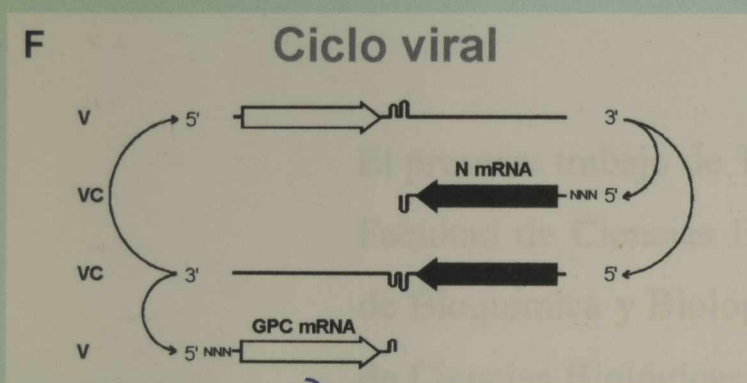
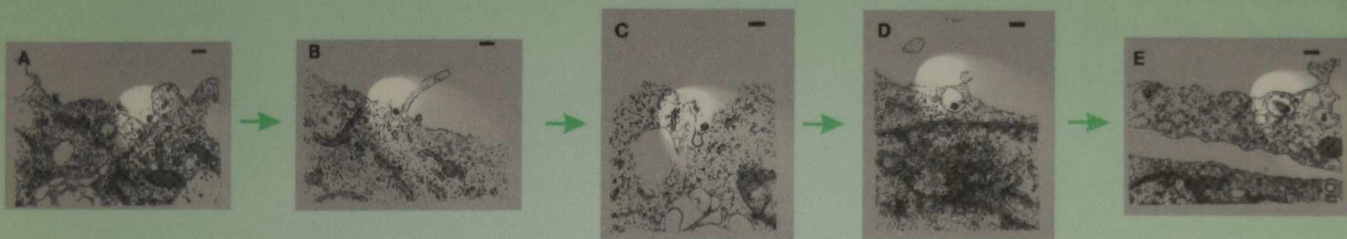




Universidad Nacional de La Plata
Facultad de Ciencias Exactas
Departamento de Ciencias Biológicas

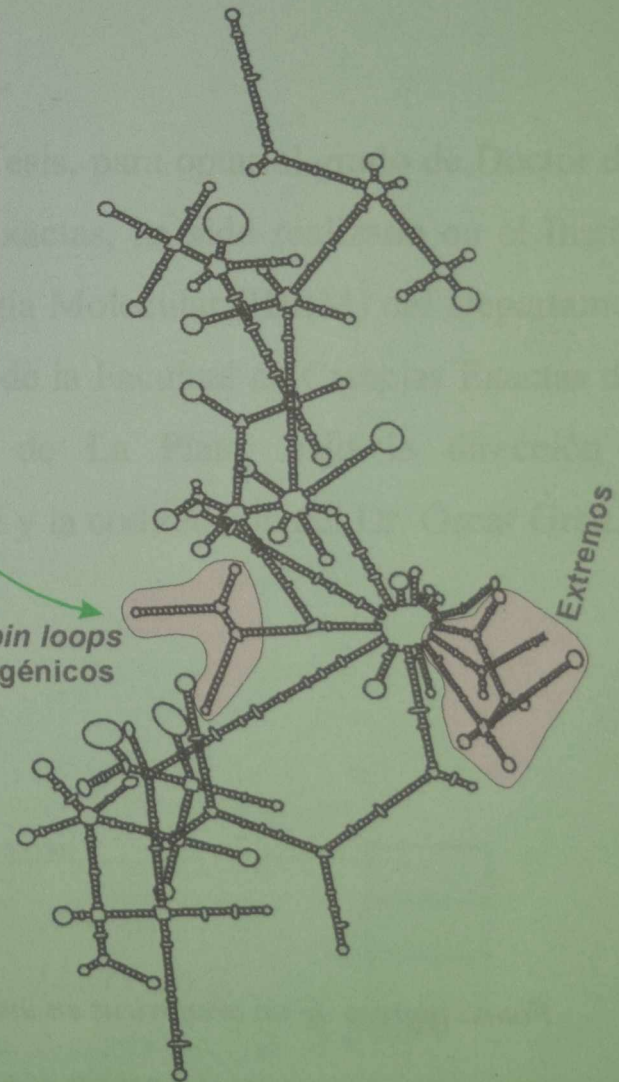
Virus Junín: clonado molecular y análisis estructural y funcional del RNA S y sus productos génicos



TESIS
AL 957 (1)
11



Dominio 3
Proteína N del virus Junín



Pablo Daniel Ghiringhelli
Tesis Doctoral
La Plata, 2002

DONACION.....
12.11.02

Fotos: ingreso de un arenavirus en una célula (tomadas de Borrow & Oldstone, 1994)

Estructura secundaria del RNA S de la cepa Candid #1 del virus Junín

Modelo de la potencial estructura del dominio 3 (zinc finger) de la proteína de la nucleocápside de la cepa MC2 del virus Junín



Biblioteca Central
Fac. Cs. Exactas
U.N.L.P

El presente trabajo de Tesis, para optar al grado de Doctor de la Facultad de Ciencias Exactas, ha sido realizado en el Instituto de Bioquímica y Biología Molecular (IBBM) del Departamento de Ciencias Biológicas de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata, bajo la dirección del Dr. Víctor Romanowski y la codirección del Dr. Oscar Grau.

DEPOSITO...

A...

Fecha 12-11-02

Inv. B.....Inv.....B

50988



Biblioteca Central
Fac. Cs. Exactas
U.N.L.P.

Profesor Pablo Daniel Ghiringhelli
Tesista

Dr. Víctor Romanowski
Director

Dr. Oscar Grau
Co-Director

TOMO I

Partes I a IV

Presentación de la información

Con el fin de facilitar la presentación y lectura, la información contenida en la presente Tesis se organizó modularmente en partes y capítulos, y en un caso en secciones.

En la **Parte I**, que comprende los *Capítulos 1 a 4*, se presenta una introducción a los virus con genoma de RNA, a la familia *Arenaviridae* y, especialmente, al virus Junín.

En las **Partes II** (*Capítulos 5 a 8*) y **III** (*Capítulos 9 a 11*), se presentan sintéticamente los resultados y aspectos informativos más elementales derivados del clonado, secuenciamiento y análisis estándar de la información de secuencia obtenida. En el *Capítulo 8*, en particular, se presentan los resultados obtenidos en ensayos de expresión de la proteína de la nucleocápside de la cepa MC2 del virus Junín en el sistema baculovirus-células de insecto y su empleo en experimentos de repliegamiento, ELISA, unión a RNA y unión a Zn⁺²; algunos de los resultados presentados en este capítulo dieron origen a planes de Tesis concretados en el IBBM en los últimos años. Estas dos partes fueron pensadas sólo como presentación de la información, razón por la cual carecen de una discusión y de conclusiones generales.

En la **Parte IV**, que comprende los *Capítulos 12 a 14*, se presenta el análisis comparativo de los RNAs S de los arenavirus y de las proteínas codificadas en los mismos, con las discusiones correspondientes. En particular, el *Capítulo 12*, que corresponde a los ácidos nucleicos de arenavirus, se subdividió en tres secciones. En los primeros seis títulos de la Sección I se detalla un análisis bioinformático que ahonda en aspectos informativos no estándar de las moléculas de ácidos nucleicos *per se*, mientras que en los últimos tres títulos se presentan aspectos estructurales de los RNAs S de arenavirus y la posible correlación con los mecanismos de recombinación. En la Sección II se discute sobre la variabilidad en los extremos de RNAs y el posible origen de los mismos y en la Sección III se reseña una análisis que correlaciona los RNAs S, el rRNA 28S y las proteínas de la nucleocápside (N). Por otra parte, en los *Capítulos 13 y 14* se discute acerca del análisis de las estructuras primarias y secundarias de las glicoproteínas y las nucleoproteínas de arenavirus, respectivamente.

En la **Parte V**, que comprende los *Capítulos 15 y 16*, se detallan los materiales empleados y se describen los métodos utilizados en el desarrollo de este trabajo.

En la **Parte VI**, que comprende los *Capítulos 17 y 18*, se reseñan algunos desarrollos que se realizaron durante el trabajo de Tesis. En el *Capítulo 17* se muestra el diseño y puesta a punto de tres técnicas experimentales y en el *Capítulo 18* se presenta el diseño y programación de varios algoritmos que han permitido realizar la mayor parte de los análisis computacionales presentados en los capítulos anteriores.

En la **Parte VII** se incluyen las conclusiones generales de la Tesis, destacando algunos de los aspectos más relevantes del trabajo.

En la **Parte VIII** se lista la bibliografía leída y utilizada durante todos estos años de trabajo.

Finalmente, la **Parte IX** incluye una serie de apéndices con información y/o breves análisis matemáticos que consideré importante incluir.

Indice

PARTE I - INTRODUCCION	1
CAPITULO 1 - VIRUS CON GENOMA DE RNA	3
<i>Una fracción del mundo del RNA</i>	3
Diversidad viral	3
RNA de simple cadena	2
RNA de doble cadena	4
Segmentación	4
Versatilidad vs. estabilidad	4
Recombinación	5
Evolución y virus emergentes	4
<i>Panorama de los virus con genoma de RNA de simple cadena y polaridad negativa</i>	6
CAPITULO 2 - ARENAVIRUS	7
<i>Historia y Taxonomía</i>	7
<i>Ecología y Epidemiología</i>	8
CAPITULO 3 - BIOLOGÍA DE LOS ARENAVIRUS	11
<i>Características de los viriones</i>	11
Ultraestructura	11
Estructura molecular	13
Morfogénesis	14
<i>Estructura y propiedades del genoma</i>	14
Genoma viral	14
Estrategia de codificación	16
Mecánica de la transcripción/replicación	17
<i>Proteínas virales</i>	18
Proteína de la nucleocápside (N)	19
Precursor de las glicoproteínas (GPC)	20
RNA polimerasa viral (L)	22
Proteína Z	24
<i>Patogénesis viral</i>	27
Infección	27
Viriones defectivos e interferencia viral	28
Infecciones agudas e infecciones persistentes	29
Citopagenicidad	30
CAPITULO 4 - VIRUS JUNIN	32
<i>El virus Junín y la fiebre hemorrágica argentina</i>	32
Epidemiología de la fiebre hemorrágica argentina	32
Fisiopatología de la infección por virus Junín	33
Principales manifestaciones clínicas de la enfermedad	34
Diagnóstico y tratamiento	36
Terapéutica	36
Diagnóstico clínico	38
Diagnóstico etiológico	38
<i>Variabilidad genética en el virus Junín</i>	40

PARTE II - RESULTADOS: VIRUS JUNIN CEPA MC2	41
CAPITULO 5 - CLONADO Y SECUENCIAMIENTO DEL RNA S - CEPA MC2	43
<i>Características de la cepa MC2</i>	43
<i>Crecimiento en células BHK-21</i>	43
<i>Purificación de RNA viral</i>	43
<i>Síntesis de cDNA, clonado y secuenciamiento del cDNA</i>	43
CAPITULO 6 - INFORMACION MOLECULAR - ACIDOS NUCLEICOS - CEPA MC2	47
<i>Secuencia nucleotídica del RNA S de la cepa MC2</i>	47
<i>Regiones codificantes</i>	48
Gen del precursor de las glicoproteínas (GPC)	48
Gen de la proteína de la nucleocápside (N)	50
<i>Regiones no codificantes</i>	51
Región intergénica	51
Extremos 5' y 3' del genoma viral	51
<i>Reconstrucción de genes</i>	52
Reconstrucción del gen GPC	52
Reconstrucción del gen N	53
CAPITULO 7 - INFORMACION MOLECULAR - PROTEINAS - CEPA MC2	54
<i>Glicoproteínas virales</i>	54
Estructura primaria del precursor de las glicoproteínas	54
Estructura del péptido señal	54
Estructura de las glicoproteínas G1 y G2	55
<i>Proteína de la nucleocápside</i>	59
Estructura de la proteína N	59
CAPITULO 8 - EXPRESIÓN DEL GEN N DE LA CEPA MC2 EN EL SISTEMA BACULOVIRUS-CELULAS DE INSECTO	63
<i>Clonado del gen N en un vector de transferencia, generación de baculovirus recombinantes y análisis del producto de expresión</i>	63
<i>Empleo del producto de expresión en ensayos de ELISA</i>	64
<i>Empleo del producto de expresión en ensayos de replegamiento y recuperación de inmunoreactividad</i>	64
<i>Empleo del producto de expresión en ensayos de unión a RNA</i>	64
<i>Empleo del producto de expresión en ensayos de unión a zinc</i>	65
<i>Empleo del producto de expresión en ensayos de unión a RNA, en presencia o ausencia de zinc</i>	65
PARTE III - RESULTADOS: VIRUS JUNIN CEPA VACUNAL CANDID #1	75
CAPITULO 9 - CLONADO Y SECUENCIAMIENTO DEL RNA S - CEPA CANDID #1	77
<i>Características de la cepa Candid #1</i>	77
<i>Crecimiento en células FRhL-2</i>	78
Preparación de un stock de trabajo	78
Preparación de viriones para estudios bioquímicos	78
<i>Purificación de RNA viral</i>	78
<i>Síntesis de cDNA, clonado y secuenciamiento del cDNA</i>	78
CAPITULO 10 - INFORMACION MOLECULAR - ACIDOS NUCLEICOS - CEPA CANDID #1	83
<i>Secuencia nucleotídica del RNA S de la cepa Candid #1</i>	83
<i>Regiones codificantes</i>	84
Gen del precursor de las glicoproteínas (GPC)	84
Gen de la proteína de la nucleocápside (N)	86

<i>Regiones no codificantes</i>	87
Región intergénica	87
Extremos 5' y 3' del genoma viral	87
CAPITULO 11 - INFORMACION MOLECULAR - PROTEINAS - CEPA CANDID #1	88
<i>Glicoproteínas virales</i>	88
Estructura primaria del precursor de las glicoproteínas	88
Estructura del péptido señal	88
Estructura de las glicoproteínas G1 y G2	89
<i>Proteína de la nucleocápside</i>	93
Estructura de la proteína N	93
PARTE IV - ANALISIS COMPARATIVO DE SECUENCIAS Y DISCUSION	97
CAPITULO 12 - ACIDOS NUCLEICOS DE ARENAVIRUS - RNAs S	99
SECCIÓN I - ASPECTOS INFORMATIVOS Y ESTRUCTURALES	99
<i>Consideraciones generales</i>	99
<i>Comparación de los RNAs S de arenavirus</i>	100
Tamaño y composición	100
Homología de secuencias	100
<i>Complejidad local de la información</i>	102
<i>Frecuencias nucleotídicas</i>	108
<i>Abundancia relativa de oligonucleótidos</i>	109
Análisis global	109
Análisis diferencial por regiones	110
Análisis diferencial por ventanas	110
<i>Discusión sobre los aspectos informativos de los RNAs S de arenavirus</i>	120
Complejidad local de composición (LCC)	120
Frecuencias A+U vs C+G	121
Abundancia relativa de oligonucleótidos	121
<i>Predicción de la estructura secundaria del RNA</i>	124
<i>Estructuras secundarias predichas para los RNAs S de arenavirus</i>	124
Estructuras secundarias completas	124
Estructuras secundarias en los extremos y en las regiones intergénicas de los RNAs S	125
Estructura secundaria de RNA y mecanismos de recombinación	134
<i>Discusión sobre la estructura secundaria de los RNAs S y posibles mecanismos de recombinación</i>	148
La necesidad, la oportunidad	148
Respuestas celulares apoptóticas	149
Peculiaridades de las replicasas	149
Clivaje mediado por otras RNAsas	150
La recombinación en sí misma	150
La estructura secundaria y la recombinación en los arenavirus	151
Un caso particular, el clon viral 6 ₂ de la cepa MC2 del virus Junín	154
Conclusiones generales	156
SECCIÓN II - EXTREMOS 5' DE LOS RNAs VIRALES	173
<i>Extremos 5' de los RNAs virales obtenidos a partir de viriones y a partir de células infectadas</i>	173
<i>Discusión sobre los extremos 5' de los RNAs genómico, antígenómico y mRNAs</i>	175

SECCIÓN III - RNAs DE ARENAVIRUS Y RIBOSOMAS	188
<i>RNAs de arenavirus, ribosomas y RNAs ribosomales</i>	188
<i>Discusión sobre los RNAs de arenavirus, ribosomas y RNAs ribosomales</i>	189
CAPITULO 13 - GLICOPROTEINAS DE ARENAVIRUS - ANALISIS COMPARATIVO Y DISCUSION	200
<i>Análisis comparativo de las proteínas GPC de las cepas MC2 y CD1 del virus Junín</i>	200
Estructura primaria	200
Estructura secundaria	200
Discusión sobre las GPCs de CD1 y MC2	200
<i>Análisis comparativo de las proteínas GPC de los arenavirus</i>	208
Estructura primaria	208
Estructura secundaria	209
Discusión sobre las GPCs de arenavirus	210
Péptido señal	210
Glicoproteína G1	211
Glicoproteína G2	213
CAPITULO 14 - PROTEINAS DE LA NUCLEOCAPSIDE DE ARENAVIRUS - ANALISIS COMPARATIVO Y DISCUSION	243
<i>Análisis comparativo de las nucleoproteínas de las cepas MC2 y CD1 del virus Junín</i>	243
Estructura primaria	243
Estructura secundaria	243
Discusión sobre las proteínas N de CD1 y MC2	243
<i>Análisis comparativo de las proteínas N de los arenavirus</i>	246
Estructura primaria	246
Estructura secundaria	246
Discusión sobre las proteínas N de los arenavirus	246
 PARTE V - MATERIALES Y METODOS	 272
CAPITULO 15 - MATERIALES	274
<i>Reactivos comerciales</i>	274
Productos químicos y enzimas	274
Material radioactivo	274
<i>Células, virus y muestras clínicas</i>	274
Cepas bacterianas	274
Líneas celulares eucariotas	274
Virus	274
Muestras clínicas	275
<i>Medios de cultivo para bacterias</i>	275
Medio de desarrollo	275
Medio para selección de recombinantes	275
Medio para el desarrollo de competencia	275
<i>Soluciones para la purificación y el análisis de los ácidos nucleicos</i>	276
Purificación de RNA	276
Preparación de plásmidos según el protocolo de lisis por álcali	276
Análisis de ácidos nucleicos mediante electroforesis en geles de agarosa	277
Southern blot	277
Reacciones enzimáticas	278
Secuenciamiento nucleotídico de DNA	278
Geles de poliacrilamida	278

CAPITULO 16 - METODOS	279
<i>Preparación de virus y células infectadas</i>	279
Propagación de virus	279
Condiciones de bioseguridad	279
Titulación de virus	280
<i>Preparación de RNA</i>	280
Control de la actividad de las RNasas	280
Extracción de RNA de células infectadas	280
<i>Preparación de baculovirus y células de insecto infectadas</i>	281
Cultivo de células	281
Propagación de virus	281
Purificación de DNA viral	281
<i>Preparación de extractos proteicos</i>	282
Para análisis por SDS-PAGE y Western blot	282
Para ensayos de replegamiento	282
<i>Ensayos de interacción proteína-Zn y/o proteína-RNA</i>	282
Interacción proteína-Zn	282
Interacción proteína-RNA	282
<i>Ensayos de ELISA</i>	282
<i>Preparación de DNA</i>	283
Mini-preparaciones de plásmidos (lisis por álcali)	283
Midi-preparaciones de plásmidos	283
Maxi-preparaciones de plásmidos	283
<i>Análisis de ácidos nucleicos por electroforesis</i>	284
Análisis de ácidos nucleicos por electroforesis en geles de agarosa	284
Análisis de DNA o RNA en geles de agarosa en condiciones nativas	284
Análisis de RNA en geles de agarosa en condiciones desnaturalizantes	285
<i>Análisis de ácidos nucleicos por hibridación</i>	286
Northern blot	286
Southern blot	286
Southern blot para fragmentos de menos de 300 pb	287
Dot blot o slot blot	287
Revelado de los híbridos	287
Prehibridación	287
Hibridación	287
Lavados de la membrana	288
Detección de los híbridos sobre la membrana	288
<i>Técnicas de clonado de DNA</i>	288
Digestión de DNA con enzimas de restricción	288
Ligación de los productos de la digestión	288
Preparación de células competentes (Método de cloruro de calcio)	289
Transformación de células competentes	289
Caracterización de los clones recombinantes	290
<i>Preparación de sondas de ácidos nucleicos</i>	290
Síntesis de los nucleótidos radioactivos	290
Preparación de sondas de DNA	291
Preparación de sondas de DNA por nick-translation	291
Preparación de sondas de DNA por extensión de oligonucleótidos	291
Preparación de sondas de RNA	292
Incorporación de marca por transferencia de fosfatos a la posición 5'	292

<i>Purificación de los ácidos nucleicos</i>	292
Precipitación alcohólica en presencia de sales	292
Purificación de ácidos nucleicos por gradiente de densidad	293
Purificación por columnas de exclusión molecular	293
Purificación por electroforesis en agarosa de bajo punto de fusión	293
Tratamiento con DNasa RQ1	294
Precipitación diferencial del RNA	294
Purificación de los oligonucleótidos marcados	294
<i>Síntesis de cDNA</i>	294
<i>Análisis del DNA por secuenciamiento nucleotídico</i>	296
Obtención de DNA de cadena simple	296
Secuenciamiento de fragmentos de DNA por el método de Sanger	296
<i>Métodos computacionales</i>	297
Gráficos cartesianos de homología	297
Búsqueda de regiones codificantes	297
<i>Cálculo de LCC (complejidad local de composición)</i>	298
Frecuencia relativa de nucleótidos (A+U vs C+G)	299
Abundancia relativa de oligonucleótidos	299
<i>Manipulación de datos</i>	300
<i>Detección de puntos potenciales de recombinación en moléculas de RNA, construcción y selección de las moléculas recombinantes basado en criterios de presencia de marcos de lectura abiertos y variaciones en la longitud de las proteínas codificadas</i>	300
<i>Análisis comparativo de estructuras secundarias de RNA (óptimas y subóptimas) y determinación de la estructura más frecuente</i>	300
<i>Determinación de anfipaticidad en estructuras secundarias y perfiles de hidrofobicidad</i>	301
<i>Modelado de estructuras proteicas por homología con homólogos remotos</i>	301
PARTE VI - MISCELANEA	302
CAPITULO 17 - PUESTA A PUNTO, MODIFICACION Y DISEÑO DE TECNICAS	304
<i>Síntesis de nucleósidos trifosfato marcados con ³²P</i>	304
<i>Un protocolo rápido para visualizar placas virales en experimentos de titulación</i>	312
<i>Ligación de oligonucleótidos a híbridos cDNA/RNA catalizada por la DNA ligasa del fago T4 y amplificación por PCR: bases para un método rápido de clonado de extremos 5' de mRNAs y RNAs virales</i>	314
CAPITULO 18 - DISEÑO Y PROGRAMACION DE RUTINAS DE COMPUTACION	324
<i>Representación gráfica de similitud (Rutina HOMGRAPH)</i>	324
<i>Complejidad local de composición (LCC)</i>	327
<i>Frecuencia relativa de nucleótidos A+U vs C+G</i>	338
<i>Abundancia relativa de oligonucleótidos</i>	341
<i>Manipulación de datos (Datmanip)</i>	346
<i>Análisis comparativo de estructuras secundarias de RNA (óptimas y subóptimas) y determinación de la estructura más frecuente</i>	346
<i>Detección de puntos potenciales de recombinación en moléculas de RNA, construcción y selección de las moléculas recombinantes basado en criterios de presencia de marcos de lectura abiertos y variaciones en la longitud de las proteínas codificadas</i>	352
<i>Análisis de anfipaticidad en estructuras secundarias de proteínas</i>	358
PARTE VII - CONCLUSIONES GENERALES	362
CONCLUSIONES GENERALES	364

PARTE VIII - BIBLIOGRAFIA	367
REFERENCIAS BIBLIOGRAFICAS	369
PARTE IX - APENDICES	394
APENDICE 1 -	396
<i>Alineamiento múltiple de los RNAs S de arenavirus</i>	396
<i>Alineamiento de los RNAs S de las cepas MC2 y Candid #1 del virus Junín</i>	412
APENDICE 2 -	417
<i>Alineamiento múltiple de los RNAs S de bunyavirus. Grupo mayoritario</i>	417
<i>Alineamiento múltiple de los RNAs S de bunyavirus. Grupo minoritario</i>	420
<i>Perfiles promedio de LCC de ambos grupos</i>	423
APENDICE 3 -	424
<i>Ligación de oligonucleótidos a híbridos cDNA/RNA, catalizada por la DNA ligasa del fago T4 y amplificación por PCR: bases para un método rápido de clonado de extremos 5' de mRNAs y RNAs virales. Detalle experimental</i>	424
APENDICE 4 -	432
<i>Gráficos de LCC individuales de los RNAs S de los arenavirus</i>	432
APENDICE 5 -	434
<i>Parte I - Transformada discreta de Fourier</i>	434
<i>Parte II - Índice de anfipaticidad</i>	437
APENDICE 6 -	440
<i>Alineamientos múltiples de las proteínas codificadas por el RNA S de los arenavirus</i>	440
APENDICE 7 -	445
<i>Análisis de períodos dominantes en estructuras secundarias predichas de las proteínas GPC y N de arenavirus</i>	445
APENDICE 8 -	447
<i>Predicciones estructurales en proteínas</i>	447
Introducción	447
Predicciones de estructura secundaria (1D) utilizando alineamientos múltiples: análisis comparativo de las predicciones realizadas con el programa GOR y con el servidor Jpred2	447
Modelado de estructuras proteicas (2D y 3D)	450
Predicciones de estructuras proteicas <i>ab initio</i> a partir de las secuencias aminoacídicas	450
Modelado por homología (Homology modelling)	450
Modelado por homología con homólogos remotos (remote homology modelling)	451
Conclusiones	451
APENDICE 9 -	454
<i>Tablas de energía libre utilizadas en el cálculo de estructuras secundarias de RNA</i>	454
APENDICE 10 -	458
<i>Esperanza probabilística de oligonucleótidos en los RNA S de arenavirus y genomas de mamíferos</i>	458
APENDICE 11 -	464
<i>Repeticiones de secuencias de la región intergénica del RNA S de la cepa Candid #1 del virus Junín y los rRNAs 18S y 28S humanos</i>	464
APENDICE 12 -	469
<i>Alineamientos múltiples de secuencias parciales de GPC correspondientes a diferentes aislamientos del virus Junín</i>	469

