y que combinados con las nuevas reglas propuestas en éste trabajo nos permiten tener una mayor información acerca de este sitio conservado en todas las RTs. Esta nueva definición será de importancia en futuros estudios experimentales ahora en proceso.

Nota: Recientemente Hsu *et al.* 1992 (*J. Bacteriol* 174: 2384-2387), publicaron la secuencia de un retrón encontrado en la bacteria *Stigmatella aurantiaca*, cuyo dominio DD cumple con las nuevas reglas propuestas en este trabajo

REFERENCIAS

- Alberts BM. (1986). The function of the hereditary materials: biological catalyses reflect the cell's evolutionary history. *Am. Zool.* 26: 781-796.
- Argos P. (1988). A sequence motif in many polymerases. *Nucleic Acids Res.* 16: 9909-9916.
- Baltimore D. (1970). Viral RNA-Dependent DNA polymerase. *Nature* 226: 1209-1211.
- Baltimore D. (1985). Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. *Cell* 40: 481-482.
- Bernad A, Blanco L, Salas M. (1990 a). Site-directed mutagenesis of the YCDTDS amino acid motif of the ϕ 29 DNA polymerase. *Gene* 94: 45-51.
- Bernad A, Lázaro JM, Salas M, Blanco L. (1990 b). The highly conserved amino acid sequence motif Tyr-Gly-Asp-Thr-Asp-Ser in alpha-like DNA polymerases is required by phage ϕ 29 DNA polymerase for protein-primed initiation and polymerization. *Proc. Natl. Acad. Sci. USA.* 87: 4610-4614.
- Besansky NJ. (1990). A retrotransposable element from the mosquito *Anopheles gambiae*. *Mol. Cell. Biol.* 10: 863-871
- Bishop M. (1978). Retroviruses. *Ann. Rev. Biochem.* 47: 35-88.
- Blackburn, E.H. (1990). Telomeres and their synthesis. *Science* 249: 489-490
- Boer PH, Gray MW. (1988) Genes encoding a subunit of respiratory NADH dehydrogenase (ND1) and reverse transcriptase-like protein (RTL) are linked to ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *EMBO J.* 7: 3501-3508
- 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.
- Chiu I-M, Yaniv A, Dahlberg JE, Gazit A, Skuntz, SF, Tronick SR, Aaronson SA. (1985). Nucleotide sequence evidence for relationships of AIDS retrovirus to lentiviruses. *Nature* 317: 366-368.
- Cohen M, Larsson E. (1988). Human endogenous retroviruses. *Bio Essays* 9: 191-196.
- Crick, FHC. (1958). On protein synthesis. In *Symp. Soc. Exp. Biol.*, XII: 138-163.
- Crick FHC. (1970). Central dogma of molecular biology. *Nature* 227: 561-563.
- Cummings DJ, Michel F, McNally KL. (1989). DNA sequence analysis of the 24.5 kilobase pair cytochrome oxidase subunit I mitochondrial gene from *Podospira anserina*: a gene with sixteen introns. *Curr. Genet.* 16: 381-406.
- Delarue M, Poch O, Tordo N, Moras D, Argos P. (1990). An attempt to unify the structure of polymerases. *Protein Eng.* 3: 461-467.
- Doolittle RF, Feng D-F, Johnson MS, McClure MA. (1989). Origins and evolutionary relationships of retroviruses. *Quater. Rev. Biol.* 64: 1-30.
- Doolittle R, Feng D-F, McClure M, Johnson M. (1990). Retrovirus phylogeny and evolution. *Current Topics in Microbiology and Immunology* 157: 1-17.

- Duncan R, Mason CL, Nagy E, Long J-A, Dobos P. (1991) Sequence analysis of infectious pancreatic necrosis virus genome segment B and its encoded VP1 protein: a putative RNA-dependent RNA polymerase lacking the Gly-Asp-Asp motif. *Virology* 181: 541-552.
- Fuettner J, Hohn T. (1987). Involvement of nucleocapsids in reverse transcription: a general phenomenon? *TIBS* 12: 92-95.
- Garret JE, Knutzon DS, Carroll D. (1989). Composite transposable elements in *Xenopus laevis* genome. *Mol. Cell. Biol.* 9: 3018-3027.
- 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: 341-409.
- Gilbert W. (1986). The RNA world. *Nature* 319: 618.
- Gojobori T, Moriyama EN, Ina Y, Ikeo K, Miura T, Tsujimoto H, Hayami M, Yokoyama S. (1990). Evolutionary origin of human and simian immunodeficiency viruses. *Proc. Natl. Acad. Sci. USA.* 87: 4108-4111.
- Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, VandePol S. (1982). Rapid evolution of RNA genomes. *Science* 215: 1577-1585.
- Huijser P, Kirchhoff C, Lankenau D-H, Hennig W. (1988). Retrotransposon-like sequences are expressed in Y chromosomal lampbrush loops of *Drosophila hydei*. *J. Mol. Biol.* 203: 689-697.
- 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.
- Jin YK, Bennetzen JL. (1989). Structure and coding properties of *Bs1*, a maize retrovirus-like transposon. *Proc. Natl. Acad. Sci. USA.* 86: 6235-6239.
- Johnson M, McClure M, Feng D - F, Gray J, Doolittle R. (1986). Computer analysis of retroviral pol genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. *Proc. Natl. Acad. Sci. USA.* 83: 7648-7652.
- Kamer G, Argos P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res.* 12: 7269-7283.
- Katz RA, Skalka AM. (1990). Generation of diversity in retroviruses. *Ann. Rev. Genet.* 24: 409-445.
- 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 their end and has homology with mammalian LINES. *Mol. Cell. Biol.* 7: 1465-1475.
- Kück U. (1989). The intron of a plastid gene from green alga contains an open reading frame for a reverse transcriptase-like enzyme. *Mol. Gen. Genet.* 218: 257-265.
- Lampson B, Sun J, Hsu M-Y, Vallejo-Ramirez 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.
- Larder BA, Purifoy JM, Powell KL, Darby G. (1987). Site-specific mutagenesis of AIDS virus reverse transcriptase. *Nature* 327: 716-717.
- Lazcano A. (1986). Prebiotic evolution and the origin of cells. *Treb. Soc. Cat. Biol.* 39: 73-103.