Indice de Figuras y Tablas

Figuras

Figura 2.1. Modos de transmisión y desarrollo de la infección del virus Machupo en <i>Calomys callosus</i>	10
Figura 3.1. Esquema de una partícula viral de arenavirus	12
Figura 3.2. Micrografías electrónicas de arenavirus	13
Figura 3.3. Micrografías electrónicas de nucleocápsides purificadas	15
Figura 3.4. Organización molecular del RNA S del virus Junín	16
Figura 3.5. Transcripción/Replicación del RNA S del virus Junín	18
Figura 3.6. Regiones variables y regiones conservadas en las proteínas L de LCM, Pichindé y Tacaribe	25
Figura 3.7. Alineamiento múltiple de las proteínas Z de LCM, Pichindé y Tacaribe	26
Figura 3.8. Esquema hipotético de interacciones entre proteínas del virión	27
Figura 5.1. Etapas del clonado viral de la cepa MC2	43
Figura 5.2. Análisis por Northern blot de clones de cDNA de la cepa MC2	44
Figura 5.3. Clonado y estrategia de secuenciación del RNA S de la cepa MC2 del virus Junín	46
Figura 6.1. Secuencia nucleotídica del RNA S de la cepa MC2 del virus Junín	47
Figura 6.2. Región intergénica del RNA S de la cepa MC2 del virus Junín	51
Figura 6.3. Extremos 5' y 3' del RNA S de la cepa MC2 del virus Junín	51
Figura 6.4. Reconstrucción del gen GPC de la cepa MC2 del virus Junín	52
Figura 6.5. Reconstrucción del gen N de la cepa MC2 del virus Junín	53
Figura 7.1. Precursor de las glicoproteínas (GPC) de la cepa MC2 del virus Junín.	54
Figura 7.2. Análisis de hidropatía y accesibilidad de la proteína GPC de MC2	56
Figura 7.3. Distribución de prolina y glicinas en las glicoproteínas G1 y G2	57
Figura 7.4. Proteína de la nucleocápside (N) de la cepa MC2 del virus Junín	59
Figura 7.5. Análisis de hidropatía y accesibilidad de la proteína N de MC2	60
Figura 7.6. Distribución de prolina y glicinas en la proteína N	61
Figura 8.1. Construcción de los plásmidos de transferencia pAcRP25-Njun y pAcUW2B-Njun	66
Figura 8.2. Cinética de expresión de la proteína N del virus Junín MC2 en células de insecto	67
Figura 8.3. Confirmación mediante Western blot de la identidad de la proteína N del virus Junín MC2 expresada en células de insecto infectadas con el recombinante Ac-N811	67
Figura 8.4. Fraccionamiento del lisado de células infectadas con el recombinante Ac-N811	68
Figura 8.5. Ensayo de ELISA utilizando el antígeno N expresado en células de insecto infectadas con el recombinante Ac-N811	68
Figura 8.6. Recuperación de inmunoreactividad de la proteína N recombinante mediante un tratamiento de replegamiento <i>in vitro</i>	69
Figura 8.7. Ensayos de unión en dot de RNA a la proteína N recombinante	70
Figura 8.8. Unión de RNA a la proteína N recombinante en un ensayo de Northwestern	71
Figura 8.9. Ensayos de zinc blot con proteínas Ns wild type de arenavirus	72
Figura 8.10. Ensayos de zinc blot con la proteína N recombin	73
Figura 8.11. Ensayos de unión de RNA en dot en presencia o ausencia de zinc	74
Figura 9.1. Historia de pasajes de la cepa Candid #1	77
Figura 9.2. Síntesis de cDNA correspondiente al RNA S de la cepa Candid #1 del virus Junín	79
Figura 9.3. Síntesis de cDNA y amplificación por PCR de regiones del RNA S de la cepa Candid #1 del virus Junín	81
Figura 9.4. Clonado y estrategia de secuenciación del RNA S de la cepa Candid #1 del virus Junín	82
Figura 10.1. Secuencia nucleotídica del RNA S de la cepa CD1 del virus Junín	83
Figura 10.2. Región intergénica del RNA S de la cepa CD1 del virus Junín	87
Figura 10.3. Extremos 5' y 3' del RNA S de la cepa CD1 del virus Junín	87
Figura 11.1. Precursor de las glicoproteínas (GPC) de la cepa Candid #1 del virus Junín	88
Figura 11.2. Análisis de hidropatía y accesibilidad de la proteína GPC de CD1	90
Figura 11.3. Distribución de prolina y glicinas en las glicoproteínas G1 y G2	91

Figura 11.4. Proteína de la nucleocápside (N) de la cepa CD1 del virus Junín	93
Figura 11.5. Análisis de hidropatía y accesibilidad de la proteína N de CD1	94
Figura 11.6. Distribución de prolínas y glicinas en la proteína N	95
Figura 12.1. Regiones variables y regiones conservadas en los RNAs S de los arenavirus Junín MC2, Junín Candid #1, Oliveros, Pichindé, Sabiá, Tacaribe, Lassa Josiah, Lassa Nigeria, LCM Armstrong, LCM WE y Mopeia	101
Figura 12.2. Análisis de complejidad local de composición.	103
Figura 12.3. Gráfico de complejidad local de composición (LCC) de los RNAs S de los arenavirus del Nuevo Mundo (Junín-Candid #1, Junín-MC2, Oliveros, Pichindé, Sabiá y Tacaribe), ajustado a un alineamiento múltiple general	104
Figura 12.4. Gráfico de complejidad local de composición (LCC) de los RNAs S de los arenavirus del Viejo Mundo (Lassa-Nigeria, Lassa-Josiah, LCM-Armstrong, LCM-WE y Mopeia), ajustado a un alineamiento múltiple general	104
Figura 12.5. Gráficos de complejidad local de composición (LCC) de los RNAs S de las cepas MC2 y Candid #1 del virus Junín	105
Figura 12.6. Cladogramas correspondientes a las relaciones filogenéticas entre los arenavirus	106
Figura 12.7. Gráficos de homología relativa de LCC de diferentes pares de RNAs S de arenavirus	107
Figura 12.8. Gráficos de frecuencia nucleotídica relativa (A+U vs C+G)	108
Figura 12.9. Abundancia relativa de dinucleótidos en los RNAs S de arenavirus, polaridad viral. Completos y regiones no codificantes	111
Figura 12.10. Abundancia relativa de dinucleótidos en los RNAs S de arenavirus, polaridad viral. Región codificante para GPC, posiciones I+II y II+III	112
Figura 12.11. Abundancia relativa de dinucleótidos en los RNAs S de arenavirus, polaridad viral. Región codificante para GPC, posición III+I. Polaridad viral complementaria. Completo	113
Figura 12.12. Abundancia relativa de dinucleótidos en los RNAs S de arenavirus, polaridad viral complementaria. Regiones no codificantes y región codificante para N, posición I+II	114
Figura 12.13. Abundancia relativa de dinucleótidos en los RNAs S de arenavirus, polaridad viral complementaria. Región codificante para N, posiciones II+III y III+I	115
Figura 12.14.A. Perfiles de abundancia relativa de dinucleótidos en los RNAs S virales de los arenavirus Junín-Candid #1, Junín- MC2 y Oliveros	116
Figura 12.14.B. Perfiles de abundancia relativa de dinucleótidos en los RNAs S virales de los arenavirus Pichindé, Sabiá y Tacaribe	117
Figura 12.15.A. Perfiles de abundancia relativa de dinucleótidos en los RNAs S virales de los arenavirus Lassa-Nigeria, Lassa-Josiah y LCM-Armstrong	118
Figura 12.15.B. Perfiles de abundancia relativa de dinucleótidos en los RNAs S virales de los arenavirus LCM-WE y Mopeia	119
Figura 12.16. Correlación entre las abundancias relativas de los dinucleótidos CG y UG	123
Figura 12.17. Correlación entre las abundancias relativas de los dinucleótidos UA y UG	123
Figura 12.18. Estructura secundaria del RNA S viral del virus Junín-MC2	126
Figura 12.19. Estructura secundaria del RNA S viral del virus Junín-Candid #1	127
Figura 12.20. Estructura tipo panhandle correspondiente al RNA S viral del virus Junín-MC2	128
Figura 12.21. Estructura tipo panhandle correspondiente al RNA S viral del virus Junín-Candid #1	129
Figura 12.22. Estructuras tipo hairpin loop correspondientes a los RNAs S virales de los virus Junín-MC2 y Junín-Candid #1	130
Figura 12.23. Estructuras tipo panhandle correspondientes a los RNAs S virales de 11 arenavirus	131
Figura 12.24. Estructuras tipo hairpin loop correspondientes a los RNAs S virales de arenavirus del Nuevo Mundo	132
Figura 12.25. Estructuras tipo hairpin loop correspondientes a los RNAs S virales de arenavirus del Viejo Mundo	133
Figura 12.26. Regiones de secuencias parcial o totalmente complementarias en la región intergénica de los RNAs S de arenavirus del Nuevo Mundo	135
Figura 12.27. Comparación de las estructuras intergénicas de los arenavirus con doble estructura tipo hairpin loop	136
Figura 12.28. Regiones de secuencias parcial o totalmente complementarias en la región intergénica de los RNAs S de arenavirus del Viejo Mundo	137

Figura 12.29. Comparación de las estructuras intergénicas de los arenavirus con simple estructura tipo hairpin loop	138
Figura 12.30. Diferentes estructuras doble cadena formadas por apareamiento parcial en la región intergénica de dos moléculas de RNA S viral del virus Junín-MC2	139
Figura 12.31. Generación de un recombinante del virus Junín-MC2 con triple estructura intergénica	140
Figura 12.32. Generación de un recombinante del virus Junín-MC2 con simple estructura intergénica	141
Figura 12.33. Estructura tipo hairpin loop alternativo correspondiente al RNA S viral del virus Junín-MC2	142
Figura 12.34. Comparación de las estructuras intergénicas de los arenavirus Junín-Candi # y Junín-MC2	143
Figura 12.35. Diferentes estructuras doble cadena formadas por apareamiento parcial en la región intergénica de dos moléculas de RNA S viral del virus Mopeia	144
Figura 12.36. Generación de un recombinante del virus Mopeia con triple estructura intergénica	145
Figura 12.37. Generación de un recombinante del virus Mopeia con simple estructura intergénica	146
Figura 12.38. Comparación de las estructuras intergénicas del arenavirus Mopeia	147
Figura 12.39. Generación de un recombinante interespecífico entre el virus Junín-MC2 y el virus Sabiá	157
Figura 12.40. Sitios potenciales de recombinación entre el RNA S viral de la cepa MC2 del virus Junín y los RNAs S virales de distintos arenavirus, mediante los mecanismos de copy choice o transesterificación	158
Figura 12.41. Sitios potenciales de recombinación entre el RNA S viral de la cepa MC2 del virus Junín y los RNAs S virales complementarios de distintos arenavirus, mediante el mecanismo de primer alignment and extension	159
Figura 12.42. Generación de un recombinante interespecífico entre el virus Junín-MC2 y el virus Oliveros	160
Figura 12.43. Recombinación mediante transesterificación (no replicativa) entre los RNAs S defectivos (al menos el de Junín MC2) de los arenavirus Junín MC2 y Oliveros	161
Figura 12.44. Recombinación mediante transesterificación (no replicativa) entre los RNAs S defectivos (al menos el de Junín MC2) de los arenavirus Junín MC2 y Oliveros	161
Figura 12.45. Recombinación mediante transesterificación (no replicativa) entre los RNAs S defectivos (al menos el de Junín MC2) RNAs S de los arenavirus Junín MC2 y Oliveros	162
Figura 12.46. Recombinación mediante transesterificación (no replicativa) entre los RNAs S defectivos (al menos el de Junín MC2) de los arenavirus Junín MC2 y Oliveros	162
Figura 12.47. Generación de un recombinante interespecífico entre el virus Junín-MC2 y el virus Whitewater Arroyo, mediante el mecanismo de copy choice	163
Figura 12.48. Generación de un recombinante interespecífico entre el virus Junín-MC2 y el virus Whitewater Arroyo, mediante el mecanismo de copy choice	164
Figura 12.49. Generación de un recombinante interespecífico entre el virus Junín-MC2 y el virus Whitewater Arroyo, mediante el mecanismo de transesterificación	165
Figura 12.50. Generación de un recombinante interespecífico entre el virus Junín-MC2 y el virus Whitewater Arroyo, mediante el mecanismo de transesterificación	166
Figura 12.51. Generación de un recombinante interespecífico entre el virus Junín-MC2 y el virus Whitewater Arroyo, mediante el mecanismo de primer alignment and extension	167
Figura 12.52. Generación de un recombinante interespecífico entre el virus Junín-MC2 y el virus Whitewater Arroyo, mediante el mecanismo de primer alignment and extension	168
Figura 12.53. Análisis de homología del tercio inicial del precursor de las glicoproteínas (GPC) de los diferentes aislamientos del virus Junín	169
Figura 12.54. Diferencias de secuencia en el clon viral 62 de la cepa MC2 del virus Junín, interpretaciones sobre su origen	170
Figura 12.55. Histogramas correspondientes a las repeticiones perfectas detectadas en el banco de datos de EST (GenBank)	171
Figura 12.56. Gráficos correspondientes a las repeticiones perfectas detectadas en el banco de datos de EST (GenBank)	172
Figura 12.57. Extremos 5' de RNAs genómicos y subgenómicos obtenidos a partir de viriones y de células infectadas	174
Figura 12.58. Extremos 5' de mRNAs de virus con genoma de RNA	178
Figura 12.59. Extremos 5' de RNAs genómicos y subgenómicos de arenavirus	182

Figura 12.60. Orígenes posibles de los extremos 5' de los RNAs genómicos y subgenómicos de arenavirus	183
Figura 12.61. Modelos alternativos de transcripción basados en el RNA viral complementario de la cepa Candid #1 del virus Junín	184
Figura 12.62. Modelos alternativos de replicación basados en el RNA viral complementario de la cepa Candid #1 del virus Junín	185
Figura 12.63. Modelo explicativo de la Aparición de un extremo 5' genómico consenso en el virus Junín	186
Figura 12.64. Modelo explicativo de la Aparición de un extremo 5' genómico diferencial en el virus Junín versus el resto de los arenavirus	187
Figura 12.65. Análisis de homología entre los RNAs S de arenavirus y el rRNA 28S humano. A y B	191
Figura 12.65. Análisis de homología entre los RNAs S de arenavirus y el rRNA 28S humano. C y D	192
Figura 12.65. Análisis de homología entre los RNAs S de arenavirus y el rRNA 28S humano. E y F	193
Figura 12.65. Análisis de homología entre los RNAs S de arenavirus y el rRNA 28S humano. G y H	194
Figura 12.65. Análisis de homología entre los RNAs S de arenavirus y el rRNA 28S humano. I y J	195
Figura 12.65. Análisis de homología entre los RNAs S de arenavirus y el rRNA 28S humano. K y L	196
Figura 12.66. Gráficos tipo dot plot correspondientes al análisis de homología entre la región intergénica del RNA S del virus Junín-Candid #1 y el rRNA 28S humano	197
Figura 12.67. Gráficos tipo dot plot correspondientes al análisis de homología entre la región intergénica del RNA S del virus Junín-Candid #1 y el rRNA 18S humano	198
Figura 12.68. Modelo de la asociación proteína N - ribosomas inmaduros y brotación en los arenavirus	199
Figura 13.1. Comparación entre las proteínas GPC de las cepas MC2 y CD1 del virus Junín	201
Figura 13.2. Análisis estructural de la potencial región transmembrana de la glicoproteína G2 de cepas del virus Junín con diferente grado de virulencia	204
Figura 13.3. Representaciones esquemáticas de las α hélices de las cepas MC2 y CD1 del virus Junín	206
Figura 13.4. Representaciones tipo helical wheel de las α hélices de las cepas MC2 y CD1 del virus Junín	207
Figura 13.5. Regiones conservadas y regiones variables en las proteínas GPC de los arenavirus Junín-Candid #1, Junín-MC2, Oliveros, Pichindé, Sabiá, Tacaribe, Lassa-Josiah, Lassa-Nigeria, LCM-Armstrong, LCM-WE y Mopeia.	218
Figura 13.6. Gráficos de cajas mostrando las frecuencias y dispersión de los distintos aminoácidos en las proteínas GPC de los arenavirus	222
Figura 13.7. Comparación de estructuras secundarias predichas para el péptido señal de los arenavirus	223
Figura 13.8. Gráficos de distribución de las estructuras secundarias en las glicoproteínas G1 de los arenavirus	224
Figura 13.9.A. Comparación de estructuras secundarias predichas para las glicoproteínas G1 de los arenavirus	225
Figura 13.9.B. Perfiles de hidropatía de las glicoproteínas G1 de los arenavirus	227
Figura 13.9.C. Comparación de perfiles de hidropatía de pares seleccionados de glicoproteínas G1 de arenavirus	229
Figura 13.9.D. Cladograma correspondiente a relaciones estructurales entre las G1 de los arenavirus	230
Figura 13.11.A. Comparación de estructuras secundarias predichas para las glicoproteínas G2 de los arenavirus	232
Figura 13.11.B. Perfiles de hidropatía de las glicoproteínas G2 de los arenavirus	234
Figura 13.11.C. Comparación de perfiles de hidropatía de pares seleccionados de glicoproteínas G2 de arenavirus	236
Figura 13.11.D. Cladograma correspondiente a relaciones estructurales entre las G2 de los arenavirus	237
Figura 13.12. Representaciones tipo helical wheel de las α hélices transmembrana de los arenavirus del Viejo Mundo	238
Figura 13.13. Representaciones tipo helical wheel de las α hélices transmembrana de los arenavirus del Nuevo Mundo	239

Figura 13.14. Frecuencia de sitios potenciales de dimerización en las proteínas G1 y G2 de los arenavirus	240
Figura 13.15. Representaciones tipo helical wheel de las α hélices $\alpha 1$ de los arenavirus del Viejo Mundo	241
Figura 13.16. Representaciones tipo helical wheel de las α hélices $\alpha 1$ de los arenavirus del Nuevo Mundo	242
Figura 14.1. Comparación entre las nucleoproteínas de las cepas MC2 y CD1 del virus Junín	244
Figura 14.2. Representaciones tipo helical wheel de las α hélices $\alpha 3$ de las cepas Candid #1 y MC2 del virus Junín	245
Figura 14.3. Regiones conservadas y regiones variables en las nucleoproteínas de los arenavirus Junín-Candid #1, Junín-MC2, Machupo, Oliveros, Pichindé, Sabiá, Tacaribe, Lassa-Josiah, Lassa-Nigeria, LCM-Armstrong, LCM-WE y Mopeia	251
Figura 14.4. Gráficos de cajas mostrando las frecuencias y dispersión de los distintos aminoácidos en las proteínas N de los arenavirus	252
Figura 14.5. Gráficos tridimensionales mostrando las frecuencias y distribución de Histidinas y Cisteínas en las proteínas N de los arenavirus	253
Figura 14.6. Gráficos de distribución de las estructuras secundarias en las nucleoproteínas de los arenavirus	255
Figura 14.7.A. Comparación de estructuras secundarias predichas para las nucleoproteínas de los arenavirus	256
Figura 14.7.B. Perfiles de hidropatía de las proteínas N de los arenavirus	261
Figura 14.7.C. Comparación de perfiles de hidropatía de pares seleccionados de proteínas N de arenavirus	262
Figura 14.7.D. Cladograma correspondiente a relaciones estructurales entre las proteínas N de los arenavirus	263
Figura 14.8. Perfiles de carga neta local de las proteínas N de los arenavirus	264
Figura 14.9.A. Sitios consenso de fosforilación y amidación en las proteínas N de arenavirus	265
Figura 14.9.B. Frecuencia de sitios potenciales de dimerización en las proteínas N de los arenavirus	266
Figura 14.10. Potenciales Zinc fingers de la región C-terminal de las proteínas N de arenavirus	267
Figura 14.11. Diagrama esquemático planar de una posible organización espacial de los motivos estructurales que conforman el dominio 3 de la proteína N del Virus Junín	269
Figura 14.12. Modelos alternativos propuestos para la transición transcripción/replicación en los arenavirus. A. Primer modelo	270
Figura 14.12. Modelos alternativos propuestos para la transición transcripción/replicación en los arenavirus. B. Segundo modelo	271
Figura 17.1. Ruta bioquímica de la síntesis enzimática de nucleósidos trifosfato marcados con ^{32}P	305
Figura 17.2. Protocolo sintético para la síntesis enzimática de ($\gamma^{32}P$) ATP	306
Figura 17.3. Protocolo sintético para la síntesis enzimática de ($\alpha^{32}P$) ATP	307
Figura 17.4. Protocolo sintético para la síntesis enzimática de ($\alpha^{32}P$) dATP	308
Figura 17.5. Protocolo sintético para la síntesis enzimática de ($\alpha^{32}P$) dCTP	309
Figura 17.6. Autorradiografía analítica de las etapas de síntesis enzimática de ($\alpha^{32}P$) rATP	310
Figura 17.7. Autorradiografía analítica de diferentes síntesis de ($\alpha^{32}P$) dATP	310
Figura 17.8. Análisis comparativo de la eficiencia de incorporación de dos ($\alpha^{32}P$) dATP de distinto origen	311
Figura 17.9. Titulación de virus eucarióticos utilizando el protocolo de tinción con azul de metileno	313
Figura 17.10. Extensión de primer, adición de nucleótidos con TdT y PCR	317
Figura 17.11. Extensión de primer, adición de oligonucleótidos sintéticos y PCR	318
Figura 17.12. Amplificación de cDNAs con nucleótidos adicionados	319
Figura 17.13. Amplificación de cDNAs ligados a oligonucleótidos	320
Figura 17.14. Southern blot de los productos de amplificación de cDNAs ligados a oligonucleótidos	320
Figura 17.15. Extensión de primer y adición de oligonucleótidos con T4 DNA ligasa	321
Figura 17.16. Amplificación de cDNAs ligados a oligonucleótidos	322

Figura 18.1. Regiones variables y regiones conservadas en las proteínas L de LCM, Pichindé y Tacaribe	326
Figura 18.2. Ejemplo de archivo de salida de la rutina de cálculo individual de LCC.	328
Figura 18.3. Ejemplo de alineamiento múltiple a emplear para ajustar los datos individuales de LCC	329
Figura 18.4. Ejemplo de archivo de salida de la rutina de ajuste	329
Figura 18.5. Ejemplo de archivo de salida de la rutina de estadística	329
Figura 18.6. Gráficos de complejidad local de composición (LCC), derivados del paquete de rutinas de cálculo y análisis múltiple de LCC	330
Figura 18.7. Gráficos de complejidad local de composición (LCC) del RNA S del virus Junín-Candid #1	331
Figura 18.8. Gráficos de complejidad local de composición (LCC)	331
Figura 18.9. Gráficos de complejidad local de composición (LCC)	332
Figura 18.10. Gráficos de LCC del grupo minoritario	332
Figura 18.11. Gráficos de LCC del grupo mayoritario	334
Figura 18.12. Estudio del rango de frecuencias para análisis de Fourier	335
Figura 18.13. Ejemplo de archivo de salida de la rutina de cálculo de la transformada de Fourier	336
Figura 18.14. Ejemplo de archivo de salida de la rutina de comparación de amplitudes de Fourier	336
Figura 18.15. Diferentes ventanas de análisis de Fourier	337
Figura 18.16. Gráficos de frecuencia nucleotídica relativa (A+U vs C+G)	339
Figura 18.17. Análisis de Montecarlo de la distribución de frecuencias nucleotídicas relativas (A+U vs C+G)	340
Figura 18.18. Análisis de abundancia relativa de los 16 dinucleótidos posibles	344
Figura 18.19. Abundancia relativa de los dieciseis dinucleótidos posibles en el RNA S del virus Junín – Candid 1	345
Figura 18.20. Análisis comparativo para la determinación de la estructura secundaria más probable en los extremos del RNA S viral del virus Junín-MC2	348
Figura 18.21. Análisis comparativo para la determinación de la estructura secundaria más probable en los extremos del RNA S viral del virus Junín-Candid #1	349
Figura 18.22. Análisis comparativo para la determinación de la estructura secundaria más probable en los extremos del RNA S viral del virus Junín-Candid #1	350
Figura 18.23. Análisis comparativo para la determinación de la estructura secundaria más probable en la región intergénica del RNA S viral del virus Oliveros	351
Figura 18.24. Ejemplo de archivo derivado de la rutina BI2seq (Blast 2 sequences), derivado de la comparación entre los RNAs S de los virus Junín-Candid #1 y Oliveros	354
Figura 18.25. Ejemplo de archivo derivado de la rutina ANPARCOD	355
Figura 18.26. Ejemplo de archivo de salida de la rutina ARMAREC4	355
Figura 18.27. Ejemplo de Tabla generada con la rutina ARMAREC4 y corregida con la rutina REV TAB3	355
Figura 18.28. Ejemplo de procesamiento ulterior para mostrar la generación de un recombinante entre los RNAs S de Junín-Candid #1 y Oliveros, en la región codificante para la proteína de la nucleocápside (N), mediante el mecanismo de copy choice.	356
Figura 18.29. Gráficos derivados de los datos generados con las rutinas descritas en este capítulo	357
Figura 18.30. Ejemplo de archivo de ingreso para la rutina ANFIPMED	359
Figura 18.31. Ejemplos de Tablas de hidrofobicidad utilizadas	359
Figura 18.32. Ejemplo de archivo de salida de la rutina ANFIPATI	360
Figura 18.33. Ejemplo de perfil de Fourier derivado del análisis con la rutina ANFIPATI	361
Figura 18.34. Ejemplo de archivo de salida de la rutina EXTRPEEK	361
Figura 18.35. Ejemplo de archivo de salida de la rutina HIDROF	361
Figura Ap1.1. Alineamiento múltiple de los RNAs S de arenavirus	396
Figura Ap1.2. Alineamiento múltiple de los RNAs S de las cepas MC2 y Candid #1 del virus Junín	412
Figura Ap2.1. Alineamiento múltiple de los RNAs S de bunyavirus, Grupo mayoritario	417
Figura Ap2.2. Alineamiento múltiple de los RNAs S de bunyavirus, Grupo minoritario	420
Figura Ap2.3. Gráfico consenso de LCC del grupo mayoritario	423
Figura Ap2.4. Gráfico consenso de LCC del grupo minoritario	423

Figura Ap3.1. Purificación de la especie de 30 nucleótidos del primer NA	425
Figura Ap3.2. Amplificación de cDNAs ligados a oligonucleótidos	427
Figura Ap3.3. Amplificación de cDNAs ligados a oligonucleótidos	429
Figura Ap3.4. Amplificación de cDNAs ligados a oligonucleótidos	431
Figura Ap4.1. Gráficos de complejidad local de composición (LCC) de los RNAs S de los arenavirus del Nuevo Mundo Junín-Candid #1, Junín-MC2, Oliveros, Pichindé, Sabiá y Tacaribe	432
Figura Ap4.2. Gráficos de complejidad local de composición (LCC) de los RNAs S de los arenavirus del Viejo Mundo Lassa-Nigeria, Lassa-Josiah, LCM-Armstrong, LCM-WE y Mopeia	433
Figura Ap5.1. Perfil de hidropatía correspondiente a la secuencia $_{96}RVERMELASGLGNLKTKF_{113}$ de la proteína N del virus Junín-MC2, predicha como α hélice con el programa PHDsec	435
Figura Ap5.2. Perfil de hidropatía correspondiente a la secuencia $_{479}SEQARQFDQQVWEKFGHLCKH_{499}$ de la proteína N del virus Junín-MC2, predicha como α hélice con el programa PHDsec	435
Figura Ap5.3. Análisis discreto de Fourier Aplicado a la secuencia numérica de valores de hidrofobicidad de la Figura Ap5.1	436
Figura Ap5.4. Análisis discreto de Fourier Aplicado a la secuencia numérica de valores de hidrofobicidad de la Figura Ap5.2	436
Figura Ap5.5. Análisis discreto de Fourier Aplicado a la secuencia numérica de valores de hidrofobicidad de la α hélice $_{160}ESGAINGAISDIFGTLVEFY_{179}$ de una proteína globular de <i>Bacillus thermophilus</i>	438
Figura Ap5.6. Análisis discreto de Fourier Aplicado a la secuencia numérica de valores de hidrofobicidad de la α hélice $_{235}GIINKAAYLISQ_{246}$ de una proteína globular de <i>Bacillus thermophilus</i>	438
Figura Ap5.7. Análisis discreto de Fourier Aplicado a la secuencia numérica de valores de hidrofobicidad de la α hélice $_{1}GIGAILKVLATGLPTLISWIKNKRKQ_{26}$ de una proteína superficial de <i>Apis mellifera</i>	439
Figura Ap5.8. Análisis discreto de Fourier Aplicado a la secuencia numérica de valores de hidrofobicidad de la α hélice $_{11}ISLIVLFVTVTIIAIALIVVLA_{32}$ de una proteína de membrana humana (Succinato isomerasa)	439
Figura Ap6.1. Alineamiento múltiple correspondiente al péptido señal del precursor de las glicoproteínas (GPC) de los arenavirus	440
Figura Ap6.2. Alineamiento múltiple correspondiente a la glicoproteína G1 de los arenavirus	441
Figura Ap6.3. Alineamiento múltiple correspondiente a la glicoproteína G2 de los arenavirus	442
Figura Ap6.4. Alineamiento múltiple correspondiente a la nucleoproteína de los arenavirus	443
Figura Ap7.1. Análisis de períodos dominantes en las α hélices predichas en el péptido señal de los arenavirus del Viejo Mundo	445
Figura Ap7.2. Análisis de períodos dominantes en las α hélices predichas en el péptido señal de los arenavirus del Nuevo Mundo	446
Figura Ap8.1. Comparación de las secuencias aminoacídicas de la glicoproteína G1 de las cepas MC2 y Candid #1 del virus Junín	449
Figura Ap12.1. Alineamiento múltiple del primer tercio de la región codificante para el precursor de las glicoproteínas (GPC) de todos los aislamientos del virus Junín con secuencias disponibles actualmente	469
Figura Ap12.2. Alineamiento múltiple del primer tercio del precursor de las glicoproteínas (GPC) de todos los aislamientos del virus Junín con secuencias disponibles actualmente	470

Tablas

Tabla 1.a. Replicación/transcripción	6
Tabla 1.b. Proteínas principales	6
Tabla 2.a. Arenavirus	9
Tabla 3.a. Características de las proteínas N	20
Tabla 3.b. Características de las proteínas GPC	21
Tabla 3.c. Características de las proteínas L	23
Tabla 3.d. Comparación de las proteínas L	24
Tabla 3.e. Características de las proteínas Z	25
Tabla 3.f. Secuencia consenso de unión de zinc en las proteínas Z	26
Tabla 4.a. Características de las cepas del virus Junín	40
Tabla 6.a. Longitud de los RNAs S de arenavirus	48
Tabla 6.b. Marcos de lectura abiertos	48
Tabla 6.c. Frecuencia de uso de codones en el gen GPC	49
Tabla 6.d. Frecuencia de uso de codones en el gen N	50
Tabla 7.a. Composición aminoacídica del péptido señal y las glicoproteínas G1 y G2	55
Tabla 7.b. Estructuras secundarias predichas para la glicoproteína G1 de MC2	58
Tabla 7.c. Estructuras secundarias predichas para la glicoproteína G2 de MC2	58
Tabla 7.d. Sitios potenciales identificados en las glicoproteínas G1 y G2	58
Tabla 7. e. Composición aminoacídica de la proteína N	60
Tabla 7.f. Estructuras secundarias predichas para la proteína N de MC2	62
Tabla 7.g. Sitios potenciales identificados en la proteína N	62
Tabla 10.a. Longitud de los RNAs S de arenavirus	84
Tabla 10.b. Marcos de lectura abiertos	84
Tabla 10.c. Frecuencia de uso de codones en el gen GPC	85
Tabla 10.d. Frecuencia de uso de codones en el gen N	86
Tabla 11.a. Composición aminoacídica del péptido señal y las glicoproteínas G1 y G2	88
Tabla 11.b. Estructuras secundarias predichas para la glicoproteína G1 de CD1	92
Tabla 11.c. Estructuras secundarias predichas para la glicoproteína G2 de CD1	92
Tabla 11.d. Sitios potenciales identificados en las glicoproteínas G1 y G2	92
Tabla 11. e. Composición aminoacídica de la proteína	94
Tabla 11.f. Estructuras secundarias predichas para la proteína N de CD1	96
Tabla 11.g. Sitios potenciales identificados en la proteína N	96
Tabla 12.a. Longitud y composición nucleotídica de los RNAs S de arenavirus	100
Tabla 12.b. Homología de secuencias nucleotídicas	101
Tabla 12.c. Homología de LCC de los RNAs S de arenavirus	102
Tabla 12.d. Abundancia relativa de oligonucleótidos cortos en los RNAs S de arenavirus	109
Tabla 12.e. Frecuencias de uso de los codones de arginina en las proteínas GPC y N de los arenavirus	122
Tabla 12.f. Estructura secundaria de los RNAs S de arenavirus	124
Tabla 12.g. Estructuras secundarias 5-óptimas de los extremos 5' y 3' virales	125
Tabla 12.h. Estructuras secundarias 5-óptimas de las regiones intergénicas virales	125
Tabla 13.a. Cambios en las secuencias de las proteínas GPC de las cepas MC2 y CD1 del virus Junín	202
Tabla 13.b.1. Identidad de secuencias aminoacídicas. - Péptido señal	219
Tabla 13.b.2. Homología de secuencias aminoacídicas. - Péptido señal	219
Tabla 13.c.1. Identidad de secuencias aminoacídicas. - Glicoproteína G1	220
Tabla 13.c.2. Homología de secuencias aminoacídicas. - Glicoproteína G1	220
Tabla 13.d.1. Identidad de secuencias aminoacídicas. - Glicoproteína G2	221
Tabla 13.d.2. Homología de secuencias aminoacídicas. - Glicoproteína G2	221
Tabla 13.e.1. Similitud entre perfiles de hidropatía de las glicoproteínas G1 de arenavirus	228
Tabla 13.e.2. Distancia entre perfiles de hidropatía de las glicoproteínas G1 de arenavirus	228
Tabla 13.f.1. Similitud entre perfiles de hidropatía de las glicoproteínas G2 de arenavirus	235

Tabla 13.f.2. Distancia entre perfiles de hidropatía de las glicoproteínas G1 de arenavirus	235
Tabla 14.a. Cambios en las secuencias de las nucleoproteínas de las cepas MC2 y CD1 del virus Junín	245
Tabla 16.a. Rango de separación en geles conteniendo diferentes cantidades de agarosa	285
Tabla 16.b. Entorno consenso de AUGs eucariotas	297
Tabla 16.c. Valores de TestCode y predicciones	298
Tabla 17.a. Reactivos específicos	304
Tabla 17.b. Comparación de tres protocolos de tinción	313
Tabla 18.a. Características de las regiones variables y las regiones conservadas	325
Tabla 18.b. RNAs S de la familia Bunyaviridae	330
Tabla 18.c. Homología nucleotídica de 13 RNAs S de bunyavirus	333
Tabla 18.d. Homología de LCC de 13 RNAs S de bunyavirus	337
Tabla Ap8.a. Certeza en la predicción de distintos métodos	448
Tabla Ap11.a. CD1 RNA S (v) vs 18S rRNA Humano	464
Tabla Ap11.b. CD1 RNA S (vc) vs 18S rRNA Humano	464
Tabla Ap11.c. CD1 RNA S (v) vs 28S rRNA Humano	465
Tabla Ap11.d. CD1 RNA S (vc) vs 28S rRNA Humano	467

PARTE I

INTRODUCCION



Biblioteca Central
Fac. Cs. Exactas
U.N.L.P

56986

Capítulo 1

Virus con genoma de RNA

The Nature of Modern RNA Suggests a Prebiotic RNA World
The RNA World (Gesteland and Atkins, 1993)

UNA FRACCION DEL MUNDO DEL RNA (basado en Atkins, 1993)

Todos los genomas de RNA modernos están confinados a ser agentes infecciosos, ya que, desde el punto de vista de las funciones moleculares básicas, no son autosuficientes y necesitan, en mayor o menor medida, de la maquinaria bioquímica especificada en los genomas de DNA. En consecuencia, la evolución de un genoma de RNA requiere la preexistencia o la coevolución de su hospedador.

Diversidad viral

Existe un gran espectro de propiedades virales, incluyendo la composición genómica. La coexistencia en la naturaleza de genomas de diverso tamaño, complejidad y naturaleza; DNA o RNA; simple o doble cadena; polaridad positiva, negativa o ambas; lineal o circular; segmentada o monopartita; posiblemente otorgue una ventaja evolutiva. Dentro de este amplio espectro, los virus con genoma de RNA son ubicuos y son los parásitos intracelulares conocidos más abundantes. Una alta proporción de ellos tiene genomas de RNA de simple cadena.

RNA de simple cadena

En la diversidad de virus con genoma de RNA de simple cadena existen las tres posibilidades de estrategias de codificación, genomas de polaridad positiva, genomas de polaridad negativa y genomas con ambas polaridades (*ambisense*).

Tanto en organismos procariotas como eucariotas, es posible encontrar virus con

genoma de RNA de simple cadena, pero especialmente prevalecen entre los virus que infectan plantas terrestres (la gran mayoría de los cuales tiene polaridad positiva).

Una posible ventaja de la polaridad positiva es el escape a la necesidad de una etapa de transcripción DNA-dependiente, lo cual le otorga una cierta independencia de la maquinaria celular en las etapas moleculares iniciales. Una posible desventaja de la polaridad positiva es que la ausencia de una etapa de transcripción dificulta un control preciso de la expresión. Sin embargo, algunos virus parecen haber superado esta dificultad a través de sofisticados mecanismos, como, por ejemplo, la generación de poliproteínas con diferentes sitios proteolíticos y la regulación de su actividad en función de la ocurrencia de la proteólisis, la generación de diferentes productos de traducción por unión de los ribosomas a sitios de reconocimiento internos (IRES, *Internal Ribosome Entry Site*), etc.

Los virus de polaridad negativa tienen una estrategia replicativa más independiente del hospedador que los virus de polaridad positiva. Como su composición genómica implica, estos genomas no son infecciosos por sí mismos, lo cual hace necesario que enzimas codificadas por el virus y empaquetadas en el virión realicen las etapas transcripcionales iniciales. Además, se necesitan señales de iniciación y terminación de la transcripción para la generación de los mRNAs subgenómicos. Generalmente, la traducción de uno o más productos virales es un requisito previo al comienzo de la replicación. La cual usa como molde un intermediario

antigenómico de la misma longitud que el RNA viral. En los virus con estrategia *ambisense* (algunos bunyavirus y arenavirus), también se generan mRNAs a partir del intermediario de polaridad antigenómica. Sin embargo, en todos ellos este intermediario de polaridad antigenómica y longitud completa tiene por función principal el actuar de molde para la generación de nuevas moléculas genómicas.

RNA de doble cadena

Todos los virus con genoma de RNA de doble cadena conocidos contienen la información genética necesaria para la realización de los eventos de transcripción. Los mRNAs producidos son traducidos, dando origen a proteínas que secuestran los mismos mRNAs en partículas subvirales. Dentro de estas partículas ocurre la síntesis de las hebras negativas y la generación de la progenie viral.

Una consecuencia lógica de la organización genómica de doble cadena es su mayor estabilidad y la relativa resistencia al ataque por nucleasas celulares.

Durante la etapa de replicación existen intermediarios de simple cadena con regiones dúplex sólo en la zona de polimerización. Esta etapa es el punto débil de la estrategia, ya que, debido a la relativa inestabilidad de los intermediarios de simple cadena, la longitud de las moléculas de RNA de doble cadena se ve limitada.

Segmentación

La división del genoma en varias moléculas independientes evita los problemas de la limitación en longitud de los genomas, pero adiciona el inconveniente de que deben existir señales repetidas en cada una de las moléculas, como ocurre, por ejemplo, con la repetición de las señales de empaquetamiento en los 10 segmentos genómicos de los reovirus.

Al mismo tiempo, la división del genoma en varios segmentos hasta cierto punto

disminuye el problema de la inexistencia de un sistema corrector de errores durante las síntesis replicativas. Es bien conocido que, cuando dos cepas de un mismo virus con genoma segmentado, ambas con mutaciones letales en diferentes cadenas genómicas, coinfectan una célula, pueden encontrarse entre la progenie tanto viriones idénticos a los originales, como viriones que son dobles mutantes o que poseen un genoma viable, carente de las mutaciones. Este fenómeno se debe a reasociaciones genómicas que ocurren sin la necesidad de que exista recombinación.

Una característica adicional de los genomas segmentados es que prescinden de la necesidad de IRES, ya que los transcriptos que se producen son, generalmente, monocistrónicos. Esta relativa simplicidad, probablemente, refleje una de las características de los genomas de RNA primitivos.

Versatilidad vs. estabilidad

Una diferencia obvia entre el DNA y el RNA es la actividad nucleofílica potencial del hidroxilo 2' de la ribosa. Podría esperarse que esta característica aumente la versatilidad a expensas de una disminución en la estabilidad.

La luz UV produce más daños fotoquímicos en el RNA que en el DNA y el pH ácido favorece la depurinación de C a U, igual que en el DNA. Dada la inexistencia de sistemas específicos, ninguno de estos cambios pueden ser corregidos. También, el RNA es fácilmente hidrolizado, especialmente en la presencia de metales de transición. Ahora bien, este problema es de importancia relativa, ya que en muchos viriones la estructura de las cápsides es muy cerrada y el entorno más íntimo del RNA es netamente hidrofóbico, con lo cual el problema de la hidrólisis se presenta cuando el RNA está "desnudo", como ocurre durante la replicación.

Sin embargo, los genomas de RNA de simple cadena tienen mayor estabilidad que la predicha, caracterizándose por ser altamente

estructurados, y esto se debe, principalmente, a que el hidroxilo 2' forma interacciones adicionales (*stacking*) que aumentan la estabilidad.

Las estructuras secundarias y terciarias conocidas, o predichas, en muchos virus con genoma de RNA de simple cadena facilitan la interacción con otras macromoléculas, dificultan el ataque por parte de nucleasas celulares, ayudan a mantener algunas regiones del genoma como simple cadena durante los eventos de replicación por secuestro de aquellas secuencias que, potencialmente, podrían interactuar con las primeras.

Lo anteriormente enunciado, importante desde el punto de vista de la estabilidad, también juega un rol principal en los eventos de recombinación, especialmente en la recombinación heteróloga.

Recombinación

En cierto número de virus con genoma de RNA se conoce la ocurrencia de eventos de recombinación, tanto homóloga como heteróloga. Estos procesos, a nivel de mecánica, son muy diferentes del mecanismo de recombinación del DNA.

Durante el progreso de la replicación, el complejo replicativo puede disociarse de un molde y asociarse con otro. Si este fenómeno ocurre en posiciones de secuencias equivalentes de genomas homólogos, el producto es un recombinante homólogo (Kirkegaard and Baltimore, 1986; Jarvis and Kirkegaard, 1992; Lai, 1992).

La alta estructuración de las moléculas de RNA posibilita la existencia de otro tipo de equivalencia entre dos moléculas de RNA. Obviamente, dicha equivalencia ocurre al nivel de estructuras, principalmente secundarias. Aquí también puede ocurrir que el complejo replicativo se disocie de un molde y se reasocie con otro, dando como resultado recombinantes heterólogos. La heterología no se refiere

solamente a la posibilidad de recombinación entre moléculas de RNA de virus distintos o entre RNAs celulares y RNAs virales (Myers *et al.*, 1990), sino también a la posibilidad de recombinación entre moléculas de RNA de un mismo virus, pero en posiciones de secuencia diferentes, dando como resultado una progenie con inserciones y/o deleciones, duplicaciones génicas, etc. (Forss and Schaller, 1982; Lai, 1992). Este último caso puede servir de explicación para la aparición de partículas defectivas interferentes cuando, por alguna causa, existen altas chances de coinfección.

Evolución y virus emergentes

La estructura genómica particular, la versatilidad, la ausencia de mecanismos de corrección y, sobre todo, la recombinación, principalmente la heteróloga, facilitan una rápida y continua evolución de los genomas de RNA. Especialmente de los genomas de RNA de simple cadena y, en particular, de aquéllos segmentados, ya que a lo antes mencionado se agrega la posibilidad de reasociaciones entre los segmentos genómicos.

Todo lo mencionado hasta aquí habla en favor de la futura aparición en la naturaleza de variantes virales o nuevos virus (virus emergentes), actualmente se conocen varios ejemplos de virus que responden a estas características (Morse, 1993). También nos habla de la historia viral y de cómo pueden haber aparecido en la naturaleza los virus actuales, ya que, en el pasado, algunos de los virus actuales podrían haber sido considerados emergentes.

PANORAMA DE LOS VIRUS CON GENOMA DE RNA DE SIMPLE CADENA Y POLARIDAD NEGATIVA

Varias familias de virus están relacionadas entre sí por las características de su material genético. Todas ellas tienen el genoma

compuesto de RNA de simple cadena y estrategia de codificación con polaridad negativa (excepto los géneros Uukuvirus y Flebovirus, dentro de la familia *Bunyaviridae*, y la familia *Arenaviridae* que son ambisentido).

Además, *Rhabdoviridae*, *Filoviridae* y *Paramyxoviridae* tienen un genoma no segmentado, mientras que *Ortomyxoviridae*, *Bunyaviridae* y *Arenaviridae* tienen un genoma segmentado.

También debe considerarse que todos son virus envueltos y, aunque el número de genes y productos proteicos es variable, se encuentran muchas similitudes en las funciones de estos últimos.

Las Tablas 1.a y 1.b resumen las características más importantes de este grupo de virus.

Tabla 1.a. Replicación/transcripción

	Replicación transcripción	Modificaciones en los mRNAs	Ejemplos
<i>Rhabdoviridae</i>	por leader	poli A 5' cap metilado	VSV Rabia
<i>Filoviridae</i>	por leader	?	Ebola Marburg
<i>Paramyxoviridae</i>	por leader editing	poli A	Sarampión Sendai
<i>Ortomyxoviridae</i>	cap snatch	5' cap poli A	Influenza
<i>Bunyaviridae</i>	cap snatch	5' cap	La Crosse TSWV Hantaan
<i>Arenaviridae</i>	cap snatch	5' cap	LCM Junin Tacaribe

Tabla 1.b. Proteínas principales

	Proteínas del complejo RNP-nucleocápside			Glicoproteínas de la membrana				Otras proteínas	
	RNA polimerasa	Proteína no estructural	Proteína de la nucleocápside	Neuraminidasa	Hemaglutinina	Proteína de fusión	Glicoproteína	Proteína de matriz	Proteínas de función desconocida
<i>Rhabdoviridae</i>	L (100)	NS (20)	N (40)	---	---	---	GP (45)	M (20)	---
<i>Filoviridae</i>	L (180)	VP35? (32-35)	NP (100) VP30 (28)	---	---	---	GP (140)	VP24 (24)	VP40 (35-40)
<i>Paramyxoviridae</i>	L (255)	P (67,5)	NP (57,8)	HN (64,3)	HN (64,3)	F (60)	---	M (39)	C (23) D (16)
<i>Ortomyxoviridae</i>	PB1 (96) PB2 (87) PA (85)	NS ₁ (26)	NP (50-60)	NA (48-63)	HA ₁ (36-40) HA ₂ (28)	---	---	M ₁ (27)	NS ₂ (14) M ₂ (15) M ₃
<i>Bunyaviridae</i>	L (250)	NS _S (10) NS _M (15)	NP (26-40)	---	---	---	G1 (120) G2 (30)	---	---
<i>Arenaviridae</i>	L (250)	Z? (11)	N (60-64)	---	---	G2 (35)	G1 (44) G2 (35)	Z? (11)	---

Los números entre paréntesis corresponden a los pesos moleculares estimados de las proteínas

Capítulo 2 Arenavirus

HISTORIA Y TAXONOMIA

La primer infección humana producida por un arenavirus fue descrita durante 1925 por un médico sueco, Arvid Wallgren. Esta enfermedad, conocida en ese momento como meningitis aséptica aguda, es la actualmente denominada coriomeningitis linfocitaria. Sin embargo, el virus asociado con la enfermedad recién fue aislado una década después.

Varios grupos realizaron el descubrimiento, de manera independiente y casi simultáneamente. Armstrong & Lillie, publicaron la primera descripción en 1934. El virus, denominado LCM por virus de la coriomeningitis linfocitaria, fue aislado luego de repetidos pasajes en monos, a partir de un homogenado de cerebro humano de un aparente caso de encefalitis de Saint Louis. El origen del inóculo viral fue confuso ya que éste podría haber estado presente en el homogenado de cerebro humano o en alguno de los monos utilizados. Los otros dos grupos ayudaron a describir con precisión la naturaleza y la procedencia de nuevos inóculos virales de LCM (Rivers and Scott, 1935; Traub, 1935).

Durante varias décadas el virus LCM permaneció sin ser clasificado dentro de ninguna familia viral.

Entre 1956 y 1963, se aislaron, en diferentes regiones del globo, tres virus, denominados Tacaribe (Trinidad), Junín (Argentina) y Machupo (Bolivia). Rápidamente se encontraron similitudes entre ellos.

Debido al escaso conocimiento de la época, y al origen de algunos de los aislamientos, inicialmente fueron ubicados dentro del grupo Arbovirus. Estos virus denominados así por *arthropod-borne-viruses*, incluyen virus de varias familias, entre ellas las familias *Reoviridae*, *Flaviviridae*, *Togaviridae*,

Bunyaviridae y *Rhabdoviridae*, y se replican dentro de artrópodos hematófagos, transmitiéndose a sus hospedadores vertebrados mediante picaduras. No obstante, nunca pudo probarse que los ácaros fueran vectores de las enfermedades producidas por los virus antes mencionados, y se supone que la detección inicial fue casual y debida a los hábitos hematófagos de los artrópodos ensayados.

Posteriormente, en 1970, en base a la caracterización serológica, bioquímica y morfológica de estos virus, se creó la familia *Arenaviridae* con el virus LCM como prototipo y los virus Tacaribe, Junín y Machupo conformando el complejo Tacaribe (Rowe *et al.*, 1970b; Johnson *et al.*, 1973; Pfau *et al.*, 1974).

Esta familia está formada hasta el presente por catorce virus aceptados por los comités internacionales de taxonomía y varios de reciente descubrimiento (Salas *et al.*, 1991; Coimbra *et al.*, 1994; M. Sabattini, comunicación personal). El nombre Arenavirus deriva del latín *arenosus* (arena) en base al aspecto granuloso que presentan los viriones al microscopio electrónico, debido a la presencia de ribosomas en su interior.

Durante los últimos años se obtuvieron datos bioquímicos sobre dos nuevos aislamientos de arenavirus que presentan suficientes diferencias con otros miembros de la familia *Arenaviridae* como para ser considerados dos nuevos virus. Estos virus, encontrados en Brasil y en Argentina, fueron denominados Sabiá y Oliveros, respectivamente (Coimbra *et al.*, 1994; R. Rico-Hesse and S. Nichol, comunicación personal).

Finalmente, a partir de reacciones serológicas y considerando relaciones estructurales entre los distintos virus del grupo, se estableció una división entre Arenavirus del

viejo mundo, aislados en Africa y Europa, y Arenavirus del nuevo mundo, aislados en el continente americano (Dalton *et al.*, 1968; Murphy *et al.*, 1969; Murphy *et al.*, 1970; Rowe *et al.*, 1970a; Pfau, 1974; Pfau *et al.*, 1974; Murphy and Withfield, 1973, 1975b; Rawls and Buchmeier, 1976; Buchmeier *et al.*, 1981). El criterio para esta agrupación fue confirmado a partir de los datos obtenidos con el uso de anticuerpos monoclonales (Ruo *et al.*, 1991) y de biología molecular (Buchmeier and Oldstone, 1981; Compans *et al.*, 1981; Compans and Bishop, 1985; Ghiringhelli *et al.*, 1991; Clegg, 1993).

El grupo de arenavirus del viejo mundo incluye a los virus Ippy, Lassa, LCM, Mobala y Mopeia (McCormick, 1990; Salas *et al.*, 1991), de los cuales sólo Lassa y LCM han sido aislados de seres humanos.

El grupo de arenavirus del nuevo mundo comprende a los virus Junín, Amapari, Flexal, Guanarito, Latino, Machupo, Paraná, Pichindé, Tacaribe, Tamiami y, los recientemente descubiertos, Sabiá y Oliveros. De todos ellos, sólo Junín, Machupo, Guanarito y Sabiá han sido aislados de seres humanos.

En general, se conoce que, además de Junín, los virus Lassa, Machupo y Guanarito son patógenos para los humanos, produciendo la fiebre de Lassa, la fiebre hemorrágica boliviana (FHB) y la fiebre hemorrágica venezolana (FHV), respectivamente. El virus Guanarito, recientemente descubierto, fue aislado de humanos durante una epidemia (en principio confundida con dengue hemorrágico) ocurrida en Venezuela.

La fisiopatología de las enfermedades producidas por los virus Junín y Machupo son muy similares y, aunque aún no ha sido extensamente estudiada, la FHV también posee una sintomatología similar.

Por otra parte, LCM, detectado en ratones y humanos (Lepine *et al.*, 1937; Rivers and Scott, 1935; Armstrong and Sweet, 1939),

es el agente etiológico de la coriomeningitis linfocitaria y, aunque se lo ha asociado a meningitis aséptica, produce en el hombre infecciones inaparentes en la mayoría de los casos. Este virus ha sido una herramienta de extrema importancia en el descubrimiento de los mecanismos inmunológicos relacionados con la infección por arenavirus (Oldstone, 1987a y 1987b).

Con excepción de LCM, que tiene una distribución cosmopolita, todos los demás arenavirus se encuentran en áreas restringidas del globo, donde constituyen endemias. De hecho, el virus LCM ha sido aislado también en el área endémica de la FHA (Maiztegui *et al.*, 1972; Sabattini, 1977). En la última década, en el área endémica de FHA se han realizado varios aislamientos de nuevos arenavirus. Uno de ellos, Oliveros, ha sido caracterizado a nivel molecular en el CDC de Atlanta (Bowen *et al.*, 1996). El otro aislamiento ha sido denominado provisionalmente Pampa, y no se descarta que sea una variedad del anterior.

ECOLOGIA Y EPIDEMIOLOGIA

Los arenavirus son agentes zoonóticos que se mantienen en la naturaleza exclusivamente en las poblaciones de su hospedadores roedores, con la excepción de Tacaribe, que fue aislado de *Artibeus* spp. (un murciélago frugívoro) (Tabla 2.a).

Las áreas endémicas de muchos de los arenavirus aislados se solapan parcial o totalmente con las áreas geográficas ocupadas por sus hospedadores primarios. Debido a esto, se supone que la diseminación de un arenavirus ocurriría en paralelo con la dispersión geográfica de su hospedador. Probablemente no existan límites precisos para las áreas endémicas, salvo debido a la existencia de barreras geográficas importantes que limiten la distribución del reservorio natural.