- Lazcano A, Guerrero R, Margulis L, Oró J. (1988a). The evolutionary transillon from RNA to DNA in early cells. *J. Mol. Evol.* 27: 283-290.
- Lazcano A, Fastag J, Gariglio P, Ramírez C, Oró J. (1988b). On the early evolution of RNA polymerase. *J. Mol. Evol.* 27: 365-376.
- Lazcano A, Fox GE, Oró J. (1992a). Life before DNA: the origin and early evolution of early Archean cells. In R. P. Mortlock (ed). *The evolution of metabolic functions*. CRC Press, Boca Ratan.
- Lazcano A, Liaca V, Cappello R, Valverde V, Oró J. (1992b). The origin and early evolution of nucleic acids polymerases. *Adv. Space Res.* 12: 207-216.
- Lazcano A, Valverde V, Hernández G, Gariglio P, Fox GE and Oró J. (1992c). The origin of RT: theoretical and experimental aspects. *J. Mol. Evol.* (en prensa).
- Leis J, Baltimore D, Bishop JM, Coffin J, Fleissner E, Goff SP, Oroszlan S, Robinson H, Skalka AM, Temin HM, Vogt V (1988) Standardized and simplified nomenclature for proteins common to all retroviruses. *J. Virol.* 62: 1808-1809.
- 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.
- Nargang FE, Bell JB, Stohl LL, Lambowitz AM. (1984). The DNA sequence and genetic organization of a *Neurospora* plasmid suggest a relationship to introns and mobile elements. *Cell* 38: 441-453.
- Ono M, Yasunaga 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.
- Oroszlan S, Lufin R. (1990). Retroviral proteinases. *Current Topics in Microbiology and Immunology* 157: 153-185.
- Poch O, Sauvaget I, Delarue M, Tordo N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *The EMBO J.* 8: 3867-3874.
- Rhee S, Hunter E. (1990). A single amino acid substitution within the matrix protein of a type D retrovirus converts its morphogenesis to that of a type C retrovirus. *Cell* 63: 77-86.
- 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.
- Schuster W, Brennicke A. (1987). Plastid, nuclear and reverse transcriptase sequence in the mitochondrial genome of *Oenothera*: is genetic information transferred between organelles via RNA?. *The EMBO J.* 6: 2857-2863.
- 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.
- Schwartz D, Tizard R, Gilbert G. (1983). Nucleotide sequence of Rous Sarcoma Virus. *Cell* 32: 853-869.
- Shinnick T, Lerner RA, Sutcliffe G. (1981). Nucleotidic sequence of Moloney murine leukaemia virus. *Nature* 293: 543-548.

- Tanese N, Goff SP. (1988). Domain structure of the Moloney murine leukemia virus reverse transcriptase: mutational analysis and separate expression of the DNA polymerase and RNase H activities. *Proc. Natl. Acad. Sci. USA.* 85: 177-1781.
- Telch N. (1982). Taxonomy of retroviruses. In R Weiss, N Telch, H Varmus and J Coffin (Eds.), RNA tumor viruses, vol. 1. *Cold Spring Harbor Laboratory*, Cold Spring Harbor, N. Y. p. 25-207
- Temin HM. (1985). Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. *Mol. Biol. Evol.* 2: 455-468.
- Temin HM. (1989). Retrons in bacteria. *Nature* 339: 254-255.
- Temin HM, Mizutani S. (1970). RNA directed DNA polymerase in virions of Rous sarcoma virus. *Nature* 226: 1211-1213.
- Varmus HE. (1983). Retroviruses. In J Shapiro (Ed.), *Mobile Genetic Elements*. Academic Press, Inc., New York. p. 411-503.
- Varmus H. (1988). Retroviruses. *Science* 240: 1427-1435.
- Varmus H, Brown P. (1989). Retroviruses. In M Howe and D Berg (eds), *Mobile DNA Elements*. American Society of Microbiology, Washington, D. C. pp. 53-108
- Voytas DF, Ausubel FM. (1988). A copia-like transposable element family in *Arabidopsis thaliana*. *Nature* 336: 242-244.
- Weiner AM, Deininger PL, Efstratiadis A. (1986). Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Ann. Rev. Biochem.* 55: 631-661.
- Weiner AM, Maizels N. (1991). The genomic tag model for the origin of protein synthesis: further evidence from the molecular fossil record. In S. Osawa y T. Honjo (eds). *Evolution of life: fossil, molecular and culture* (Spring-Verlag; New York), pp. 51-56.
- Wintersberger U and Wintersberger E. (1987). RNA makes DNA: a speculative view of the evolution of DNA replication mechanisms. *TIG* 3: 198-202.
- Xiong Y, Eickbush TH. (1988a). 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 YH. (1988b). Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. *Mol. Biol. Evol.* 5: 675-690.
- Xiong Y, Eickbush YH. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *The EMBO J.* 9: 3353-3362.