Tabla 2.a. Arenavirus

Virus	Hospedador principal	Localización
Viejo Mundo		
LCM	<i>Mus musculus</i>	cosmopolita
Lassa	<i>Mastomys huberti</i>	Africa oeste
Mopeia	<i>Mastomys natalensis</i>	Africa sur
Mobala	<i>Praomys</i> spp.	Africa central
Ippy	<i>Arvicanthus</i> spp.	Africa central
Nuevo Mundo		
Junín	<i>Calomys musculus</i> <i>Calomys laucha</i> <i>Akodon azarae</i> <i>Necromys benefactus</i>	Argentina
Machupo	<i>Calomys callosus</i>	Bolivia
Guanarito	<i>Zygodontomys breviceauda</i>	Venezuela
Tacaribe	<i>Artibeus</i> spp. *	Trinidad
Amapari	<i>Oryzomys goeldii</i> <i>Neacomys guianae</i>	Brasil
Flexal	<i>Oryzomys</i> spp.	Brasil
Pichindé	<i>Oryzomys albigularis</i> <i>Thomasomys fuscatus</i>	Columbia
Latino	<i>Calomys callosus</i>	Bolivia Brasil
Paraná	<i>Oryzomys buccinatus</i>	Paraguay
Tamiami	<i>Sigmodon hispidus</i>	Florida, USA
Sabiá	?	Brasil
Oliveros	<i>Necromys benefactus</i>	Argentina

* murciélago frugívoro

depende fundamentalmente de su capacidad para establecer infecciones crónicas en los mismos. Durante toda la vida del animal infectado crónicamente hay una producción viral importante, que puede ser liberada al medio a través de fluidos corporales como la saliva y la sangre y también de las excretas, orina y excrementos.

El modo de transmisión del virus dentro de la población de hospedadores ha sido estudiado por muchos autores en distintos sistemas animales, principalmente, con los arenavirus LCM, Lassa, Junín, Pichindé y Machupo.

En este último virus, se han identificado dos tipos diferentes de respuesta a la infección por parte del roedor (dependientes de diferencias genotípicas), lo cual hace que en la población de roedores coexistan tres estados, roedores no infectados, roedores de tipo A (con viremia crónica y con fertilidad reducida) y roedores de tipo B (inmunocompetentes y con fertilidad normal). También se ha comprobado que la transmisión viral entre individuos de la población puede realizarse tanto de forma horizontal como vertical (Figura 2.1).

La densidad de las poblaciones de *Calomys* fluctúa cíclicamente en función de la proporción de virus que exista en el medio ambiente y que favorecen la generación de roedores de tipo A o tipo B. La relativa infertilidad de los roedores de tipo A infectados crónicamente se traduce en una progresiva disminución de este genotipo en la población y, consecuentemente, de la proporción de virus existente en el medio ambiente. Probablemente, los *Calomys* de tipo A poseen alguna ventaja competitiva sobre los de tipo B en ausencia del virus Machupo, lo cual lleva a la recuperación de la población con este genotipo y posibilita el comienzo de un nuevo ciclo (Johnson *et al.*, 1965; Justines and Johnson, 1969; Johnson *et al.*, 1973; Peters *et al.*, 1974; Webb *et al.*, 1975; Murphy and Walker, 1975; Peters *et al.*, 1987; Childs and Peters, 1993).

El ejemplo descrito anteriormente es, probablemente, una de las razones de los ciclos de epidemia que se observan en casi todos los arenavirus endémicos que afectan al ser humano.

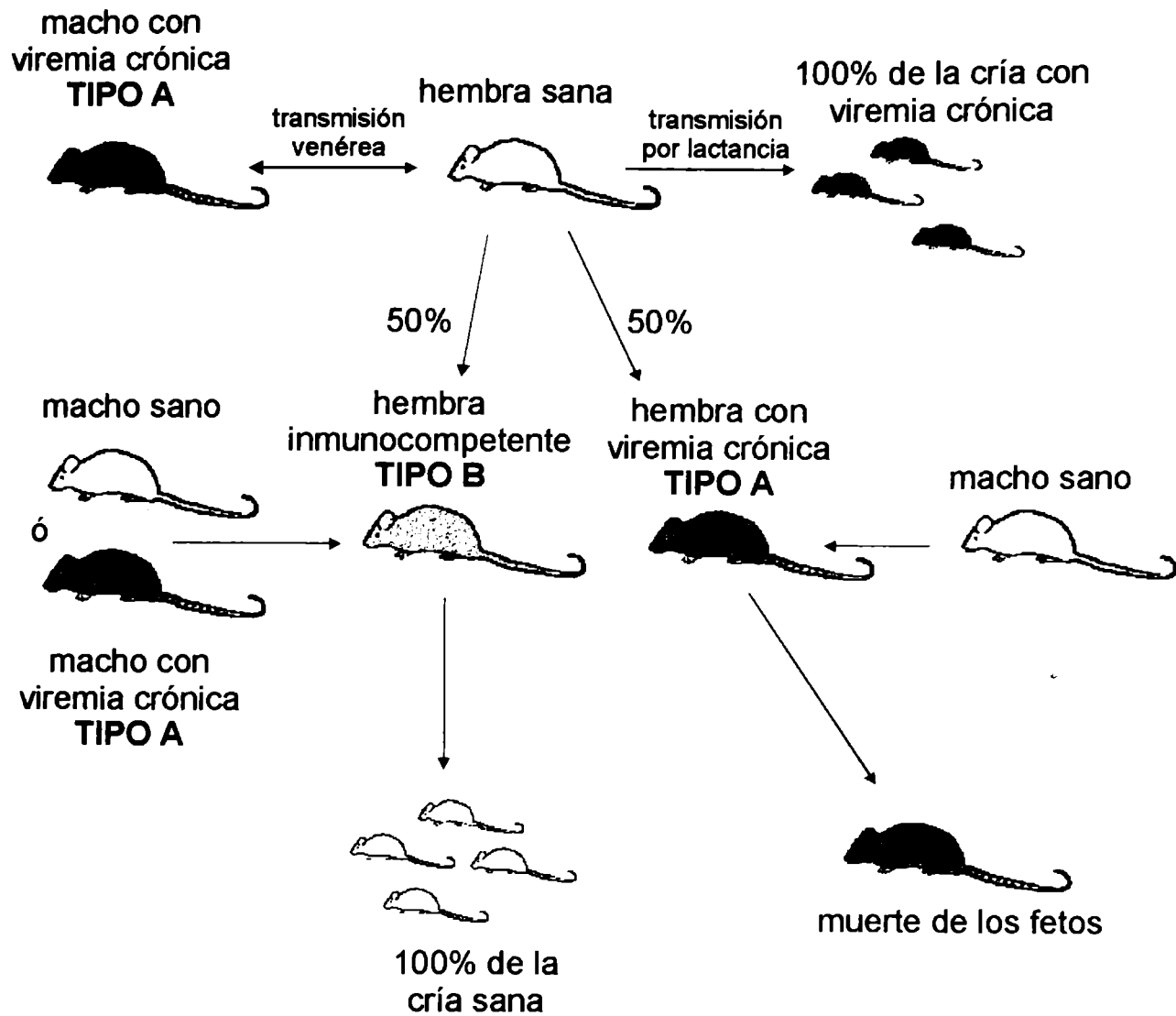


Figura 2.1. Modos de transmisión y desarrollo de la infección del virus Machupo en *Calomys callosus*. Se muestran los tres estados: roedores no infectados, roedores de **TIPO A** (con viremia crónica y con fertilidad reducida) y roedores de **TIPO B** (inmunocompetentes y con fertilidad normal). Las hembras recientemente infectadas mediante transmisión horizontal, producen 100% de la cría de **TIPO A** mediante transmisión vertical durante la lactancia. La misma hembra, si sobrevive a la infección, tiene el 50% de probabilidad de desarrollar una respuesta inmune completa y, por tanto, pertenecer al **TIPO B**, o de desarrollar una infección crónica y, por tanto, pertenecer al **TIPO A**. Las hembras de **TIPO B**, inmunocompetentes, dan siempre origen a progenie sana. Las hembras de **TIPO A**, crónicamente infectadas, generalmente no tienen progenie, ya que los fetos mueren durante el desarrollo.

Capítulo 3

Biología de los arenavirus

CARACTERÍSTICAS DE LOS VIRIONES

Ultraestructura

Los viriones de la familia *Arenaviridae* son partículas de forma variable, con un diámetro aproximado de 50 a 300 nm (**Figura 3.1**) (Dalton *et al.*, 1968; Gschwender *et al.*, 1975; Murphy *et al.*, 1969, 1970; Murphy and Whitfield, 1975; Ofodile *et al.*, 1973; Speir *et al.*, 1970). En general, las partículas pequeñas a medianas son de forma esférica, mientras que las más grandes son pleomórficas (**Figura 3.2**). En el virus Junín, los viriones son redondos u ovals con un diámetro promedio de 110-130 nm (Lascano and Berría, 1970, 1974). A partir de datos obtenidos de experimentos de ultracentrifugación se ha calculado un coeficiente de sedimentación de aproximadamente 500 S para las partículas virales de varios arenavirus.

Para el virus Junín suficientemente purificado, la densidad de flotación en CsCl fue medida en 1,20 g/ml (Romanowski, 1981). Otros autores han obtenido valores desde 1,12 g/ml hasta 1,17 g/ml, dependiendo del grado de purificación de los viriones, y de la composición de la solución utilizada durante la centrifugación de los viriones para calcular su densidad de flotación (Martínez Segovia and Díaz, 1968; Help *et al.*, 1970; Coto *et al.*, 1972; Johnson *et al.*, 1973; Añón *et al.*, 1976).

En las microfotografías electrónicas, los viriones se observan recubiertos por una unidad de membrana electrodensa que contiene proyecciones externas en su superficie. Estas proyecciones aparecen dispuestas de manera dispersa en la envoltura membranosa y, cuando se observan preparaciones por tinción negativa, tienen estructura claviforme (Murphy *et al.*, 1970). El interior de los viriones aparece

constituido por una materia amorfa compuesta principalmente por citoplasma celular y ribosomas, que están presentes en cantidad variable, como gránulos electrodensos de 20-25 nm de diámetro (**Figuras 3.1 y 3.2**). Su origen celular fue demostrado tanto por métodos genéticos como bioquímicos (Pedersen, 1970, 1971, 1973; Pedersen and Königshofer, 1976; Carter *et al.*, 1973; Farber and Rawls, 1975; Añón *et al.*, 1976; Vezza *et al.*, 1977, 1978) y su funcionalidad fue demostrada en sistemas de traducción *in vitro* (Chinault *et al.*, 1981; Giménez and Compans, 1980; Buchmeier *et al.*, 1981).

La función de estos ribosomas en el virión es desconocida, ya que, cuando la síntesis del rRNA en células infectadas es previamente inhibida por actinomicina D, la infectividad de los viriones, encapsidados sin ribosomas, no se ve afectada (Carter *et al.*, 1973) y, en el arenavirus Pichindé, se descartó su participación en los mecanismos de transcripción/replicación viral aprovechando una línea celular mutante con ribosomas *ts* (temperatura sensible) (Leung and Rawls, 1977). Por ello, se supone que los ribosomas son incluidos dentro del virión debido a un mecanismo de encapsidación poco específico. Sin embargo, el descubrimiento de mRNAs dentro del virión (Iapalucci *et al.*, 1989b; Salvato and Shimomaye, 1989), mantiene abierta la discusión. El hecho de que en los virus Tacaribe y LCM se encuentre uno de los mRNAs dentro de la partícula viral podría llevar a pensar que se estaría en presencia de un sistema de traducción detenido y, por lo tanto, existiría una asociación con los ribosomas. En el momento de la infección, este sistema se activaría utilizando factores celulares, constituyendo así, un evento muy temprano

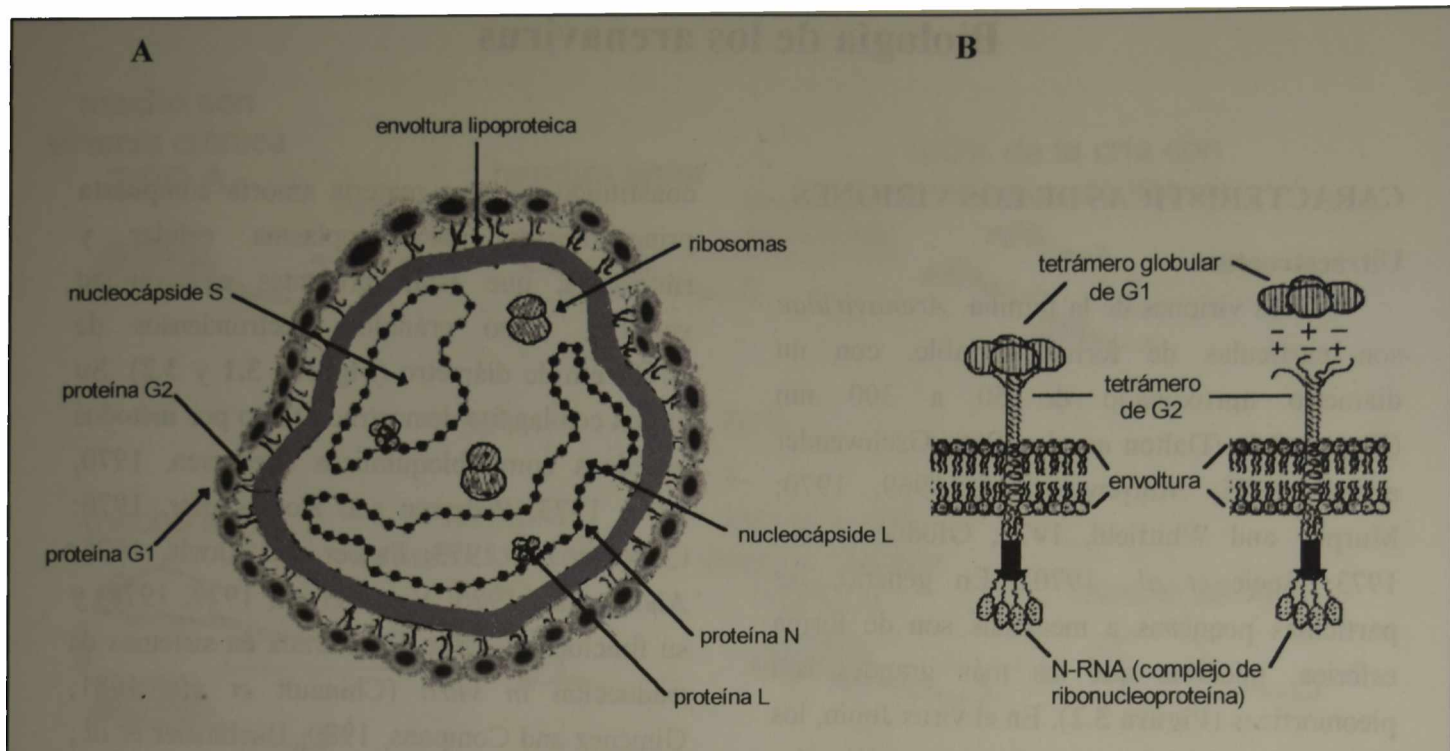


Figura 3.1. Esquema de una partícula viral de arenavirus. A. Las partículas de los arenavirus son pleomórficas, envueltas en una bicapa lipoproteica, con las glicoproteínas virales (G1 y G2) asociadas a la misma y componentes internos que incluyen dos nucleocápsides circulares de aspecto de cuentas de collar y ribosomas. Las nucleocápsides están constituidas por RNA (S o L) y proteínas, una mayoritaria (N) y otras minoritarias (L y Z). B: Modelo de la interacción entre los componentes mayoritarios del virión. G2 forma un complejo tetramérico, que posee tres dominios. El dominio transmembrana de unos 18 aminoácidos, conecta el dominio C-terminal interno (asociado con la nucleocápside) con el dominio N-terminal externo que forma un bastón asociado con un complejo tetramérico de G1 mediante interacciones iónicas.

dentro de los fenómenos de transcripción y replicación virales.

Es muy difícil observar otros detalles estructurales del interior de los viriones, sin embargo, después de la ruptura de la envoltura viral, se liberan las nucleocápsides (complejos de ribonucleoproteínas), que se ven como estructuras elongadas, similares a hebras, de 9-15 nm de diámetro (Veza *et al.*, 1977). En preparaciones teñidas negativamente con fosfotungstato de sodio (Gard *et al.*, 1977) las nucleocápsides tienen un diámetro de 3-4 nm y apariencia de collar de perlas, con estructuras globulares espaciadas con una periodicidad de 5 nm. En los virus Tacaribe y Tamiami la densidad de flotación en CsCl de estas estructuras es de 1,31-1,36 g/ml (Palmer *et al.*, 1977; Gard *et al.*, 1977).

En muchas preparaciones, las nucleocápsides purificadas aparecen como

estructuras circulares cerradas (Figura 3.3) correspondientes a una de las dos clases de tamaño predominantes, aproximadamente 640 nm y 1300 nm de longitud. La disposición circular de las nucleocápsides ha sido atribuida a la complementariedad de secuencias de los extremos 5' y 3' de los RNAs genómicos (Young and Howard, 1983; Auperin *et al.*, 1984b; Romanowski *et al.*, 1985; Bishop and Auperin, 1987). Aún cuando no se ha probado experimentalmente la existencia de estas estructuras en los arenavirus, en el caso de los RNAs genómicos del virus influenza (Hsu *et al.*, 1987) y el bunyavirus La Crosse (Raju and Kolakofsky, 1989), que contienen secuencias complementarias en los extremos 5' y 3', se efectuaron experimentos de entrecruzamiento químico que permitieron comprobar la existencia de estructuras similares a las propuestas para las nucleocápsides de los arenavirus.

Estructura molecular

Los arenavirus son virus envueltos, con una bicapa lipoproteica de origen celular y un genoma constituido por dos especies de RNA que codifican para las proteínas virales (Martínez Segovia and Grazioli, 1969; Añón *et al.*, 1976; Riviere *et al.*, 1985a) (Figura 3.1). Estos RNAs son de cadena simple y poseen una longitud aproximada de 7 kb (RNA L) y de 3,5 kb (RNA S).

En el momento de la brotación, el virus queda rodeado por la envoltura lipoproteica. En varios miembros de la familia *Arenaviridae* se ha estudiado la composición proteica y lipídica de la envoltura. Las proteínas celulares, que están presentes en escasa proporción, resultan excluidas por grupos de las glicoproteínas virales G1 y G2. La proteína G1 se encuentra formando tetrámeros globulares en el exterior del virión, unida por enlaces no covalentes a G2. Esta es una proteína tetramérica transmembrana que, en el interior del virión, se conecta con la nucleocápside viral. De esta manera, G1 y G2 constituirían las típicas espículas observadas por

microscopía electrónica en la cubierta viral. Respecto de los lípidos, se ha encontrado que la composición lipídica cuantitativa de la envoltura viral y de la membrana plasmática de las células infectadas por el virus son diferentes (Rosas, 1984; Rosas *et al.*, 1988). Esta relación entre la brotación del virus y la membrana plasmática de la célula hospedadora ha dado lugar a diversos estudios acerca de la influencia de la infección viral en los procesos mediados por membranas tales como la variación de la permeabilidad a iones y la disminución de la actividad de la bomba de Na^+/K^+ (sensible a ouabaina) y de la bomba de Ca^{+2} (Rey *et al.*, 1988; Rey, 1990).

Localizadas en el interior del virión se encuentran ribonucleoproteínas de origen celular (ribosomas) y viral (nucleocápsides). Las nucleocápsides están formadas por la proteína N asociada a los RNAs virales. La proteína N ha sido identificada como un polipéptido mayoritario de 60-68 kDa (Martínez Segovia and De Mitri, 1977; Gard *et al.*, 1977; Grau *et al.*, 1981; Romanowski, 1981; González *et al.*, 1983; Clegg and Lloyd, 1983). Además, se

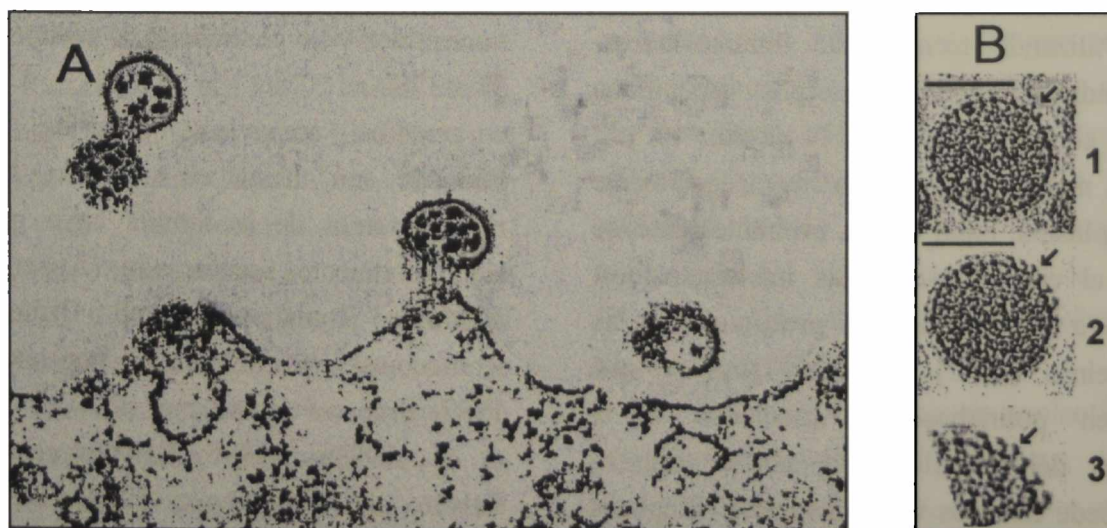


Figura 3.2. Micrografías electrónicas de arenavirus. A. Secciones de células BHK-21 infectadas donde se ven viriones libres y el proceso de brotación. Se observan partículas virales de 110 nm conteniendo gránulos electrondensos de 20 nm. B.1. Detalle de la envoltura viral (tomada de Burns and Buchmeier, 1993). La barra representa 100 nm. B.2. La misma imagen procesada por computadora para diferenciar las espículas en forma de clavija del fondo que rodea al virión. B.3. Sección de la imagen 2 aumentada y con mayor procesamiento. En los tres casos se indica con una flecha una espícula de glicoproteínas virales.

encuentran otras dos proteínas virales, la RNA polimerasa (replicasa), llamada proteína L, de aproximadamente 200 kDa y una proteína de 11 kDa, denominada Z por contener motivos estructurales postulados de tipo *zinc finger*, que aún no tiene una función biológica asignada.

Cada RNA genómico viral posee dos marcos de lectura abiertos, no superpuestos y de polaridades opuestas (genoma ambisentido, Auperin *et al.*, 1984b; Romanowski, 1993). En el RNA L se encuentran los marcos de lectura para la RNA polimerasa viral y para la proteína Z. En el RNA S se encuentran los marcos de lectura para las proteínas estructurales más importantes del virión, la proteína N y el precursor de las glicoproteínas de cubierta (GPC), que es procesado proteolíticamente durante el ciclo replicativo viral, para producir las dos glicoproteínas de la envoltura del virión (G1 y G2).

Morfogénesis

La brotación a nivel de la membrana plasmática de la célula infectada es la única etapa de la morfogénesis viral que ha sido observada por microscopía electrónica (Dalton *et al.*, 1968; Murphy *et al.*, 1970).

Utilizando técnicas de inmunofluorescencia indirecta con anticuerpos específicos dirigidos contra la proteína N (Zeller *et al.*, 1988), la misma es detectada mayoritariamente en el citoplasma, lugar donde, probablemente, se produce el ensamblaje de las nucleocápsides. Respecto a la síntesis del precursor de las glicoproteínas (GPC) es lógico suponer que ocurre en polirribosomas asociados a la membrana del retículo endoplásmico rugoso, como sucede en otros virus envueltos (Stephens & Compans, 1988). En el lumen del retículo ocurre el procesamiento proteolítico que da origen a las glicoproteínas virales G1 y G2, las cuales son transportadas al aparato de Golgi, donde tienen lugar los eventos asociados a la glicosilación. El movimiento final de las

glicoproteínas hacia la membrana plasmática es mediado, probablemente, por un proceso de transporte vesicular.

Por algún mecanismo desconocido, las glicoproteínas G1 (externa) y G2 (transmembrana) se disponen en la membrana plasmática en grupos (*patches*), excluyendo a las proteínas de membrana celulares. Las nucleocápsides podrían interactuar con el dominio citoplásmico de G2, con (Salvato *et al.*, 1992) o sin (Bruns *et al.*, 1986) la intervención de la proteína Z. Una vez concretadas estas interacciones moleculares, se produce la brotación de las partículas virales.

ESTRUCTURA Y PROPIEDADES DEL GENOMA

Genoma viral

Los Arenavirus poseen un genoma segmentado, formado por dos tipos de RNA de cadena simple. Estas especies de RNA son de tamaños diferentes. El RNA menor (denominado S, *small*), tiene alrededor de 3500 nucleótidos de longitud y un coeficiente de sedimentación de 22 a 25S, mientras que el mayor (denominado L, *large*), tiene aproximadamente 7000-7500 nucleótidos y un coeficiente de sedimentación de 31 a 37S.

Los arenavirus Tacaribe, LCM y Pichindé son, hasta el momento, los únicos representantes de la familia cuyo genoma ha sido totalmente secuenciado (Auperin *et al.*, 1984a,b; Romanowski and Bishop, 1985; Romanowski *et al.*, 1985; Bishop and Auperin, 1987; Franze-Fernández *et al.*, 1987, Iapalucci *et al.*, 1989a, 1989b; Salvato *et al.*, 1989; Salvato and Shimomaye, 1989). La secuencia nucleotídica completa del RNA S fue determinada, además, para los virus Lassa (Clegg and Oram, 1985; Auperin *et al.*, 1986; Auperin and McCormick, 1989; Clegg *et al.*, 1990), Junín (Ghiringhelli *et al.*, 1987, 1989, 1991a, 1991b, 1997), Mopeia (Wilson and

Clegg, 1991) y Oliveros (Bowen, *et al.*, 1996), así como la mitad 3' del RNA S del virus Machupo (Griffiths *et al.*, 1992).

Aunque para asegurar la infectividad, los viriones deben poseer, al menos, una copia completa de cada especie de RNA, la proporción de cada uno de los RNAs no es equimolar en distintas preparaciones de viriones. Es bastante común encontrar que el número de moléculas del RNA S sea mayor al número de moléculas del RNA L. De hecho, se ha comprobado experimentalmente mediante métodos genéticos y bioquímicos la existencia de relaciones molares RNA S/RNA L de 1,5 en Junín (Añón *et al.*, 1976), 2,9-5,0 en LCM (Pedersen, 1973; Romanowski and Bishop, 1983) y 1,8-2,0 en Pichindé (Carter *et al.*, 1973; Farber and Rawls, 1975; Dutko *et al.*, 1976; Vezza *et al.*, 1978), lo cual deriva en la existencia de partículas virales diploides para el RNA S.

En todas las preparaciones virales se encuentran otras especies de RNA que acompañan al genoma. Estos RNAs, inicialmente caracterizados por su coeficiente de sedimentación, 4-6S, 18S y 28S, corresponden a rRNAs originarios de los ribosomas presentes en el virión (Abelson *et al.*, 1969; Pedersen, 1970; Carter *et al.*, 1973; Añón *et al.*, 1976; Vezza *et al.*, 1977; Ramsingh *et al.*, 1980; Dutko *et al.*, 1981). Como ya se indicó, estos ribosomas no parecen tener ninguna función en el ciclo replicativo viral y su presencia dentro de los virus, podría deberse a un sistema de encapsidación y brotación poco específico. Efectivamente, los arenavirus son capaces de encapsidar también algunas de las especies de RNAs utilizadas como intermediarios en la replicación viral, tales como la forma antígenómica del RNA S, aunque su proporción no supera el 5% de la masa del RNA S genómico. Además, en algunas preparaciones del virus Pichindé, se ha detectado por *Northern blot* un RNA 15S, que según algunos autores (Farber and Rawls, 1975; Dutko *et al.*, 1976;

Auperin *et al.*, 1984a; Bishop, 1990), sería el mRNA de la proteína N, aunque esto aún no ha sido comprobado experimentalmente. También, se ha descrito la presencia de un RNA de bajo peso molecular en preparaciones de viriones de los virus Tacaribe y LCM, en cantidades

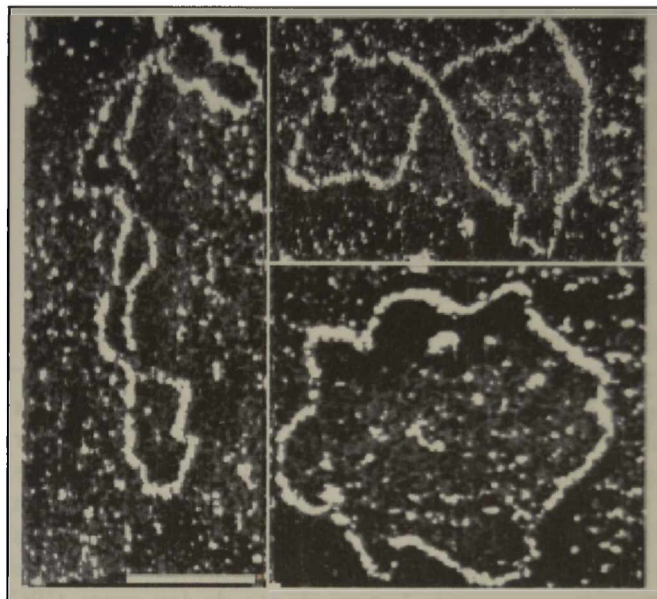


Figura 3.3. Micrografías electrónicas de nucleocápsides purificadas. Se observan tres nucleocápsides aisladas en diferentes estados de enrollamiento. La barra blanca representa 200 nm (tomada de Young and Howard, 1983).

molares comparables al RNA S viral. La caracterización del mismo permitió comprobar que corresponde al mRNA transcripto del gen Z; es probable que este RNA se encuentre asociado a los ribosomas en la partícula viral y pueda ser traducido en una etapa muy temprana de la replicación del virus en una célula infectada (Iapalucci *et al.*, 1989a; Salvato and Shimomaye, 1989).

Los datos de secuencia nucleotídica viral han permitido ubicar dos marcos de lectura abiertos en cada RNA y, en ambos casos, los mismos son de sentido opuesto, lo que resulta en una doble polaridad del RNA. Por ello, Auperin *et al.* (1984b) propusieron el nombre de *ambisense* en contraposición con los anteriormente descritos *positive* y *negative sense*, para describir la estrategia de

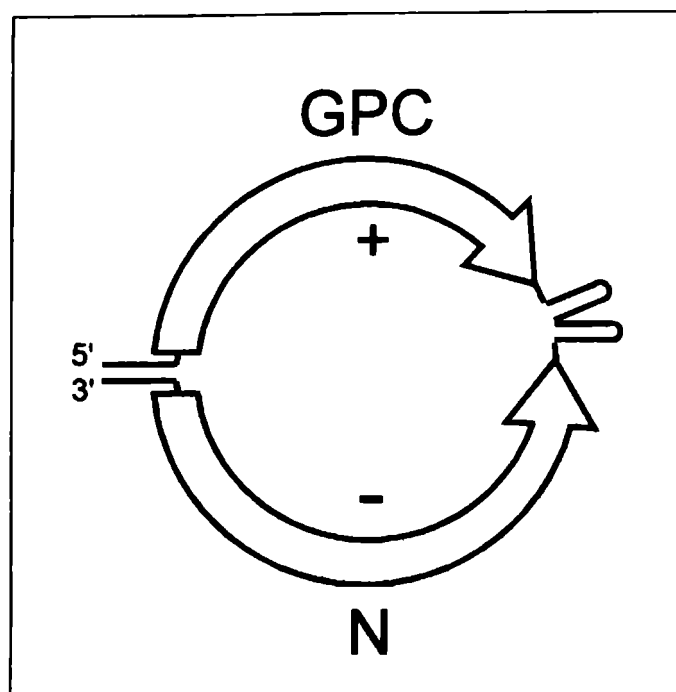


Figura 3.4. Organización molecular del RNA S del virus Junín. El esquema interpreta las observaciones por microscopía electrónica de nucleocápsides y las predicciones de los datos de secuencia nucleotídica. El RNA S se muestra en forma circular cerrada con los extremos 5' y 3' apareados en una estructura tipo *panhandle*. Con flechas abiertas se indican los marcos de lectura abierta de los genes GPC (polaridad positiva) y N (polaridad negativa). En la región intergénica se muestran las estructuras tipo horquilla (*hairpin loop*). Las estructuras doble cadena no están dibujadas a escala.

codificación de estos virus. La doble polaridad es una estrategia de codificación que también presentan algunos géneros de bunyavirus, una familia muy relacionada en este aspecto a los arenavirus (Elliot, 1990).

Estrategia de codificación

Las especies moleculares S y L codifican para distintas proteínas (Veza *et al.*, 1978; Leung *et al.*, 1979; 1981; Harnish *et al.*, 1983). El RNA S codifica para las proteínas de la nucleocápside viral (N) y el precursor de las glicoproteínas de superficie (GPC). El RNA L codifica para la replicasa viral (L, RNA polimerasa-RNA dependiente) y la proteína Z, sin función asignada hasta el momento (Figura 3.4). De esta manera, casi toda la capacidad codificante de los RNAs virales está aprovechada, presentando solamente regiones no

codificantes en los extremos 3' y 5' y en las regiones intergénicas (Figura 3.4).

Los mRNAs de los arenavirus se caracterizan por presentar regiones 5' no codificantes de longitud variable y poseer en el extremo nucleótidos extras no codificados en el molde (Garcín and Kolakofsky, 1990, 1992). Los codones de iniciación han sido identificados en su mayoría por comparación con la secuencia consenso para mRNAs eucarióticos, CC{A/G}CCAUGG, predicha por Kozak (1978, 1983, 1984).

En el extremo 3' de los RNAs genómicos S y L de los arenavirus que han sido secuenciados, se halla una secuencia de 19 nucleótidos común a todos ellos (3' GCGUGUCCUAGGAUCCG 5', Auperin *et al.*, 1982a, 1982b). Se supone que esta secuencia podría tener alguna importancia en el reconocimiento del molde por parte de la RNA polimerasa en el inicio de la replicación. Esta característica es conocida dentro de los togavirus (Ou *et al.*, 1981), paramixovirus (Blumberg *et al.*, 1981) y bunyavirus (Elliot, 1990) y, aparentemente, los filovirus poseen una estrategia similar (Kiley *et al.* 1986; Feldman *et al.*, 1993). Además, otra característica común importante de esta secuencia, es que las primeras bases del extremo 5' viral son complementarias a la misma. Cuando se analizan secuencias particulares, se observa que la complementariedad entre los extremos 5' y 3' de los RNAs S y L, se extiende más allá de la secuencia consenso en todos los arenavirus, y, de acuerdo a las energías libres calculadas, su formación es altamente probable. En este sentido, es posible que la estabilización de estas estructuras tipo *panhandle* sea la responsable de la estructura circular que se observa por microscopía electrónica en las nucleocápsides virales (Figura 3.3). Además, la formación de esta estructura durante la replicación del virus en la célula infectada podría ser la señal que marca el comienzo de la encapsidación viral.

Mecánica de la transcripción/replicación

Estudios recientes, utilizando agentes lisosomotrópicos, han demostrado que los virus Pichindé, Lassa y Mopeia penetran en la célula mediante lisosomas ácidos (Glushakova and Lukashevich, 1989; Harnish *et al.*, 1993). Aún cuando no se conocen con precisión los eventos moleculares que ocurren una vez que el virus ha entrado a la célula, el camino a seguir incluye, por un lado, la transcripción y traducción de las proteínas virales, la replicación del genoma viral, el ensamblado de la nucleocápside y, finalmente, la brotación de la partícula viral y, por otra parte, la interacción del virus con la célula hospedadora estableciendo infecciones persistentes o líticas, dependiendo de la línea celular.

Los mRNAs de las proteínas N y L son transcritos utilizando como molde la región 3' del RNA viral (de polaridad negativa) S o L, respectivamente. En cambio, ya que la región 5' viral es de polaridad positiva, los mRNAs de GPC y Z son transcritos utilizando como molde la región 3' del RNA viral complementario (antigenómico) S o L, respectivamente. El RNA viral complementario (antigenómico) de longitud completa es un intermediario obligado en la replicación. Hasta el presente no se ha podido probar que las regiones 5' de los RNAs, viral o viral complementario, de longitud completa, actúen directamente como mensajeros, lo que hace suponer que la interacción con los ribosomas no depende solamente de las secuencias codificadas.

La región intergénica no codificante tiene como característica la capacidad de formar una estructura secundaria muy estable en forma de horquilla cuya posible función sería la de actuar como terminador de la transcripción (Romanowski *et al.*, 1985; Franze Fernández *et al.*, 1987; Ghiringhelli *et al.*, 1991a; Iapalucci *et al.*, 1991; Rivera Pomar, 1991). En la mayoría de los arenavirus, esta estructura es simple pero, Ghiringhelli *et al.* (1987) comunicaron una

probable estructura de doble horquilla en la región intergénica del RNA S del virus Junín. Más recientemente, Wilson and Clegg (1991) describieron una estructura similar en el RNA S del virus Mopeia, en tanto que Iapalucci *et al.* (1991) reinterpretaban su versión original de la secuencia intergénica del virus Tacaribe considerándola una doble horquilla, del mismo tipo que la descrita para el virus Junín.

Los mecanismos de transcripción y replicación viral que se encuentran esquematizados en la **Figura 3.5**, han sido estudiados con mayor detalle para el RNA S. Se supone que el RNA L sigue una estrategia similar. Una vez que el virus infecta a una célula, la RNA polimerasa que ingresa con las nucleocápsides comienza la transcripción de la región 3'. Cuando la polimerasa alcanza las estructuras intergénicas en horquilla, se disocia del RNA molde, dando origen a una especie subgenómica que es el mRNA de la proteína N. El extremo 5' de este RNA posee un número variable de bases extra, cuya secuencia no se encuentra especificada por el molde viral. Este hecho llevó a proponer que la síntesis de los mRNAs comenzaría con un mecanismo del tipo de *cap-snatching* (Krug, 1981; Garcin and Kolakofsky, 1990). El extremo 3' de los mRNAs de N, en los virus Tacaribe (Iapalucci *et al.*, 1991) y Junín (Rivera Pomar *et al.*, 1993), contiene la copia de la primer estructura en horquilla de la región intergénica que encuentra la RNA polimerasa en la dirección de la transcripción. Si la RNA polimerasa no se detiene en la región intergénica, el resultado es una copia completa del RNA S. Esta nueva especie, de secuencia antigenómica, sirve como molde para la transcripción de las dos especies de RNA S de secuencia genómica: el mRNA de la proteína GPC y el RNA S genómico viral de longitud completa. En el virus Tacaribe, el mRNA de GPC también incluye en su extremo 3' una copia de la primer estructura en horquilla de la región intergénica que encuentra la RNA

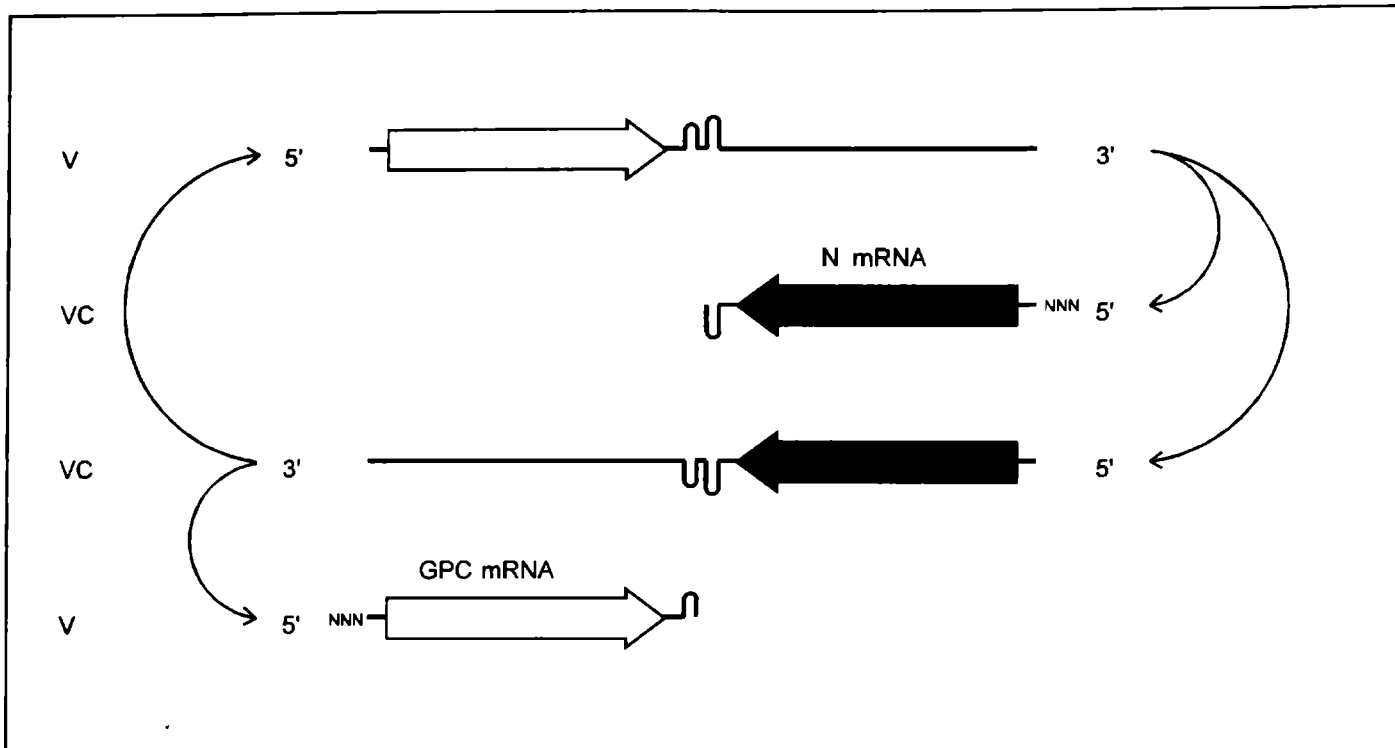


Figura 3.5. Transcripción/Replicación del RNA S del virus Junín. El RNA S genómico (o viral, V) del virus Junín contiene un marco de lectura para el gen GPC. Este RNA es transcrito a dos formas de RNAs antigenómicos (o viral complementarios, VC) que contienen el marco de lectura del gen N: el mRNA de N y el RNA S antigenómico. Los marcos de lectura de los genes GPC y N, se muestran como rectángulo vacío y lleno, respectivamente, con las puntas en flecha indicando la dirección de la transcripción y traducción. En la región intergénica se muestran las estructuras en horquilla (fuera de escala). El símbolo "NNN" indica los nucleótidos independientes del molde agregados a los extremos 5' de los mRNAs. Las flechas delgadas indican las relaciones molde/transcrito entre las diferentes especies de RNA.

polimerasa en la dirección de la transcripción (Iapalucci *et al.*, 1991). Por otra parte, si la síntesis proteica se encuentra inhibida, el mRNA de N es la única especie que puede ser sintetizada, producto de la transcripción del RNA S (Franze Fernández *et al.*, 1987).

La aparición de las otras especies de RNA virales es dependiente de la síntesis proteica, lo cual implica necesariamente que el mecanismo de antiterminación de la transcripción del RNA S depende de la síntesis proteica. Este efecto de antiterminación podría ser llevado a cabo por la proteína N, alguna otra proteína viral o algún polipéptido celular con una alta velocidad de recambio.

Experimentos en los que se estudiaron los eventos tempranos de la replicación y transcripción del RNA S del virus Junín (Rivera Pomar *et al.*, 1993) indican que la principal responsable de la acción antiterminadora *in vivo* es la proteína N.

PROTEINAS VIRALES

En las partículas virales purificadas de los distintos arenavirus sobre los cuales se ha trabajado hasta el momento, se han detectado tres proteínas mayoritarias (Ramos *et al.*, 1972; Gard *et al.*, 1977; Martínez Segovia and De Mitri, 1977; Vezza *et al.*, 1977; Buchmeier *et al.*, 1977, 1978, 1981; Buchmeier and Oldstone, 1978, 1979, 1981; Buchmeier and Parekh, 1987; Gangemi *et al.*, 1978; Saleh *et al.*, 1979; Cresta *et al.*, 1980; Compans *et al.*, 1981; Grau *et al.*, 1981; Harnish *et al.*, 1981a,b; Kiley *et al.*, 1981; Leung *et al.*, 1981; Boersma *et al.*, 1982; Bruns *et al.*, 1983a,b, 1986; Young *et al.*, 1981; Clegg and Lloyd, 1983). Estas parecen ser las proteínas estructurales más importantes del virión. Además, se han identificado otras dos proteínas, de características no estructurales que, resultan encapsidadas en proporciones variables dentro de la partícula viral (Iapalucci *et al.*, 1989a,b; Salvato and Shimomaye, 1989).

Estas proteínas son la proteína de la nucleocápside viral, N, las glicoproteínas estructurales G1 y G2 (originadas a partir del precursor de las glicoproteínas, GPC), y las proteínas L (identificada como la RNA polimerasa-RNA dependiente o replicasa) y Z (una proteína recientemente caracterizada, con un dominio postulado de tipo *zinc finger*).

Proteína de la nucleocápside (N)

La proteína N es el constituyente principal de la nucleocápside y es la más abundante del virión, representando un 70% de la masa proteica viral. En el virus Pichindé se ha estimado que hay unas 1500 moléculas de proteína N por virión (Veza *et al.*, 1977), la mayor parte de las cuales está asociada a los RNAs virales, por lo cual es indudable que la función principal de la proteína N es netamente estructural. El conocimiento de las secuencias aminoacídicas de diferentes Ns permite estimar una longitud promedio de 564 aminoácidos y un peso molecular promedio de aproximadamente 63 kDa (Tabla 3.a).

La responsabilidad o participación de la proteína N en otra u otras funciones virales no es muy clara hasta el momento.

El ensayo de fijación de complemento fue utilizado para establecer relaciones entre los distintos arenavirus (Rowe *et al.*, 1970a; Casals *et al.*, 1975). Estos estudios permitieron dividir a las especies de arenavirus en dos grupos antigénicamente relacionados, los arenavirus del Nuevo Mundo (o complejo Tacaribe) y los arenavirus del Viejo Mundo. Gschwender and Lehmann-Grube (1973) determinaron que la fijación de complemento estaba asociada con una fracción soluble obtenida a partir de células infectadas con LCM. Posteriormente, Buchmeier *et al.* (1977) y De Mitri & Martínez Segovia (1980) demostraron que la proteína N o productos de menor peso molecular, supuestamente generados por proteólisis intracelular son los antígenos responsables de la fijación de complemento.

Por otra parte, se ha mostrado que la proteína N de los virus Junín, Pichindé y Lassa puede sufrir un procesamiento o degradación produciendo un polipéptido de 25-30 kDa (Harnish *et al.*, 1981a; Young *et al.*, 1987; Sánchez *et al.*, 1989). Este producto puede liberarse por lavado de las células infectadas con el virus Junín y detectarse libre en el medio de cultivo sobrenadante, de donde puede suponerse asociado a la superficie celular. Por el contrario, en células infectadas con Pichindé la localización del mismo sería intranuclear (Young *et al.*, 1987).

Además, se ha descrito la existencia de, al menos, un epítopo de N en la superficie del virión de LCM, así como en la superficie de células infectadas (Bruns *et al.*, 1986; Zeller *et al.*, 1986, 1988). Esta observación está soportada, en teoría, por el hecho de que N presenta regiones hidrofóbicas que podrían anclarla en la membrana celular. La proteína N ha sido localizada en el citoplasma de las células infectadas ya sea por inmunomicroscopía óptica o electrónica. Young *et al.* (1987) describieron que la proteína N estaba presente en el núcleo de las células infectadas con Pichindé, aspecto no demostrado en los demás arenavirus.

Un análisis detallado de las secuencias aminoacídicas de las proteínas N permite reconocer la presencia de varios sitios potenciales de N-glicosilación y de fosforilación por proteinquinasas, caseinquinasas y tirosinquinasas. Todos los estudios sobre proteínas virales glicosiladas demuestran que estos sitios potenciales en la proteína N no son utilizados *in vivo*. En cambio, al menos en LCM, Tacaribe y Pichindé, la proteína N se encuentra en dos fracciones, como proteína fosforilada y no fosforilada (Boersma and Compans, 1985; Bruns *et al.*, 1986). Esta fosforilación depende de una actividad de proteinquinasa asociada al virión que utiliza como sustrato a la proteína N (Howard and

Buchmeier, 1983). Esta actividad aún no ha sido asignada a ninguna proteína del virión.

Experimentos recientes sobre la replicación y transcripción del RNA S de distintos arenavirus (Lee *et al.*, 2000; Tortorici *et al.*, 2001a; López *et al.*, 2001) muestran que la proteína N es la principal responsable de la acción antiterminadora *in vivo*.

Precursor de las glicoproteínas (GPC)

El conocimiento de la secuencia aminoacídica de varios precursores de las glicoproteínas permite estimar una longitud promedio de 494 aminoácidos y un peso molecular promedio aproximado de 56 kDa en su forma no glicosilada. (Tabla 3.b). Este precursor está constituido por tres regiones, un péptido señal, la glicoproteína G1 y la glicoproteína G2.

El procesamiento del precursor GPC ocurre en dos etapas (Burns and Buchmeier, 1993). El péptido señal es procesado previo al transporte de la glicoproteína desde el retículo endoplásmico. En cambio, el procesamiento proteolítico de GPC para dar G1 y G2, ocurre

entre el Golgi medio y el trans-Golgi (Wright *et al.*, 1990). Buchmeier *et al.* (1987) determinaron que el sitio de procesamiento proteolítico en el precursor de la glicoproteína de LCM se encuentra en las proximidades de la secuencia básica, Arg₂₆₂-Arg₂₆₃. Este sitio está seguido por una serie de 23 residuos hidrofóbicos. Esta secuencia característica se conserva en las dos cepas estudiadas de LCM (ARM y WE), Lassa (Auperin *et al.*, 1986) y Pichindé, y es reemplazada por la secuencia Arg-Lys en Junín (Ghiringhelli *et al.*, 1991) y Arg-Thr en Tacaribe (Franze Fernández *et al.*, 1987). La proteasa que procesa las glicoproteínas se ha confirmado como intracelular, no codificada por el genoma viral. El sitio de procesamiento proteolítico fue localizado exactamente a 3 aminoácidos de distancia de la primer Arg de la secuencia conservada en la mayoría de los arenavirus. En el virus Tacaribe, la falta de esta secuencia conservada (Franze-Fernández *et al.*, 1987) y la detección por electroforesis de una única glicoproteína de peso molecular de 44 kDa, llevó a pensar en una composición y procesamiento de

Tabla 3.a. Características de las proteínas N

Virus	Cepa	Longitud en aminoácidos	Peso molecular calculado (kDa)	Referencias
Junín	MC2	564	63027	Ghiringhelli <i>et al.</i> , 1989 Ghiringhelli <i>et al.</i> , 1991
Junín	Candid # 1	564	63123	Ghiringhelli, P.D., resultados no public.
Lassa	GA391	570	62913	Clegg and Oram, 1985 Clegg <i>et al.</i> , 1990
Lassa	Josiah	569	62882	Auperin and McCormick, 1988
LCM	Armstrong	558	62177	Salvato <i>et al.</i> , 1988
LCM	WE	558	62182	Romanowski <i>et al.</i> , 1985 Romanowski and Bishop, 1985
Machupo	AA288-77	564	63299	Griffiths <i>et al.</i> , 1992
Mopeia	800150	570	63434	Wilson <i>et al.</i> , 1991
Oliveros	-----	558	62154	Bowen <i>et al.</i> , 1996
Pichindé	3739	561	63012	Auperin <i>et al.</i> , 1984 Bishop and Auperin, 1987
Sabiá	SPH114202	562	63122	González <i>et al.</i> , 1996
Tacaribe	T.RVL.II 573	570	63789	Franze Fernández <i>et al.</i> , 1987

glicoproteínas diferente al resto de la familia. Sin embargo, mediante el secuenciamiento del extremo amino de las glicoproteínas estructurales del virus Tacaribe, se mostró que el procesamiento proteolítico del precursor GPC produce dos glicoproteínas G1 y G2 de pesos moleculares aparentes muy próximos, que no se alcanzan a resolver como bandas distintas por electroforesis en gel de poliacrilamida.

Las glicoproteínas, G1 y G2, derivadas del procesamiento del precursor GPC se hallan en cantidades equimolares en el virión y son proteínas de superficie, observables al microscopio electrónico como espículas claviformes de 5 a 10 nm en la membrana (Murphy *et al.*, 1970; Lascano and Berría, 1974; Murphy and Whitfield, 1975; Young *et al.*, 1981). G1 es una proteína periférica, mientras que G2 es transmembrana (Di Simone *et al.*, 1994). Estas proteínas probablemente están involucradas en la interacción virión-célula a través de un receptor aún no identificado.

Las dos glicoproteínas pueden ser removidas del virión por tratamientos con enzimas proteolíticas y detergentes no iónicos (Gard *et al.*, 1977, Vezza *et al.*, 1977; Buchmeier and Oldstone, 1978; Romanowski, 1981). Estudios de neutralización han mostrado que los anticuerpos monoclonales obtenidos

contra G1 son más efectivos que aquéllos contra G2 (Buchmeier *et al.*, 1981). Estos datos soportan la idea de que G1 es más externa. Para el virus Junín, Romanowski (1981) demostró el carácter extrínseco de G1 en viriones altamente purificados por iodación *in vitro*. En el caso del virus LCM, el tratamiento de los viriones con detergentes en ausencia de agentes reductores provocó la liberación de estructuras oligoméricas, constituidas únicamente por complejos de G1, formando como máximo tetrámeros (Wright *et al.*, 1989, Burns and Buchmeier, 1991). Los tratamientos con detergentes mostraron además, que G1 es una proteína periférica mientras que G2 es una proteína integral de membrana. La proteína G2 del virus LCM posee 233 aminoácidos separados en tres dominios. Los 164 aminoácidos del extremo amino terminal forman el dominio más importante, externo al virión, que es el que interacciona con G1. Los restantes 69 aminoácidos se dividen en dos dominios, uno de anclaje a la membrana, constituido por un grupo de 20 a 25 aminoácidos hidrofóbicos y el otro, de 44 a 49 aminoácidos, que conforman el dominio interno de G2, es el que interacciona con la nucleocápside viral.

La organización macromolecular de los complejos estructurales de las glicoproteínas del

Tabla 3.b. Características de las proteínas GPC

Virus	Cepa	Longitud en aminoácidos	Peso molecular calculado (kDa)	Referencias
Junín	MC2	481	55118	Ghiringhelli <i>et al.</i> , 1991
Junín	Candid # 1	485	55559	Ghiringhelli <i>et al.</i> , 1996
Lassa	GA391	491	55744	Clegg <i>et al.</i> , 1990
Lassa	Josiah	491	55752	Auperin <i>et al.</i> , 1986 Auperin and McCormick, 1988
LCM	Armstrong	498	56215	Salvato <i>et al.</i> , 1988
LCM	WE	498	56235	Romanowski <i>et al.</i> , 1985
Mopeia	800150	489	55987	Wilson <i>et al.</i> , 1991
Olivero	-----	518	59380	Bowen <i>et al.</i> , 1996
Pichindé	3739	503	57278	Auperin <i>et al.</i> , 1984 Bishop and Auperin, 1987
Sabiá	SPH114202	488	55877	González <i>et al.</i> , 1996
Tacaribe	T.RVL.II 573	483	55554	Franze Fernández <i>et al.</i> , 1987

virus LCM fue estudiada, además, con agentes químicos de entrecruzamiento. Mediante la utilización de agentes de entrecruzamiento impermeables a las membranas, se demostró que cada glicoproteína se encontraba formando complejos de homotetrámeros. Cuando los agentes utilizados podían atravesar las membranas biológicas se observó que G2 formaba, además, complejos con la proteína N (**Figura 3.1 B**). El extremo carboxilo terminal de G2 se encuentra cargado positivamente. En ese sentido, 5 de los 12 residuos del extremo carboxilo de la G2 de LCM son lisina o arginina, y es muy probable que estos residuos sean los responsables de la interacción con N (Burns and Buchmeier, 1991).

Estos estudios han establecido un modelo estructural para las glicoproteínas del virus LCM. Las espículas estarían formadas por una cabeza externa constituida por un tetrámero de G1 unido por puentes disulfuro, y por una parte integral a la membrana constituida por G2 (Burns *et al.*, 1990). Este modelo se corresponde con las espículas observadas en microscopía electrónica (**Figura 3.2**).

Respecto de la cadena de polisacáridos es poco lo que se conoce (Boersma *et al.*, 1982, Wright *et al.*, 1990, Scolaro *et al.*, 1990), salvo que medidas de incorporación de azúcares marcados radiactivamente han determinado un mayor grado de glicosilación de G1 (Ramos *et al.*, 1972; Vezza *et al.*, 1977; Buchmeier *et al.*, 1978) y que el precursor GPC es rico en manosa mientras que G1 y G2 son ricas en glucosamina, galactosa y fucosa (Wright *et al.*, 1990). Hasta ahora no se ha detectado la presencia de otros grupos en asociación covalente con las glicoproteínas.

Los análisis de biosíntesis de la glicoproteína muestran que después de la proteólisis enzimática en el Golgi medio se produce el procesamiento de oligosacáridos en el trans-Golgi produciendo G1 y G2 con 5 y 2 oligosacáridos complejos, respectivamente, ligados por asparragina (Wright *et al.*, 1990).

El estudio de las secuencias aminoacídicas derivadas de la información nucleotídica de distintos arenavirus ha demostrado diferentes grados de homología, regiones consenso y, en particular, la conservación de algunos de los sitios potenciales de N-glicosilación y de los posibles sitios de procesamiento proteolítico del precursor. Además, la glicoproteína, presenta como rasgo interesante, epitopes conservados con la proteína de la nucleocápside entre los arenavirus africanos entre sí, por una parte, y los del Continente Americano, por otra (Buchmeier and Parekh, 1987).

Los mapas peptídicos obtenidos por digestiones enzimáticas de las glicoproteínas de los virus Pichindé y LCM son diferentes (Buchmeier and Oldstone, 1979; Harnish *et al.*, 1981a, 1981b; 1983). En 1990, Scolaro *et al.* realizaron experimentos de mapeo de péptidos que indican que la glicoproteína de una mutante atenuada del virus Junín presenta modificaciones estructurales que la diferencian de la cepa parental más patógena. Estas diferencias pueden estar correlacionadas con diferencias en el tropismo viral y/o la interacción inicial con el potencial receptor celular.

RNA polimerasa viral (L)

La proteína L es un polipéptido no glicosilado (Harnish *et al.*, 1981b; 1983) con una longitud promedio de 2200 aminoácidos y un peso molecular promedio aproximado de 250 kDa (**Tabla 3.c**). El tamaño de esta proteína es muy similar al de la proteína L de virus con genoma de RNA de polaridad negativa (Schubert *et al.*, 1984; Shioda *et al.*, 1986; Yusoff *et al.*, 1987; Tordo *et al.*, 1988; Galinski *et al.*, 1988; Blumberg *et al.*, 1988).

Esta proteína se halla asociada a la nucleocápside viral formando un complejo ribonucleoproteico y la actividad que se le atribuye es la de RNA polimerasa-RNA dependiente ("replicasa") análoga a la proteína L de rhabdovirus y paramixovirus. Mutantes del

RNA L han mostrado capacidad de alterar la morfología de las placas de lisis y la virulencia y letalidad del virus LCM (Riviere *et al.*, 1985b). De esto se puede inferir una función en los mecanismos de replicación y establecimiento de infecciones persistentes (Singh *et al.*, 1987). Siendo la proteína L codificada por un gen de gran tamaño es un blanco fácil de mutaciones que alteren sus funciones (Meier *et al.*, 1984; Schubert *et al.*, 1984) dando lugar a formas que, por ejemplo, generen productos de transcripción defectuosos relacionados con las infecciones persistentes (Francis and Southern, 1988). De todos modos existen al respecto otras teorías diferentes (Bruns *et al.*, 1990).

Un análisis de la secuencia de aminoácidos deducida a partir de los genes clonados de los virus Tacaribe, LCM y Pichindé (Iapalucci *et al.*, 1989a; Salvato *et al.*, 1989) muestra algunas coincidencias con aquéllas que corresponden a replicasas virales.

Uno de estos motivos, presente en 26 RNA polimerasas RNA dependientes, 20 DNA polimerasas RNA dependientes y 11 DNA polimerasas DNA dependientes, consiste en una secuencia consenso de 15 aminoácidos con los residuos Asp-Asp en el centro, flanqueados por ciertos aminoácidos, algunos de los cuales son hidrofóbicos (Kamer and Argos, 1984; Kornberg and Baker, 1992). En los tres arenavirus se encuentran varias secuencias que correlacionan con la descrita: 8 en LCM, 7 en Pichindé y 10 en Tacaribe. De todos estos sitios, solamente uno se encuentra conservado en secuencia y posición en un alineamiento

múltiple. Se conservan, no sólo los residuos Asp-Asp sino además, algunos de los aminoácidos vecinos, siendo el consenso propio de arenavirus: Tyr-Thr-Ser-Ser-Asp-Asp-{Gln/Glu}-{Ile/Val} (1317-1324 en LCM, 1299-1306 en Pichindé y 1325-1331 en Tacaribe). Los cuatro aminoácidos centrales de este consenso (Ser-Ser-Asp-Asp) coinciden con el tetrapéptido central de una secuencia de 15 residuos altamente conservada, característica de la RNA polimerasa del virus de influenza.

Por último, otro motivo descrito es el pentapéptido Gln-Gly-Asp-Asn-Gln (Poch *et al.*, 1990), el cual está ausente en las tres RNA polimerasas.

Si bien no hay una correspondencia total con las regiones consenso que indican Kamer y Argos (1984) para replicasas de virus y Tordo *et al.* (1988) y Galinski *et al.* (1988) para replicasas de virus de RNA de polaridad negativa, su tamaño, la existencia de algunos de los motivos y su presencia en los complejos ribonucleoproteicos hacen suponer que son un tipo particular de replicasas (Salvato *et al.*, 1989).

Un aspecto llamativo es que, si bien la proteína L presenta estas regiones consenso con otras replicasas, el análisis en bancos de datos revela homologías poco explicables. Así, la replicasa del virus de la estomatitis vesicular presenta mayor homología con la proteína N del virus LCM (Salvato *et al.*, 1989) que con la replicasa misma, aunque esto podría considerarse como un caso de evolución convergente.

Tabla 3.c. Características de las proteínas L

Virus	Cepa	Longitud en aminoácidos	Peso molecular calculado (kDa)	Referencias
LCM	Armstrong	2210	254550	Salvato <i>et al.</i> , 1989
Pichindé	3739	2190	252091	Harnish (comunicación personal)
Tacaribe	T.RVL.II 573	2210	252211	Iapalucci <i>et al.</i> , 1989a

Otro aspecto destacable es la escasa homología (identidad y cambios conservativos) que se encuentra cuando se efectúan alineamientos, múltiples o de a pares, de las secuencias nucleotídicas o aminoácidas de las RNA polimerasas de los arnavirus. Como se puede observar en la **Tabla 3.d**, los porcentajes de identidad apenas superan el 25% derivado de similitudes aleatorias (States and Boguski, 1991), mientras que al considerar también los cambios conservativos, estos valores se elevan significativamente. Es importante indicar que, tanto en identidad como en homología, la similitud entre los dos arnavirus del nuevo mundo (Pichindé y Tacaribe) es mayor que la que existe entre cualquiera de ellos y el arnavirus del viejo mundo (LCM). Un análisis más detallado del alineamiento múltiple permite descubrir regiones muy conservadas, con un alto porcentaje de similitud, y regiones muy variables (**Figura 3.6**).

Por otra parte, la proteína L de LCM presenta una región que podría estar asociada con una actividad de proteinquinasa, ya que la secuencia Asn-Leu-Lys-Leu-Gly-Asp la asemeja a la proteinquinasa de serinas CDC²⁸ (Salvato *et al.*, 1989). Esto concuerda con la actividad postulada para la proteína L del virus Pichindé por Howard and Buchmeier (1983), aunque la secuencia potencial identificada en LCM no está presente, tal cual, ni en Pichindé ni en Tacaribe. En los tres arnavirus existen varias secuencias similares en distintas posiciones, pero la simple comparación de secuencias no permite, hasta el momento, predecir si la actividad enzimática existe y cuál es la secuencia aminoácida responsable de la misma.

De todos modos, esta proteína no ha sido totalmente purificada hasta hoy como para analizar la actividad de polimerasa (y las otras supuestas) en un sistema *in vitro*. En el virus LCM se han realizado estudios de la actividad de replicasa en viriones (Fuller-Pace and Southern, 1988) y en extractos de células

infectadas (Fuller-Pace and Southern, 1989). Más recientemente, trabajos preliminares con los virus Pichindé (Oubiña *et al.*, 1990) y Tacaribe (Garcín and Kolakofsky, 1992) permitieron establecer la actividad de polimerasa a partir de complejos ribonucleoproteicos purificados de células BHK21 infectadas que son capaces de sintetizar RNA viral *in vitro*.

Tabla 3.d. Comparación de las proteínas L

	Identidad (%)			Homología (%)	
	PIC	TAC		PIC	TAC
LCM	33	34	LCM	53	54
PIC		42	PIC		63

Los estudios realizados sobre los mecanismos de replicación del virus Tacaribe han llevado a atribuirle a la proteína L una posible actividad de nucleasa (Raju *et al.*, 1990, Garcín and Kolakofsky, 1990) y quizás de enzima de *capping* (Garcín and Kolakofsky, 1990). Hasta ahora, ninguna de estas actividades postuladas tiene gran apoyo experimental o derivado del análisis de secuencia

Proteína Z

Esta proteína, detectada en principio como una proteína no identificada de un peso molecular aparente de 14 kDa en el virus LCM (Buchmeier and Parekh, 1987), está codificada por el gen Z del RNA L. Casi todo lo que se conoce proviene de los estudios de clonado llevados a cabo por Iapalucci *et al.* (1989b) en el virus Tacaribe, por Salvato and Shimomaye (1989) en LCM y, más recientemente, en Pichindé. Estos trabajos permitieron identificar un marco de lectura abierto de polaridad viral para una proteína de longitud promedio de 93 aminoácidos y un peso molecular promedio aproximado de 10,6 kDa (**Tabla 3.e**).

Es llamativa la escasa conservación de residuos idénticos entre las tres proteínas Z, aunque cuando se consideran también los

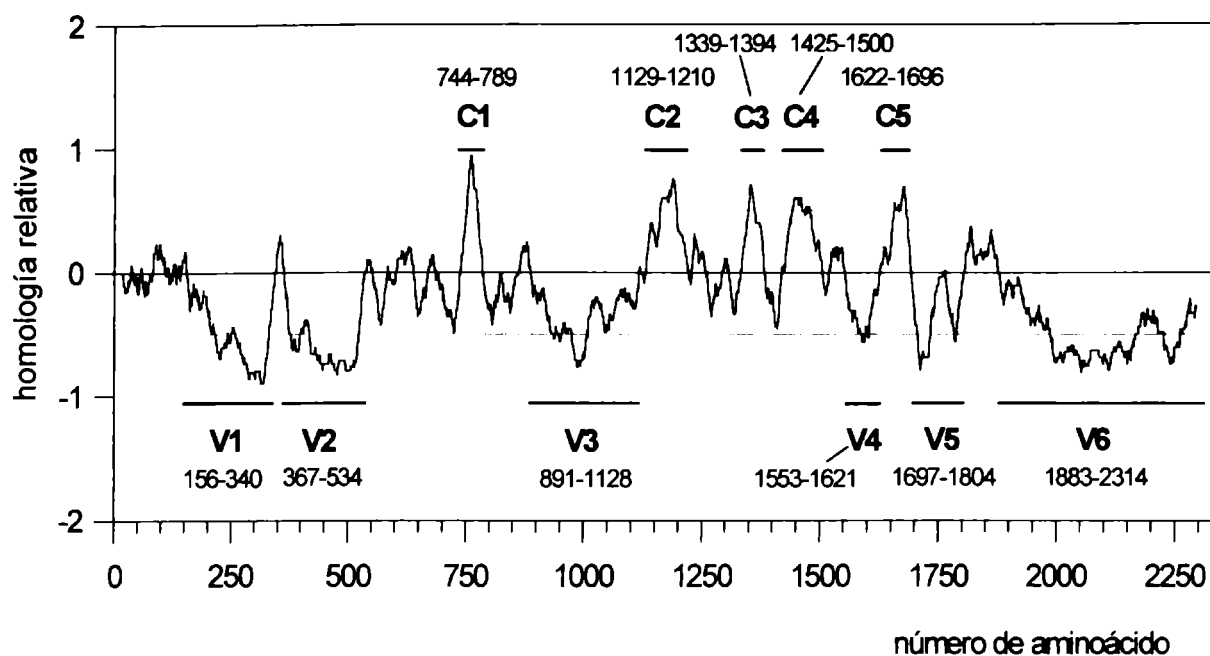


Figura 3.6. Regiones variables y regiones conservadas en las proteínas L de LCM, Pichindé y Tacaribe. El gráfico, derivado de la rutina HOMGRAPH (Ghiringhelli, P.D., 2002, ver Métodos y Miscelánea), muestra 6 regiones variables y 5 regiones conservadas. V1: identidad= 9,1%, variabilidad= 71,4%; V2: identidad= 6,0%, variabilidad= 80,3%; V3: identidad= 16,0%, variabilidad= 63,4%; V4: identidad= 8,7%, variabilidad= 66,7%; V5: identidad= 15,7%, variabilidad= 69,5%; V6: identidad= 9,7%, variabilidad= 71,1%; C1: identidad= 78,4%, variabilidad= 10,8%; C2: identidad= 56,1%, variabilidad= 18,3%; C3: identidad= 52,7%, variabilidad= 19,7%; C4: identidad= 60,6%, variabilidad= 19,7%; C5: identidad= 46,7%, variabilidad= 20,0%. Las posiciones señaladas corresponden al alineamiento múltiple, que incluye *gaps*, y no a la numeración real de cada secuencia. En la región C3 se encuentra la secuencia consenso Tyr-Thr-Ser-Ser-Asp-Asp-{Gln/Glu}-{Ile/Val} descrita en el texto.

cambios conservativos, los porcentajes de homología se incrementan a valores significativos (Figura 3.7).

Salvato & Shimomaye (1989) describen la presencia de una estructura de tipo *zinc finger* (muy parecida a la del factor de transcripción IIIA) en la región carboxilo terminal y, además, han demostrado experimentalmente que la proteína Z de LCM une Zn^{+2} . Del análisis comparativo de las secuencias de las tres proteínas Z conocidas surge que, aún cuando las cisteínas involucradas en el potencial *zinc finger*

están conservadas en cantidad y posición, la Phe que ha sido definida como crucial en el TFIIIA, sólo existe en LCM. La observación detallada del alineamiento múltiple de la Figura 3.7 indica que prácticamente todas las cisteínas están conservadas. La cantidad de las mismas, la distribución y la conservación en los tres arnavirus, sugiere que en lugar de un dominio de tipo *zinc finger*, el dominio de unión de Zn^{+2} podría ser de tipo *ring finger*, presente en algunas proteínas que tienen actividad regulatoria (Figura 3.7 y Tabla 3.f).

Tabla 3.e. Características de las proteínas Z

Virus	Cepa	Longitud en aminoácidos	Peso molecular calculado (kDa)	Referencias
LCM	Armstrong	90	10198	Salvato and Shimomaye, 1989
Pichindé	3739	95	10800	Harnish (comunicación personal)
Tacaribe	T.RVL.II 573	95	10850	Iapalucci <i>et al.</i> , 1989b

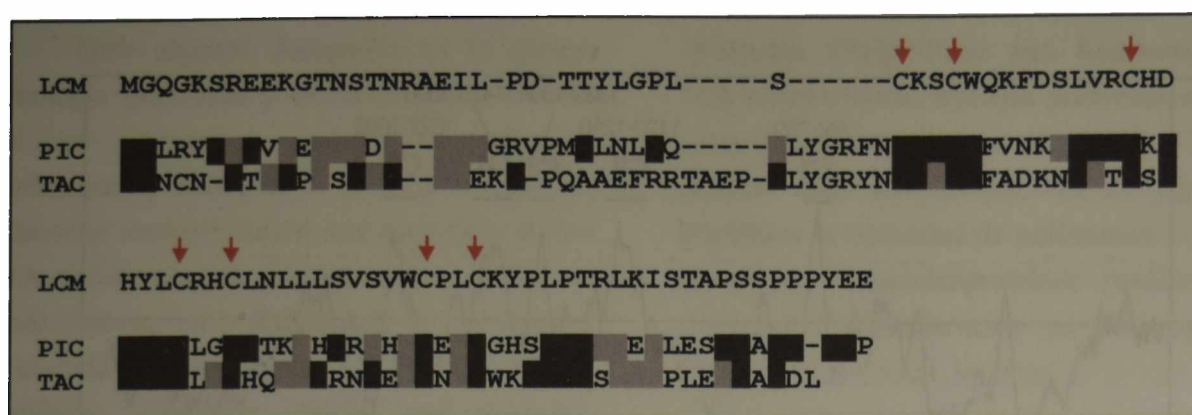


Figura 3.7. Alineamiento múltiple de las proteínas Z de LCM, Pichindé y Tacaribe. La comparación se realizó tomando a la proteína Z de LCM como secuencia de referencia. Se observa que la conservación exacta de aminoácidos es pequeña (LCM/PIC I=31,7%; LCM/TAC I=27,9%; PIC/TAC I=35,6%), mientras que los porcentajes se incrementan al considerar los cambios conservativos (LCM/PIC H=50,0%; LCM/TAC H=43,3%; PIC/TAC H=46,5%). Con flechas rojas se indican las 7 cisteínas conservadas en posición y relacionadas con la estructura de tipo *ring finger* (ver **Tabla 3.f**). La figura deriva de la utilización de la rutina HOMFIG (diseñada y escrita por Ghiringhelli y Rubbi, 1996, no publicada)

Las características generales del polipéptido la asemejan a proteínas de unión a RNA con una posible acción regulatoria de la transcripción en etapas tempranas. Esto se refuerza con el hecho que el mRNA de Z se halla encapsidado en los viriones, de tal modo que podría ser traducido en cuanto el virus entra a la célula sin necesidad de mediar la formación del RNA L viral complementario para su transcripción.

Homologando a la proteína Z con *Tat*, Salvato & Shimomaye (1989) le atribuyen una

posible función regulatoria de la transcripción como antiterminador o como proteína transregulatoria de genes celulares. Esto permitiría explicar los efectos de alteración de los patrones de transcripción de las células infectadas con el virus LCM (Klavinskis and Oldstone, 1989) o la inducción de la dediferenciación celular (Oldstone *et al.*, 1977). En ese sentido, Garcín *et al.* (1993) sugieren que la proteína Z es requerida para la replicación viral y también, aunque en menor grado, para la síntesis de los mRNAs subgenómicos.

Tabla 3.f. Secuencia consenso de unión de zinc en las proteínas Z

Identificación	Banco de datos	Cys totales	Secuencia consenso potencial (Ghiringhelli, P.D., 2002, esta Tesis)	Bibliografía
A47380	PIR	10	C-X ₂ -C-X ₁₂ -C-X ₄ -C -X ₂ -C-X ₁₁ -C-X ₂ -C	Lovering <i>et al.</i> , 1993
HSU15637	GenBank	29	C-X ₂ -C-X ₁₁ ^a -C-X ₄ -C -X ₂ -C-X ₁₁ -C-X ₂ -C	Hu <i>et al.</i> , 1994
HUMCAPIA	GenBank	25	C-X ₂ -C-X ₁₁ ^a -C-X ₄ C -X ₂ -C-X ₁₁ -C-X ₂ -C	Sato <i>et al.</i> , 1995
GMZINCF	GenBank	9	C-X ₂ -C-X ₁₅ -C-X ₄ -H ^b -X ₂ -C-X ₁₀ -C-X ₂ -C	Schauser <i>et al.</i> , 1994
LJRINGP	GenBank	9	C-X ₂ -C-X ₁₅ -C-X ₄ -H ^b -X ₂ -C-X ₁₀ -C-X ₂ -C	Schauser <i>et al.</i> , 1994
HSU09825	GenBank	14	C-X ₂ -C-X ₁₁ -C-X ₄ -C -X ₂ -C-X ₁₃ -C-X ₂ -C	Chu, 1994
HUMERFP	GenBank	29	C-X ₂ -C-X ₁₁ -C-X ₄ -C -X ₂ -C-X ₁₃ -C-X ₂ -C	Inoue <i>et al.</i> , 1993
HUMMEL18	GenBank	11	C-X ₂ -C-X ₁₂ -C-X ₄ -C -X ₂ -C-X ₁₀ -C-X ₂ -C	Ishida <i>et al.</i> , 1993
MUSRAF2A	GenBank	30	C-X ₂ -C-X ₁₁ -C-X ₄ -C -X ₂ -C-X ₁₁ -C-X ₂ -C	Rothe <i>et al.</i> , 1994
U01161	GenBank	9	C-X ₂ -C-X ₂₀ -C-X ₄ -C -X ₂ -C-X ₁₆ -C-X ₂ -C	Senkevich <i>et al.</i> , 1993
Z - LCM	-----	7	C-X ₂ -C-X ₉ -C-X ₅ -C -X ₂ -C-X ₁₀ -C-X ₂ -C	Salvato <i>et al.</i> , 1989
Z - Pichindé	-----	7	C-X ₂ -C-X ₉ -C-X ₅ -C -X ₂ -C-X ₁₀ -C-X ₂ -C	Harnish (com. pers.)
Z - Tacaribe	-----	9	C-X ₂ -C-X ₉ -C-X ₅ -C -X ₂ -C-X ₁₀ -C-X ₂ -C	Iapalucci <i>et al.</i> , 1989

^a X₁₁= X₄-C-X₆

^b La H puede reemplazar a C en la coordinación de Zn²⁺

Otra posible función de la proteína Z podría ser la interacción entre el RNA viral y otras proteínas que contribuyan al ensamblado del virión, como son G1 y G2 según se muestra en la **Figura 3.8** (Salvato, 1993).

PATOGENESIS VIRAL

Infección

La etapa inicial en un ciclo de infección viral es la unión de la partícula viral a la membrana de la célula hospedadora. Esta unión depende de la interacción entre proteínas superficiales del virión (VAP: *Viral Attachment Proteins*) y un componente normal de la membrana plasmática, que actúa como receptor viral. El receptor viral puede ser diferente en distintas variedades celulares del organismo infectado, aunque interactúe con la misma proteína viral; como así también, puede ocurrir que cambios en la proteína viral de superficie se traduzcan en diferencias de afinidad por distintos receptores potenciales en las diferentes variedades celulares, derivando en variaciones en el tropismo celular (Lentz, 1990).

Hasta el presente, no son muchos los sistemas virales en los cuales se ha identificado con claridad el receptor celular.

La segunda etapa es el ingreso del virus dentro de la célula, lo cual puede ocurrir por fusión directa de la envoltura viral a la membrana plasmática, como ocurre en ciertos paramyxovirus, retrovirus y herpesvirus, o por un proceso de viropexis donde los viriones ingresan dentro de vesículas en las cuales ocurre un proceso de fusión de membranas dependiente del pH (Chopin and Scheid, 1980; Marsh and Helenius, 1989; Helenius *et al.*, 1989).

En los arenavirus, es en LCM (virus prototipo de la familia) donde se han estudiado con más detalle las interacciones con las membranas celulares y el proceso de entrada en la célula hospedadora.

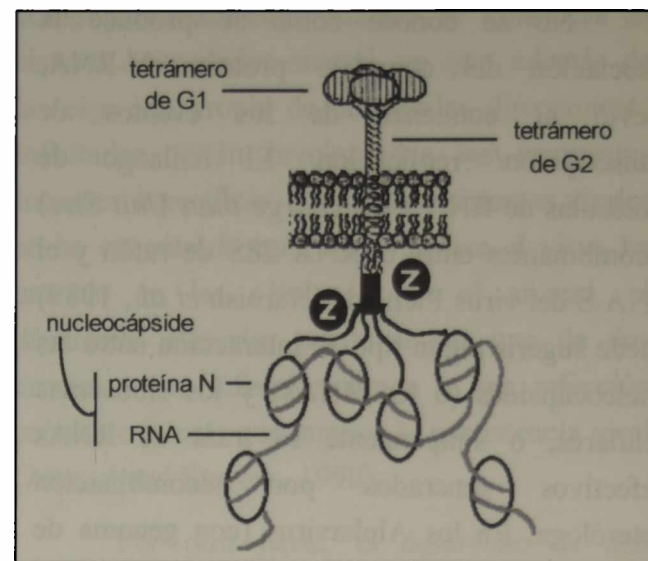


Figura 3.8. Esquema hipotético de interacciones entre proteínas del virión. Este esquema deriva de un conjunto de información obtenido principalmente en LCM. La estequiometría de L:N:G1:G2:Z es 30:1500:650:650:450; G1 y Z son hidrofóbicas y, en experimentos de entrecruzamiento químico, Z se puede encontrar unida a N (Salvato *et al.*, 1992). Los componentes del esquema no están dibujados a escala.

La interacción inicial entre el virus LCM y cultivos celulares de fibroblastos de ratón, hamster y mono ha sido estudiada por Borrow y Oldstone (1992, 1994). Ellos demostraron que, en estos sistemas, la glicoproteína G1 es la única responsable de la interacción inicial, uniéndose a un componente de membrana de naturaleza proteica y de un peso molecular aparente de 120-140 kDa. Esta proteína normalmente está glicosilada, aunque la glicosilación no es imprescindible para la unión del virus.

El ingreso en la célula se produce a través de grandes vesículas de paredes lisas y no asociadas con catrina o microfilamentos. Posteriormente, la progresiva acidificación del endosoma que contiene al virión hace que a pH 5,3-5,5 (endosoma tardío) se produzca la fusión de la envoltura viral con la membrana endosomal, un cambio de conformación irreversible en G1 que lleva a la disociación del complejo G1-G2, la formación de un poro (Di Simone *et al.*, 1994; Di Simone and Buchmeier, 1995) y la liberación de las nucleocápsides en el citoplasma.

No se conoce como se produce la disociación del complejo proteína N-RNA, previo al comienzo de los eventos de transcripción/ replicación. El hallazgo de moléculas de RNA LUS (*Large than Unit Size*) recombinantes entre el RNA 28S de ratón y el RNA S del virus Pichindé (Harnish *et al.*, 1989) puede sugerir algún tipo de interacción entre las nucleocápsides (o los RNAs) y los ribosomas celulares, o simplemente se trata de RNAs defectivos generados por recombinación heteróloga. En los Alphavirus (con genoma de RNA positivo) Sindbis (Wengler and Wengler, 1984) y Semliki Forest (Singh and Helenius, 1992) se ha probado que la disociación de la proteína de la nucleocápside (C) del RNA genómico es dependiente de la interacción de la proteína C con los ribosomas y específicamente con el RNA 28S.

Posteriormente a la disociación del complejo proteína N-RNA se producen los eventos de transcripción, traducción y replicación que llevan, finalmente, a la encapsidación (mecanismo del cual se conoce poco o nada) y la brotación (ya descritos con anterioridad).

Viriones defectivos e interferencia viral

En cualquier ciclo de infección viral, una proporción variable de la progenie generada son formas mutantes, con alteraciones en su patogenicidad y con la capacidad de interferir en el desarrollo de los viriones estándar.

En distintos sistemas virales, es posible encontrar partículas interferentes que presentan desde un simple cambio en la RNA replicasa (Inokuchi and Hirashima, 1987), hasta formas incompletas, con deleciones más o menos importantes o con rearrreglos en su genoma (Whitaker-Dowling and Younger, 1987). Estas partículas se liberan al medio de la misma manera que los virus normales (Fenner *et al.*, 1974). Para diferenciarlos de los virus infectivos, Huang & Baltimore (1970)

propusieron el nombre de partículas defectivas interferentes (DI) y las definieron en base a cuatro características: 1) pierden parte del genoma; 2) presentan las proteínas normales de los virus infectivos; 3) se replican sólo en presencia de virus infectivos homólogos o muy relacionados y 4) interfieren con la replicación de las partículas infectivas (Huang and Baltimore, 1977). Barret & Dimmock (1986) publicaron una revisión del tema con descripciones detalladas de este fenómeno.

En las partículas DI generadas durante las infecciones producidas por virus con genoma de RNA (Lazzarini *et al.*, 1981), las alteraciones son variables. Por ejemplo, en el virus Sindbis afectan la información necesaria para la replicación y el ensamblado (Levis *et al.*, 1986), en el fago Q β se encontraron en el gen de la replicasa (Inokuchi and Hirashima, 1987) y en los coronavirus se encontraron dispersas en el genoma, pero afectando principalmente a la replicasa (Makino *et al.*, 1988).

Las partículas DI interfieren con la replicación del virus estándar del cual provienen sin inhibirla totalmente, de modo tal que puedan seguir produciéndose las formas DI a expensas de la maquinaria replicativa de los virus infectivos, y suelen estar asociadas a los virus estándar en las infecciones persistentes.

Este fenómeno puede observarse también en los arenavirus. En infecciones realizadas con multiplicidades de infección de más de 1 ufp por célula, el efecto lítico (Welsh and Pfau, 1972; Popescu *et al.*, 1976; Dutko and Pfau, 1978; Pevear and Pfau, 1989) y/o citopático (Dutko and Pfau, 1978; López and Franze-Fernández, 1985; López, 1986; Rey *et al.*, 1988) del virus se atenúa al punto de tener monocapas de células con aspecto normal y una reducida producción de viriones infectivos estándar. Infecciones con diluciones sucesivas de los mismos inóculos permiten aumentar la producción de virus infectivos estándar y recuperar la capacidad de formar placas de lisis

debido a que, a multiplicidades de infección menores que 0,01 ufp por célula, la probabilidad de coinfección de una partícula infectiva estándar y una DI resulta despreciable. Este fenómeno, denominado autointerferencia, también puede ocurrir en cultivos celulares durante coinfecciones con altas multiplicidades de diferentes arenavirus (Giménez and Compans, 1980) y, además, puede observarse como sobrevida de animales infectados con altas dosis de, por ejemplo, virus Junín (Help and Coto, 1980).

Las partículas DI de los arenavirus se generan en una alta proporción durante las infecciones agudas y a partir de células persistentemente infectadas. Estas partículas fueron descritas en detalle para los virus LCM (Welsh and Pfau, 1972; Welsh *et al.*, 1972; Welsh *et al.*, 1975; Popescu and Lehmann-Grube, 1977; Pedersen, 1979; Welsh and Buchmeier, 1979; Martínez Peralta *et al.*, 1981) Tacaribe, (Giménez and Compans, 1980; D'Aiutolo and Coto, 1986) y Pichindé (Dutko *et al.*, 1976, Dutko and Pfau, 1978). Estas se caracterizan por tener viriones de menor tamaño que los infectivos, igual composición proteica, aunque con una estequiometría diferente y la existencia de formas de RNA S y L de tamaño subgenómico y de ambas polaridades (genómica y antigenómica), que evidenciarían algún tipo de delección (Romanowski and Bishop, resultados no publicados; Stocker *et al.*, 1994).

El conocimiento acerca de la generación de las partículas DI de los arenavirus y la comprensión de los mecanismos moleculares de la interferencia es aún muy escaso. Bruns *et al.* (1990) han propuesto un modelo de replicación del RNA S del virus LCM en células BHK, que permitiría explicar la formación de las partículas defectivas interferentes.

Infecciones agudas e infecciones persistentes

La interacción de un virus con células en cultivo es compleja y, con más razón, la

complejidad aumenta cuando se considera el sistema hospedador animal, ya que, además de la respuesta propia de las células directamente infectadas, están involucradas las respuestas inmunes específicas. En muchos sistemas virales se ha encontrado que, una vez que el virus ha entrado en las células o en el animal, el desarrollo posterior puede seguir uno de dos caminos: una infección aguda o una infección crónica o latente que implica la persistencia viral (Ahmed and Stevens, 1990).

Por otra parte, el desarrollo de una infección persistente es una condición *sine qua non* para el mantenimiento del virus en la población de reservorios naturales (Meyer and Southern, 1997).

Holland *et al.* (1978) estudiaron la importancia de las partículas DI, las mutaciones y las respuestas inmunes del hospedador en el desarrollo de infecciones persistentes por virus envueltos con genoma de RNA, encontrando que ninguna de ellas es, por sí sola, suficiente para explicar la persistencia.

Sin embargo, una regla que parece ser general es la asociación entre las infecciones persistentes y la ausencia de CTLs (*Cytotoxic T Lymphocyte*) específicos.

En los arenavirus en particular, la interacción con las células en cultivo puede producir tanto infecciones estándar, con o sin efecto citopático, como infecciones persistentes. En los sistemas animales estudiados, las infecciones producen patrones de enfermedad variados, dependiendo de la vía y el momento de entrada de los virus (Staneck *et al.*, 1972; Staneck y Pfau, 1974; Hotchin, 1974; Van der Zeijst *et al.*, 1983; Coto *et al.*, 1979; D'Aiutolo and Coto, 1986; Oldstone *et al.*, 1988).

En este sentido, los estudios sobre la persistencia en el virus LCM, han permitido coleccionar una gran cantidad de datos relacionados con este fenómeno. Una de las cuestiones de mayor importancia es el esclarecimiento de la incapacidad del sistema inmune de un organismo

para reconocer e interferir con la acción de los virus (Zinkernagel and Doherty, 1974, 1977; Welsh and Oldstone, 1977; Zinkernagel *et al.*, 1982; Oldstone, 1987b; Whitton *et al.*, 1988a, b).

Aunque estos aspectos son tan variados y complejos que serían suficientes para una tesis *per se*, es importante mencionar algunos hallazgos recientes, a nivel molecular, relacionados con la persistencia.

Varios autores han trabajado sobre este tema, caracterizando a nivel molecular el tropismo celular y la respuesta inmunológica de clones de LCM (ARM) y LCM (WE) con diferentes patrones de infección. Una síntesis muy apretada de estos resultados muestra que el simple cambio de un aminoácido en la proteína GPC de LCM (ARM) (Phe₂₆₀⇒Leu₂₆₀) es responsable de una selección clonal órgano-específica, de la capacidad para infectar macrófagos selectivamente y necesario, pero no suficiente, para eliminar la respuesta citotóxica (CTL). Otro cambio, hallado en la RNA polimerasa viral (Lys₁₀₇₉⇒Gln₁₀₇₉, Clon 13 o Lys₁₀₇₉⇒Asn₁₀₇₉, Clon 28b), está asociado con el fenotipo CTL⁻, el tropismo hacia los macrófagos y la capacidad de replicar muy eficientemente en los mismos (Salvato *et al.*, 1988; Matloubian *et al.*, 1990; Salvato *et al.*, 1991; Ahmed *et al.*, 1991; Matloubian *et al.*, 1993). Pircher *et al.* (1990), encontraron en LCM (WE) variantes que escapan a la respuesta citotóxica mediante cambios en el epítipo T.

La persistencia a nivel celular es demostrable *in vitro* a partir de células que sobreviven a infecciones líticas de arenavirus y que pueden subcultivarse indefinidamente. Estas células se caracterizan por tener una morfología y crecimiento igual a las no infectadas, por presentar una mayor expresión de la proteína N con una drástica disminución de la expresión de la GPC y ser resistentes a infecciones con virus homólogos, no así, con heterólogos (Oldstone and Buchmeier, 1982; Francis *et al.*, 1987).

Como punto destacable, el establecimiento *in vitro* de una infección persistente, trae aparejada la producción cíclica de partículas DI e infectivas (Popescu and Lehmann Grube, 1977).

Citopatogenicidad

La infección viral se halla asociada a cambios en la morfología y fisiología celular. El conjunto de fenómenos asociados a la infección viral que determinan estos cambios se denomina efecto citopático.

En el virus Junín, al igual que en los demás arenavirus, la infección presenta dos etapas: la etapa inicial se caracteriza por una marcada citopatogenicidad, con destrucción celular y alta producción de viriones. Esta etapa se inicia a partir de focos de células infectadas con un marcado efecto citopático que altera su morfología, rodeadas de otras células de aspecto normal (López, 1986), extendiéndose luego al resto de la monocapa (Tkaczewski and Boxaca, 1970; Tkaczewski and Correa, 1972). Las células que sobreviven a esta etapa de la infección reconstituyen la monocapa con involución del efecto citopático visible al microscopio y establecen las infecciones persistentes.

El efecto citopático más marcado en las infecciones de arenavirus es el cambio de morfología celular. Este ocurre en ciertas líneas celulares susceptibles como HeLa, VERO, CV-1 y MDCK. En cambio, células como la línea BHK-21 no presentan un efecto citopático marcado. Sin embargo, se han reportado cambios morfológicos en etapas agudas a alta multiplicidad de infección con el virus LCM (Romanowski, comunicación personal) o Tacaribe (Garcín and Kolakofsky, 1990). Este efecto se caracteriza por una extensa vacuolización del citoplasma y la aparición de cuerpos de inclusión polimórficos basófilos irregularmente distribuidos en el citoplasma o cerca de la membrana nuclear (Tkaczewski and Correa, 1972). Estos cuerpos de inclusión, vistos al microscopio electrónico, son

estructuras membranosas que incluyen partículas de aspecto ribosomal o ribonucleoproteicas (Murphy *et al.*, 1970, 1973) o solamente estructuras multilamelares como se observó en LCM (Mannweiler and Lehman Grube, 1973). Frecuentemente, aparecen células gigantes de nucleolo marcadamente distinguible (Tkaczewski and Boxaca, 1970). A nivel bioquímico se han descrito como efecto citopático la disminución en el nivel de acetilcolinesterasa en células de neuroblastoma infectadas con el virus LCM (Oldstone *et al.*, 1977, 1978), modificaciones en los niveles de transcripción de genes (como disminución de los niveles de expresión de la hormona de crecimiento), aunque sin alterar las funciones de mantenimiento celular o *house-keeping* (Klavinskis and Oldstone, 1989), alteraciones en las bombas de Ca^{2+} y Na^+/K^+ sensible a la ouabaína (Rey *et al.*, 1988, Rey, 1990).

Capítulo 4 Virus Junín

EL VIRUS JUNIN Y LA FIEBRE HEMORRAGICA ARGENTINA

El virus Junín es el agente etiológico de la Fiebre Hemorrágica Argentina (FHA), la cual es una enfermedad endémica caracterizada por alteraciones cardiovasculares, hematológicas, renales, inmunológicas y neurológicas. La FHA es una enfermedad estacional, que presenta el máximo número de casos durante el mes de mayo (Weissenbacher and Damonte, 1983).

Los primeros casos registrados fueron descritos a partir de una epidemia ocurrida en los años 1953 y 1955 en Bragado, provincia de Buenos Aires (Aribalza, 1955) y, desde entonces, se ha informado una gran variación en el número de casos anuales, entre 300 y 3000. Entre 1958 y 1962, el área endémica de la enfermedad ocupaba una superficie de 16.000 km² en el norte de la provincia de Buenos Aires, extendiéndose rápidamente hacia las provincias de Córdoba, Santa Fe y La Pampa (Maiztegui and Sabattini, 1977), cubriendo en la actualidad más de 150.000 km² en una de las regiones agrícola-ganaderas más importante de la Argentina (Maiztegui *et al.*, 1986; D. Enría, comunicación personal).

La etiología viral de la enfermedad fue establecida en 1958 por dos grupos en forma independiente (Parodi *et al.*, 1958; Parodi *et al.*, 1959a; Pirotsky *et al.*, 1959). Los aislamientos iniciales de virus fueron obtenidos de roedores y ácaros (Parodi *et al.*, 1959a; Parodi *et al.*, 1959b) y a partir de sangre u órganos obtenidos por necropsias en el Hospital Regional de la ciudad de Junín, en la Provincia de Buenos Aires. En esta época se lo consideró un arbovirus al creer que la FHA era mediada por artrópodos (Mettler *et al.*, 1963). Sin embargo, no pudo probarse que los ácaros fueran vectores

de la enfermedad, por lo que se supone que fue una detección casual en un artrópodo de hábitos hematófagos.

El agente etiológico fue llamado virus Junín debido al nombre de la ciudad donde se realizó el primer aislamiento.

Epidemiología de la fiebre hemorrágica argentina

El virus causante de la Fiebre Hemorrágica Argentina generalmente ingresa al hospedador humano a través de las mucosas (conjuntival, inhalatoria u orofaríngea) o por cortes o raspaduras de la piel por contacto con semillas, tallos de malezas o por inhalación de aerosoles o polvo u otros materiales contaminados con la orina o la saliva de un roedor infectado (Aribalza, 1955; Rugiero *et al.* 1964a,b,c; MacKenzie *et al.*, 1964; Johnson *et al.*, 1965; Peters *et al.*, 1974; Maiztegui, 1975; Weissenbacher *et al.*, 1987). Se ha investigado la posibilidad del contagio interhumano, pero sólo se demostró en muy pocos casos, en todos los cuales había existido contacto muy íntimo con el enfermo.

Los reservorios naturales de casi todos los arenavirus son los roedores (Arata & Gratz, 1975). La excepción llamativa a esta regla es el virus Tacaribe, el cual fue aislado inicialmente a partir de un quiróptero frugívoro. Todavía es una incógnita cual fue el camino evolutivo que siguió un posible arenavirus ancestral para pasar de un roedor a un quiróptero herbívoro. En el caso del virus Junín el reservorio principal es *Calomys musculinus*, un roedor con hábitos rurales, aunque también ha sido aislado de *Calomys laucha* (Maiztegui, 1975), *Akodon azarae* y *Orizomys flavescens*. Ocasionalmente se ha encontrado la infección en otros roedores con hábitos urbanos como *Mus musculus*

(Sabattini *et al.*, 1977). En particular *Mus musculus* es el reservorio natural más importante para el virus LCM que, siendo cosmopolita, ha sido aislado dentro del área endémica del virus Junín.

En los roedores infectados, el virus cumple un ciclo que asegura su mantenimiento en la naturaleza. En general se encuentran altos títulos virales en casi todos los órganos y en los fluidos corporales como la sangre y, en particular, la saliva del animal. En la mayoría de los arenavirus, los roedores presentan infecciones crónicas inaparentes, con eliminación persistente del virus al medio ambiente (Staneck *et al.*, 1972; Rawls *et al.*, 1981; Lehman Grube *et al.*, 1983). De esta manera, se contagian los roedores sanos y, ocasionalmente, se infecta el hombre.

Aunque se detectan casos de FHA durante cualquier época del año, esta enfermedad tiene una distribución estacional que abarca desde marzo hasta julio. Este hecho está relacionado con el número de roedores dentro del área endémica. La población de *Calomys* aumenta considerablemente durante estos meses del año, probablemente debido a un cambio en las condiciones ambientales que favorecen su desarrollo. En ese sentido se han hecho especulaciones epidemiológicas que vinculan el cambio en el tipo y las condiciones de cultivo en la pampa húmeda durante la segunda guerra mundial, con la emergencia de la FHA como enfermedad humana. Se supone que la prevalencia de la FHA en varones y, en particular, dentro de la población rural está también estrechamente relacionada con los hábitos de sus reservorios y con la relación ocupacional de la exposición. Durante el primer brote de FHA, la enfermedad afectó principalmente a personas que trabajaban en la cosecha de cereales. En los campos sembrados con cereales se encontró una mayor densidad de roedores así como oportunidades, durante la

cosecha, para la transmisión del virus entre el roedor y el hombre.

Existe una hipótesis que relaciona directamente el aumento progresivo del área endémica de FHA con el aumento del área ocupada por estos roedores. Sin embargo, las verdaderas causas de la extensión progresiva del área endémica de FHA aún no se conocen. Prueba de ello es que varias especies de *Calomys* han sido aisladas en diferentes regiones rurales fuera del área endémica, e incluso se ha demostrado que algunos de ellos estaban infectados (Parodi *et al.*, 1961; Weissenbacher *et al.*, 1985; Mills *et al.*, 1991).

Fisiopatología de la infección por virus Junín

El conocimiento actual sobre la patogénesis de la FHA, proviene principalmente de la experimentación en ratones lactantes y en cobayos (Bruno Lobo *et al.*, 1968; Coto *et al.*, 1972; Garay, 1975). Sin embargo, muchas de las alteraciones fisiológicas y patológicas observadas en estos animales luego de la inoculación con el virus Junín, no se correlacionan con las manifestaciones de la enfermedad en el hombre. Estas limitaciones fueron superadas en FHA, usando *Macacca mulatta* como modelo de la infección en el hombre. La reproducción de la enfermedad en este modelo animal, con manifestaciones clínicas similares a las desarrolladas en humanos, permitió confirmar que las respuestas biológicas de distintas cepas aisladas de pacientes con diferentes cuadros clínicos de FHA eran específicas del virus utilizado (McKee, 1985).

Se ha demostrado que, en la fase aguda de la infección, el virus Junín se replica activamente en linfomonocitos de sangre periférica, sugiriendo que éste es el sitio primario de replicación viral (Ambrosio *et al.*, 1986). Aunque aún no se han realizado estudios definitivos, la evolución de la enfermedad parece estar relacionada con el título del virus en sangre

o en tejidos. Los estudios de microscopía electrónica han mostrado inclusiones citoplásmicas y un daño celular inespecífico en todos los órganos examinados. Los estudios de inmunofluorescencia sugieren que los daños celulares podrían estar relacionados con antígenos virales pero no con inmunoglobulinas o fracciones del complemento. Los estudios clínicos revelan la activación del sistema de complemento, pero no existe evidencia de la formación del complejo inmune (de Bracco *et al.*, 1978). Probablemente el daño celular sea causado directamente por la replicación viral más que por la respuesta inmune (Maiztegui, 1975; Maiztegui *et al.*, 1975). En ese sentido, se ha encontrado que la presencia de la nucleoproteína viral en el citoplasma de células en cultivo mimetiza el efecto citopático producido por la infección con el virus Junín (Rivera Pomar *et al.*, 1991).

La función de los órganos, excepto para el sistema endotelial, permanece intacta. El período crítico de shock es corto, durando apenas entre 24 a 48 hs. El daño hepático es escaso, produciendo una hepatitis mediana (Mando, 1977) y la función renal está bien mantenida. Aún cuando la hemorragia es prominente, no es la causa del shock y de la muerte. Sin embargo, la ruptura capilar es significativa, produciendo una pérdida de proteínas y de volumen intravascular, mucho más pronunciada que la pérdida de eritrocitos. De hecho, el progreso de la enfermedad se ve asociado con una proteinuria elevada y una deshidratación junto con hemoconcentración. Las observaciones clínicas sugieren que hay una disfunción endotelial vascular, y la subsiguiente falla circulatoria, sin que exista una clara evidencia de replicación viral en el endotelio con el correspondiente daño o ruptura endotelial.

Principales manifestaciones clínicas de la enfermedad

Las fiebres hemorrágicas sudamericanas tienen un comienzo insidioso de malestar

general, con fiebre, mialgia y anorexia. En el caso de FHA el período de incubación dura entre 6 y 14 días, mientras que en los primeros 4 días (período de invasión) aparecen los síntomas clínicos tempranos más característicos, que incluyen decaimiento, anorexia, cefalea y mialgias. Esta sintomatología fue descrita a partir de una epidemia ocurrida en los años 1953 y 1955 en Bragado, provincia de Buenos Aires (Aribalzaga, 1955).

Después de un período de estado (desde que comienzan los síntomas) de 6 a 11 días, se puede producir, en los pacientes con fiebre hemorrágica, la remisión espontánea de la enfermedad. En caso contrario aparece un conjunto de síntomas poco específicos, que caracterizan el comienzo de la segunda etapa de la enfermedad. En esta etapa la enfermedad puede aparecer como un síndrome hemorrágico agudo, cuyas manifestaciones más comunes son epistaxis y hematemesis (Melcon and Herskovits, 1981), o como un síndrome neurológico agudo (Rugiero *et al.*, 1960). Los informes acerca de infecciones subclínicas son escasos. Después del comienzo de los síntomas hasta que se produce la curación o muerte transcurre un período de entre 12 y 15 días.

En la etapa aguda de la enfermedad continúan los síntomas de malestar, fiebre elevada, mialgia severa, anorexia, y aparecen, además, dolor lumbar, sensibilidad abdominal, conjuntivitis, dolor retroorbital, generalmente, con fotofobia y constipación. Después de los dos o tres primeros días de la enfermedad, frecuentemente ocurren náuseas y vómitos. La temperatura es alta, generalmente alrededor de 40°C, con muy pocas fluctuaciones. En el examen físico el paciente aparece como intoxicado y los signos más comunes en la primer semana de la enfermedad son inyección conjuntival, edema palpebral discreto y exantema eritematoso de cara, cuello y tercio superior del tronco. Después del cuarto o quinto día de la enfermedad es muy común encontrar

petequias aunque, en general, en escasa cantidad y circunscriptas a las regiones axilares y al tercio superior de los brazos. Además, aunque la faringitis es poco común, el enantema bucofaríngeo es una de las manifestaciones más características de la enfermedad aguda. Las manifestaciones hemorrágicas que aparecen en casi la mitad de los enfermos, presentan la mayoría de los signos clínicos tempranos más característicos de la FHA. En particular, después de una semana de infección, las encías se presentan tumefactas, congestivas y en algunos casos con hemorragias. Además aparece, en la mayoría de los pacientes, una línea de enrojecimiento en la encía, a nivel de la inserción dentaria, denominada ribete gingival, cuya presencia es un indicador clínico importante de la enfermedad. En ese sentido, en la mujer, se presenta una metrorragia precoz, que es generalmente escasa, pero constante, y se constituye en un signo de gran valor para la sospecha de diagnóstico de FHA.

En general, a nivel del aparato cardiovascular, se observa una bradicardia relativa e hipotensión postural. Además, pueden presentarse arritmias que son transitorias y benignas. Hay diferentes grados de deshidratación y oliguria.

Los virus causantes de las fiebres hemorrágicas sudamericanas no parecen afectar en gran medida las vías aéreas superiores. Algunos enfermos pueden presentar una tos seca, pero sin dolor de garganta. Además, aunque se puede observar enantema faríngeo, no se detectan alteraciones bronco-pulmonares, lo que adquiere especial importancia en el diagnóstico diferencial de otros procesos infecciosos agudos.

Al comienzo de la segunda semana de evolución, entre el día noveno y undécimo, en la mayoría de los casos se produce una franca mejoría, desaparecen o disminuyen las manifestaciones hemorrágicas y neurológicas, y se

restablece la diuresis. A nivel del laboratorio, comienzan a subir los glóbulos blancos y las plaquetas y desaparece la proteinuria. En estos casos, se inicia un período de convalecencia, que puede prolongarse más de un mes. Aunque esta enfermedad cura, generalmente sin secuelas, es frecuente observar transitoriamente durante este período, diferentes manifestaciones como hipoacusia, irritabilidad, inestabilidad emocional y caída del cabello.

Se reconocen varias formas clínicas de FHA: leves o comunes y graves. Las formas graves, a su vez, se clasifican en predominantemente hemorrágicas, predominantemente neurológicas o mixtas, siendo estas últimas las más frecuentes. Estas formas graves se observan en el 30% al 40% de los pacientes no tratados con plasma de convalecientes y son, casi siempre, mortales.

Alrededor de la mitad de los pacientes de FHA, desarrollan un síndrome neurológico agudo. En general, durante los primeros 4 o 5 días el examen cuidadoso sólo permite detectar temblores en los dedos de las manos y en la lengua. Después de los primeros 7 días de evolución de la FHA es común observar un decaimiento general en los pacientes, incluyendo la somnolencia y cierto grado de lentitud y desorientación mental. También hay hiperestesia cutánea, hipotonía muscular e hiporreflexia osteotendinosa. En los casos graves se intensifican el estado de confusión mental y los temblores, con excitación psicomotriz marcada, *delirium tremens*, convulsiones generalizadas y finalmente, coma. Sin embargo, la complicación del sistema nervioso central permanece oscura no existiendo, hasta el presente, evidencia de infección directa de los tejidos nerviosos. Además, los síntomas meníngeos y las alteraciones del fluido cerebrospinal son raros. Por otra parte, en los convalecientes de FHA que han recibido inmunoterapia se ha descrito un

síndrome neurológico tardío que se caracteriza por síntomas cerebelares (Melcom & Heskovits, 1981; Maiztegui *et al.*, 1979; Enría *et al.* 1986).

Casi la mitad de los pacientes con FHA presentan manifestaciones hemorrágicas en la segunda etapa de la enfermedad. Estas hemorragias pueden manifestarse en las superficies mucosas o en la piel como petequias con un aumento previo del volumen del paquete eritrocitario. En general, se observan hemorragias gingivales, epistaxis, metrorragia, petequias, equimosis, melena y hematuria (Maiztegui, 1975). Además, en esta etapa, se mantiene la presencia de eritema en la cara, cuello y tórax, así como también petequias en las encías. Sin embargo, la hemorragia no es la causa del shock y de la muerte. Por otra parte, no se presentan linfadenopatías ni esplenomegalia y la ictericia es poco frecuente, en los estadios terminales de algunas formas graves. En los casos severos aparece, además, el compromiso de los sistemas gastrointestinal y nervioso, con náuseas, vómitos, temblores y convulsiones. Los síntomas son acompañados, generalmente, por uremia, proteinuria, hematuria y oliguria. El daño renal se localiza principalmente en las células tubulares distales y en los conductos colectores y, en pocas ocasiones, afecta los glomérulos o los túbulos proximales (Cossio *et al.*, 1975). Esta localización fue confirmada mediante la observación de valores normales tanto para la velocidad de filtración glomerular, como para el flujo de plasma renal y la eliminación de creatinina (Maiztegui, 1975). Sólo en pocos casos se describió la existencia de falla renal (Agrest *et al.*, 1969).

En los casos fatales se demostró colapso vascular con shock hipotensivo, hipotermia y edema pulmonar. La persistencia del shock hipovolémico, a pesar de la presencia de dilatadores del volumen intravascular, sugiere que esto es debido al daño de la función endotelial y a la pérdida de fluido dentro de los

espacios extravasculares. Las observaciones clínicas llevan a la conclusión que la falla circulatoria, asociada con la pérdida de función endotelial, es importante (Ruggiero *et al.*, 1964a). En el examen microscópico se puede observar una alteración general en las células endoteliales y un edema leve de las paredes vasculares, junto con engrosamiento capilar y hemorragia perivascular (Child *et al.*, 1967; Elsner *et al.*, 1973). Generalmente se observan grandes áreas de hemorragia intraalveolar o bronquial, sin evidencia de la existencia de algún proceso inflamatorio. Además, en la mitad de los casos, aparece una neumonía asociada con bronquitis necrotizante o embolia pulmonar. También existe hemorragia y un infiltrado linfocitario del pericardio junto con alguna evidencia electrocardiográfica de miocarditis intersticial (Ruggiero *et al.*, 1964b). Los nódulos linfáticos están agrandados y obturados mostrando una hiperplasia de las células reticulares. Además son comunes la hemorragia esplénica y la congestión medular con hemorragia pélvica y pericapsular.

A pesar de que en un único caso se encontraron evidencias de productos de degradación del fibrinógeno y ausencia de fibrinólisis (Agrest *et al.*, 1969), la coagulación intravascular diseminada, descrita en algunos casos, no parece ser un rasgo característico de esta patología (Molinas and Maiztegui, 1981; Weissenbacher *et al.*, 1987).

Otras complicaciones, que se observan frecuentemente, sobre todo en el caso de pacientes que no han sido tratados con plasma de convaleciente, son las infecciones bacterianas y micóticas agregadas, que pueden producir una sepsis generalizada. En estos pacientes, la tasa de mortalidad llega hasta un 30%.

Diagnóstico y tratamiento

Terapéutica

La respuesta inmune contra el virus Junín (Cossio *et al.*, 1975; de Bracco *et al.*,

1978) puede ser muy efectiva en la eliminación del virus durante la enfermedad aguda y además, puede ser suficiente para proteger contra futuras infecciones. Los anticuerpos detectados en pacientes que comienzan a recuperarse después de una infección aguda son principalmente anticuerpos neutralizantes. En ese sentido, la eficacia terapéutica del tratamiento de los pacientes de FHA con plasma inmune (plasma de convalecientes de FHA) está directamente relacionada al título de anticuerpos neutralizantes en el plasma suministrado (Enría *et al.*, 1984). Por otra parte, aunque se ha mostrado alguna evidencia, en humanos, de respuesta inmune mediada por células, se desconoce su rol en la eliminación del virus y en la protección subsiguiente.

En la década del 70 se realizaron importantes avances en el tratamiento de la FHA. Los resultados de los estudios originales sobre la utilización de la administración pasiva de anticuerpos como tratamiento para la infección de virus Machupo en primates (Eddy *et al.*, 1975), sugirieron que esta terapia podría tener éxito en humanos para el tratamiento de FHA. Poco tiempo después, un ensayo realizado sobre un muestreo estadístico de pacientes de FHA en el Instituto Nacional de Estudios sobre Virosis Humanas en Pergamino, estableció la eficacia de la administración de plasma de la fase convaleciente como tratamiento. Esta terapia redujo drásticamente la tasa de mortalidad del 30% a menos del 1% sólo cuando se aplicó antes del octavo día después del comienzo de la infección (Maiztegui *et al.*, 1979). Esta forma de tratamiento neutraliza la viremia, acorta la duración del período febril, disminuye la intensidad de la mayoría de las alteraciones y reduce la incidencia de las infecciones bacterianas agregadas.

Hasta el momento, la administración de plasma inmune, obtenido de pacientes que han sobrevivido a la infección por virus Junín, se ha convertido en la única terapia efectiva contra la

FHA. Sin embargo esta mejora no se logró sin un costo, ya que 10% de los pacientes tratados con plasma inmune desarrollaron el síndrome neurológico tardío (Enría *et al.*, 1986). Este síndrome, no se ha observado en pacientes no tratados y, hasta el momento, no ha podido detectarse la presencia de virus Junín en preparaciones de líquido cefalorraquídeo de los pacientes afectados. El síndrome neurológico tardío comienza entre cuatro y siete semanas después del comienzo de la enfermedad aguda y termina en menos de una semana más. Este síndrome se caracteriza por fiebre, cefalea, ataxia, temblores y está asociado con una pleocitosis moderada y la presencia de anticuerpos contra el virus Junín en el líquido cefalorraquídeo (Enría *et al.*, 1986). El cuadro neurológico es, en general, benigno, pero puede provocar secuelas y en casos poco frecuentes la muerte del paciente. El título de los anticuerpos neutralizantes contra el virus Junín produce un pico más tardío, aunque de mayor valor, en el caso de los pacientes con síndrome neurológico tardío, comparado con el que se obtiene para los pacientes tratados que no desarrollan el síndrome o con los pacientes no tratados (Weissenbacher *et al.*, 1987). No parece existir ninguna correlación entre el día de la administración de la terapia o la dosis de anticuerpos neutralizantes suministrada, y la aparición del síndrome neurológico tardío.

Además la terapia conlleva otro riesgo que consiste en la posible transmisión al paciente de algún agente infeccioso transportado por sangre. Esto hace imprescindible el ensayo del plasma utilizado para la detección de agentes infecciosos, tales como los virus que producen hepatitis, el HIV y el *Trypanosoma cruzi*.

Por otra parte, la terapia depende de una provisión limitada del plasma que donan voluntariamente los convalecientes de FHA. Previamente a su utilización, se debe obtener el título de anticuerpos neutralizantes presentes en el plasma donado. Una última complicación es

que desde el comienzo de la aplicación masiva de la terapia, se ha observado una disminución en el título de anticuerpos neutralizantes detectados en los individuos receptores del plasma inmune (D. Enría, comunicación personal).

Con respecto a la prevención de la enfermedad, se ha desarrollado una vacuna mediante un proyecto de colaboración entre el Ministerio de Salud Pública de la República Argentina y los laboratorios de USAMRIID en Maryland, Estados Unidos. Esta es una vacuna a virus vivos de virulencia atenuada (Barrera Oro & Eddy, 1982).

La vacunación consiste en la administración de una cepa de virus Junín de virulencia atenuada, llamada Candid #1, para provocar la respuesta inmune protectora en los pacientes. La cepa vacunal fue derivada de la misma cepa original de la cual, 20 años antes, se obtuvo la cepa atenuada XJ-Cl₃ (Weisenbacher *et al.*, 1987). La cepa atenuada vacunal demostró no poseer efectos adversos en primates, sin evidencia de persistencia viral. Esta vacuna, que ha pasado exitosamente una serie de pruebas clínicas en Estados Unidos y en Argentina, incluyendo un estudio realizado con 6000 voluntarios, es un adelanto para la protección de individuos con muy alto riesgo de contraer la enfermedad. De hecho la vacunación es una protección importante tanto para los trabajadores rurales en el área endémica como para el personal del sistema de salud que trabaja en estrecho contacto con el virus. Sin embargo, la alta frecuencia de mutación de los virus cuyo genoma está constituido por RNA (Holland *et al.*, 1982), es un argumento a ser tenido en cuenta en la evaluación del riesgo que implica el uso de una vacuna a virus vivos, toda vez que no se conocen los determinantes de la atenuación de la virulencia a nivel molecular. A ello deben sumarse además, la alta frecuencia de recombinación y reasociación de los virus con

genoma segmentado (Vezza *et al.*, 1980) y la aparición de viriones con genomas poliploides (Romanowski and Bishop, 1983). Por otra parte, aunque los intentos de desarrollo de una vacuna con virus inactivados no tuvieron éxito, esta estrategia está siendo reevaluada por otros investigadores.

Diagnóstico clínico

El diagnóstico de FHA es primariamente establecido por examen clínico, análisis generales de laboratorio bioquímico y anamnesis del paciente. Los síntomas y signos iniciales de la FHA son similares a los de muchas otras enfermedades infecciosas. Sin embargo, la velocidad de eritrosedimentación, que es normal o está disminuida y la leucopenia permiten descartar la mayoría de las infecciones bacterianas. Además, el recuento de plaquetas es inferior a 100.000 y los tiempos de sangría y de retracción de coágulo están concordantemente prolongados. También, se ha descrito la presencia de un inhibidor de la agregación de plaquetas circulante (Cummins *et al.*, 1990) junto con niveles disminuidos de los factores de coagulación II, V, VII, VIII y X y de fibrinógeno. A pesar de ello, las alteraciones en las funciones de coagulación son menores.

En FHA se han observado, además, altos niveles de interferón en pacientes con una enfermedad severa llegando hasta 64.000 UI/ml (Levis *et al.*, 1984), y una disminución en el complemento. Los niveles de interferón en los pacientes con FHA se relacionan con la severidad de la enfermedad y con su progreso (Levis *et al.*, 1985). Contrariamente, los títulos virales en suero son bajos, pero la infección es aparentemente pantrópica (Weissenbacher *et al.*, 1975).

Diagnóstico etiológico

La gran diversidad de formas en que se presenta la FHA, y la alta frecuencia de cuadros clínicos confusos, hace imperativa la necesidad

de contar con un ensayo de laboratorio específico que pueda aplicarse en el período temprano de la enfermedad.

Los inmunoensayos son específicos, aunque su utilidad es limitada ya que la presencia de anticuerpos se detecta tardíamente durante la convalecencia. En efecto, aunque la conversión serológica ocurre, en general, después del primer mes, su detección se utiliza como un ensayo etiológico definitivo, recién a los 60-90 días después de la internación del paciente, sólo con el fin de confirmar el diagnóstico etiológico de FHA.

El ensayo de ELISA es el más comúnmente utilizado en el diagnóstico serológico de FHA (Meegan *et al.*, 1986; Barrera Oro *et al.*, 1990). También se utilizan técnicas de inmunofluorescencia indirecta (análoga a la desarrollada para la fiebre hemorrágica boliviana, Peters *et al.*, 1973) o de neutralización (Webb *et al.*, 1969). Estos ensayos utilizan antígenos virales obtenidos por medio de cultivo de tejidos infectados. La técnica de fijación de complemento es muy poco utilizada en la actualidad (Casals, 1977). Debido al riesgo intrínseco, derivado de la manipulación de virus vivos, se están intentando desarrollar inmunoensayos que utilizan antígenos producidos en el laboratorio, por medio de la tecnología del DNA recombinante (Ghiringhelli *et al.*, resultados no publicados).

Por otra parte, la viremia presente en las primeras dos semanas después de la infección, no alcanza títulos suficientes que permitan detectar antígenos virales por ELISA en la sangre de los pacientes, en el período agudo de la enfermedad. Los bajos títulos de virus Junín en las etapas tempranas de la FHA hacen necesaria la puesta a punto de un sistema diagnóstico con una sensibilidad de detección directa muy alta o con una capacidad de amplificación suficiente para obtener una señal detectable. El aislamiento del virus es más consistente a partir de sangre y suero (Lascano

et al., 1981), aunque también puede ser obtenido de muestras bucales o faríngeas en la etapa aguda y además, ha sido aislado de leche materna (Maiztegui *et al.*, 1973). Si bien la inoculación de animales experimentales (como ratón lactante o hámster) o cultivos celulares, con muestras de sangre de pacientes de FHA, permite amplificar la "señal" original, por replicación del virus en estos sustratos, con una sensibilidad de aproximadamente 50%, el proceso y la detección e identificación del mismo llevan, cuanto menos, un par de semanas.

Este nivel de detección puede ser mejorado hasta un 98% si se emplean para el aislamiento viral células mononucleares de sangre periférica (CMSP) en un procedimiento que requiere el co-cultivo de las mismas con una monocapa de células Vero (Ambrosio *et al.*, 1986). Sin embargo el procedimiento es lento y de manipulación compleja, por lo que sólo sirve para confirmar el diagnóstico clínico.

Por último, es importante tener en cuenta que las manipulaciones con virus Junín, ya sea para su detección o para la obtención de antígenos virales para los ensayos serológicos, poseen una complicación adicional ya que requieren la utilización de condiciones de seguridad especiales.

En ese sentido, los ensayos de diagnóstico basados en la detección de los ácidos nucleicos ofrecen la posibilidad de trabajar con el agente patógeno inactivado en una etapa temprana después de la extracción de la muestra disminuyendo, de esta manera, los riesgos a los que se expone el personal de laboratorio encargado de llevar a cabo el análisis. Además, estos ensayos son, en general, más sensibles y tempranos que los que se basan en la detección de antígenos o de anticuerpos de respuesta, por medio de técnicas inmunológicas.

Actualmente se dispone de un test mediante PCR que permite un diagnóstico muy precoz, sensible y preciso (Lozano *et al.*, 1993; Lozano, 1995).

VARIABILIDAD GENÉTICA EN EL VIRUS JUNÍN

Desde su identificación como agente etiológico de la Fiebre Hemorrágica Argentina, continuamente se han realizado aislamientos a partir de pacientes y de roedores. Los ensayos iniciales en animales modelo permitieron encontrar una alta correlación entre los patrones de virulencia en humanos y en cobayos adultos, por lo cual se los ha utilizado como herramienta principal para caracterizar la patogenia de los aislamientos (Coto *et al.*, 1993).

Acorde con su letalidad en cobayos adultos, se pueden mencionar 3 grupos de diferente virulencia: cepas muy patogénicas, cepas de patogenia intermedia y cepas de escasa patogenia (con diferentes grados de atenuación).

Al grupo de las altamente patogénicas pertenecen la cepa prototipo XJ, de origen humano (Parodi *et al.*, 1958) y la cepa AN 9446, aislada de un roedor en 1967.

Al grupo de las de patogenia intermedia pertenece la cepa MC2, aislada de un roedor (Berría *et al.*, 1967). Por último, al grupo de las atenuadas pertenecen la cepa XJCl₃, derivada de XJ por sucesivos pasajes en laboratorio (Guerrero *et al.*, 1969), la cepa IV 4454, aislada de un caso humano leve (Contigiani and Sabattini, 1977) y la cepa Candid #1, derivada de XJCl₃ por pasajes sucesivos en laboratorio (Barrera Oro and Eddy, 1982).

Algunos de los índices de patogenia medidos se resumen en la **Tabla 4.a** (tomada de Coto *et al.*, 1993).

En síntesis, tanto en poblaciones naturales como en cepas obtenidas en laboratorio, existe una considerable variabilidad en cuanto a sus características patogénicas. Esta variabilidad necesariamente tiene una base molecular que puede comenzar a ser dilucidada recién en el momento en que se disponga de información detallada a este nivel.

Tabla 4.a. Características de las cepas del virus Junín

	Cepa			
	IV4454	XJ	XJCl ₃	MC2
Patogenicidad en cobayos (% de mortalidad)	30	100	20	80
Título de neutralización con antisuero contra XJCl ₃	407	584	9058	901
Citopatología				
% de células viables	18	22	78	90
Morfología de placas	lítica	lítica	turbia	muy turbia
% de inhibición de la síntesis proteica	65	35	25	0
Producción máxima en células Vero	9x10 ⁶	5x10 ⁶	4x10 ⁶	8x10 ⁵

PARTE II

RESULTADOS:

VIRUS JUNIN CEPA MC2

Capítulo 5

Clonado y secuenciamiento del RNA S - Cepa MC2

CARACTERISTICAS DE LA CEPA MC2

La cepa MC2 fue aislada en la localidad de Monés Cazón (Buenos Aires) a partir de un roedor de campo infectado (Berría *et al.*, 1967). Esta cepa es de patogenicidad intermedia para cobayos, ratones y humanos. En el sistema modelo de cobayos muestra un 70% de letalidad y su tropismo está dirigido principalmente al tejido linfático y nervioso (Weissenbacher *et al.*, 1987; Candurra *et al.*, 1989).

Un inóculo de la cepa MC2 fue cedido gentilmente por investigadores del Instituto Nacional de Microbiología "Dr. Malbrán" (De Mitri and Martínez Segovia, 1971). Para su utilización en las etapas posteriores, esta cepa fue clonada por plaqueo, como se muestra en la **Figura 5.1** (Romanowski, 1981).

CRECIMIENTO EN CELULAS BHK-21

Con el fin de obtener viriones para purificar el RNA, se infectaron monocapas de células BHK-21 (50% de confluencia) con un inóculo obtenido a partir de una placa de lisis aislada. El inóculo, de bajo número de partículas interferentes, se diluyó en medio de cultivo y se usó para infectar con multiplicidades de infección entre 0,01 y 0,1 ufp/cél.

En los días 3, 4 y 5 post-infección, de mayor título, se colectaron los sobrenadantes, cada 6 u 8 horas, para la posterior purificación de viriones. Dado el escaso rendimiento de este virus, de crecimiento lento y baja productividad, resultó necesario realizar crecimientos a gran escala, utilizando hasta 16 frascos de 175 cm² infectados para obtener una cantidad suficiente de viriones para purificar el RNA.

PURIFICACION DE RNA VIRAL

Los sobrenadantes de cultivos infectados

fueron concentrados por ultracentrifugación.

El RNA viral fue purificado a partir de los viriones concentrados, utilizando tiocianato de guanidinio (Chomczynski and Sacchi, 1987). y una posterior centrifugación a través de un colchón de CsCl 5,7 M para eliminar los restos de DNA contaminante.

El análisis en geles de agarosa con 10 mM hidróxido de metilmercurio como agente desnaturalizante permite observar los RNAs virales S y L, y los rRNAs 18 y 28S, derivados de los ribosomas encapsidados en el virión.

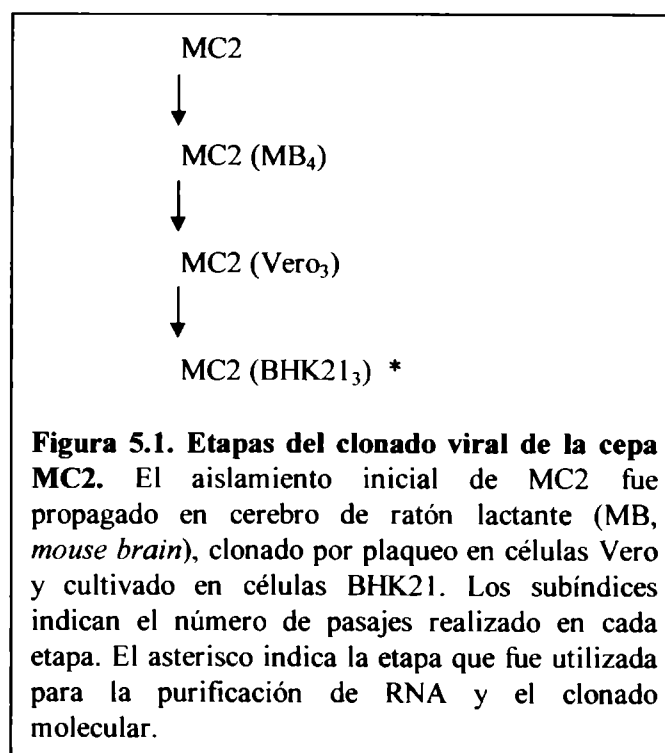


Figura 5.1. Etapas del clonado viral de la cepa MC2. El aislamiento inicial de MC2 fue propagado en cerebro de ratón lactante (MB, mouse brain), clonado por plaqueo en células Vero y cultivado en células BHK21. Los subíndices indican el número de pasajes realizado en cada etapa. El asterisco indica la etapa que fue utilizada para la purificación de RNA y el clonado molecular.

SINTESIS DE cDNA, CLONADO Y SECUENCIAMIENTO

Las síntesis de cDNA se realizaron inicialmente utilizando transcriptasa reversa de AMV y nucleasa S1 (Maniatis *et al.*, 1982). Como los resultados obtenidos no fueron muy alentadores, se pasó a utilizar la técnica descrita por Gubler & Hoffman (1983; Gubler 1988), que emplea una transcriptasa reversa, RNasa H y DNA polimerasa I.

En las síntesis iniciales se utilizó el *primer* 5' CGCACAGTGGATCCTAGGC 3' (*primer* universal de arenavirus, posición 1 a 19 VC), complementario a la secuencia del extremo 3' viral de los RNAs S y L de otros arenavirus (Auperin *et al.*, 1982a).

El cDNA de doble cadena se modificó con la DNA polimerasa del bacteriofago T4 para producir extremos romos y se insertó en el sitio Hinc II del plásmido pUC 9 (Vieira and Messing, 1982) utilizando la DNA ligasa de T4. Alternativamente, y con el objeto de aprovechar la mayor eficiencia de la reacción de ligación sobre extremos cohesivos, se modificó el cDNA con la metilasa Eco RI, se adicionaron *linkers* Eco RI a los extremos del cDNA, se digirió con la endonucleasa Eco RI y se utilizó el sitio correspondiente del plásmido pUC 18 (Vieira and Messing, 1982) para su clonado.

Con los plásmidos recombinantes se transformaron *E. coli* MC 1061 competentes y se detectaron secuencias virales por hibridación de colonias sobre papel de filtro Whatman 541 (adaptado de Grunstein & Hogness, 1975). Inicialmente, se usó como sonda un cDNA marcado con $\alpha(^{32}\text{P})\text{dATP}$ de alta actividad específica por extensión del *primer* arriba mencionado.

De esta manera, se obtuvieron clones positivos correspondientes al extremo 3' del RNA viral. La hibridización en *Northern blots* permitió identificar que algunos de ellos correspondían al RNA L y otros al RNA S.

Uno de los clones de mayor longitud (pJUN63, 947 bp) se caracterizó con endonucleasas de restricción (ver Métodos) y se determinó su secuencia nucleotídica por el método de Maxam & Gilbert (1977). También, se determinó la secuencia de otro clon (pJUN9), de 308 pares de bases, superponible con el primero. La secuencia obtenida resultó idéntica a la zona homóloga del clon pJUN63.

Con el objeto de identificar nuevos recombinantes del RNA S entre los clones de

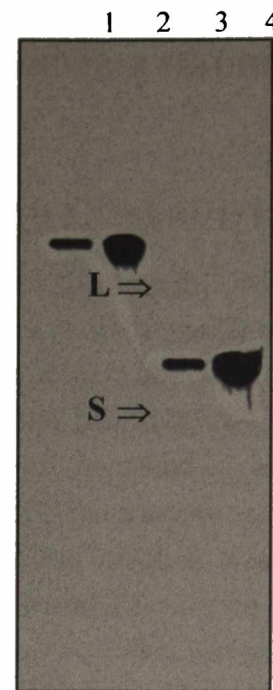


Figura 5.2. Análisis por Northern blot de clones de cDNA de la cepa MC2. RNA purificado de viriones fue sometido a electroforesis en gel de agarosa 1 % con hidróxido de metilmercurio como desnaturizante (50 ng, calles 1 y 3, 500 ng, calles 2 y 4). El RNA fue transferido a una membrana de nylon GeneScreen e hibridado con sondas derivadas de los clones de cDNA pJUN251 (calles 1 y 2) y pJUN9 (calles 3 y 4). Las posiciones de los RNAs L y S del virus Junín (que se indican) y las de los RNAs ribosomales 18 y 28S fueron visualizadas con bromuro de etidio previo a la transferencia.

cDNA, se utilizó como sonda un fragmento de restricción distal del clon pJUN63 marcado por *nick translation* (Maniatis *et al.*, 1982). Este intento dio resultados negativos. Por ello, se diseñó un nuevo *primer* complementario a los nucleótidos 847-865 VC para iniciar una nueva síntesis de cDNA, específica del RNA S, que permitiera avanzar en la secuencia del RNA S del virus Junín. Con este oligonucleótido, de secuencia 5' GGTGGAAATATGTTGGAGA 3' (sintetizado gentilmente por E. Neizert, EMBRABIO), se generó una nueva serie de clones recombinantes por medio de las técnicas ya descritas. El más largo de estos clones (pJUN42) incluía una secuencia de 983 nucleótidos correspondiente al RNA S del virus Junín, parcialmente superpuesto (63 nt) con el extremo distal del clon pJUN63.

El clon pJUN42 fue caracterizado con endonucleasas de restricción y subclonado en la forma replicativa de los bacteriofagos M13 mp18 y mp19 (Yanish-Perrón *et al.*, 1985). Utilizando una modificación del método dideoxi de Sanger *et al.* (1977) se determinó la secuencia nucleotídica de los subclones. Las zonas ricas en pares GC se secuenciaron sustituyendo el dGTP por dITP, obteniéndose lecturas claras y precisas. Los mejores resultados se obtuvieron cuando se usó la DNA polimerasa del fago T7 (Sequenase™, USB) en lugar del clásico fragmento de Klenow de la DNA polimerasa I de *E. coli*. Esta metodología de secuenciamiento se empleó también para confirmar la secuencia del clon pJUN63 y eliminar algunas imprecisiones.

En base a los datos de secuencia obtenidos anteriormente, se diseñó el *primer* 5' CGGACTGGGGGGCCACACC 3', complementario a los nucleótidos 1807-1825 (VC). Los sucesivos intentos de obtener cDNA iniciando con este primer fueron negativos, a pesar de las modificaciones introducidas a la técnica de Gubler & Hoffman (1983), probablemente porque la región complementaria al *primer* corresponde a la parte central de las estructuras tipo *hairpin-loop* intergénicas, que poseen gran estabilidad. Por esta razón, se volvió a utilizar el *primer* complementario a 847-865 (VC) para sintetizar cDNA sobre una nueva preparación de RNA. La síntesis de cDNA y las modificaciones posteriores, incluyendo la adición de *linkers* Eco RI se realizaron de la manera descrita anteriormente. En esta etapa se utilizó como vector pUC13, digerido con Eco RI.

Con los plásmidos recombinantes se transformaron *E. coli* DH5 α competentes y se detectaron secuencias virales por hibridación de colonias sobre papel de filtro Whatman 541 (Grunstein & Hogness, 1975). Como sonda se utilizó un DNA de simple cadena marcado con

$\alpha(^{32}\text{P})\text{dATP}$ de alta actividad específica, generado por extensión del *primer* universal de M13 sobre un molde de DNA de cadena simple correspondiente a un subclón del virus Junín con secuencia ya conocida (nt. 1272-1849, VC). De esta manera se detectaron 34 clones positivos. El análisis de los mismos permitió encontrar varios clones con insertos mayores de 1 kpb. Uno de ellos, pJUN433, con un inserto de 1565 pb fue caracterizado con enzimas de restricción, subclonado en la forma replicativa de los bacteriofagos M13 mp18 y mp19 y secuenciado de la manera ya descrita. Es de destacar que en los subclones secuenciados correspondientes a los 983 bp de superposición entre los clones pJUN433 y pJUN42 no se hallaron diferencias de secuencia.

La secuencia parcial del extremo distal (nt 2410 VC) permitió diseñar un *primer* complementario a los nucleótidos 2255-2271 (VC), 5' ATTGCAGTAAGGGACAC 3', para iniciar una nueva síntesis de cDNA que permitiera llegar al extremo 5' viral.

Con este oligonucleótido se generó una nueva serie de clones recombinantes por medio de las técnicas ya descritas. La hibridación con sondas derivadas del clon pJUN433 permitió identificar 31 clones positivos. Uno de los clones más largos, pJUN58, incluía una secuencia de 1125 nt. que comprendían el extremo 5' viral del RNA S (confirmado posteriormente en el laboratorio por extensión de *primer* sobre RNA viral). El clon pJUN58 fue caracterizado, subclonado y secuenciado de la manera ya descrita.

En resumen, a partir de cinco clones parciales, con diferente grado de superposición (pJUN9, pJUN63, pJUN42, pJUN433 y pJUN58), se determinó la secuencia completa del RNA S de la cepa MC2 del virus Junín (3400 nt) (Figura 5.3).

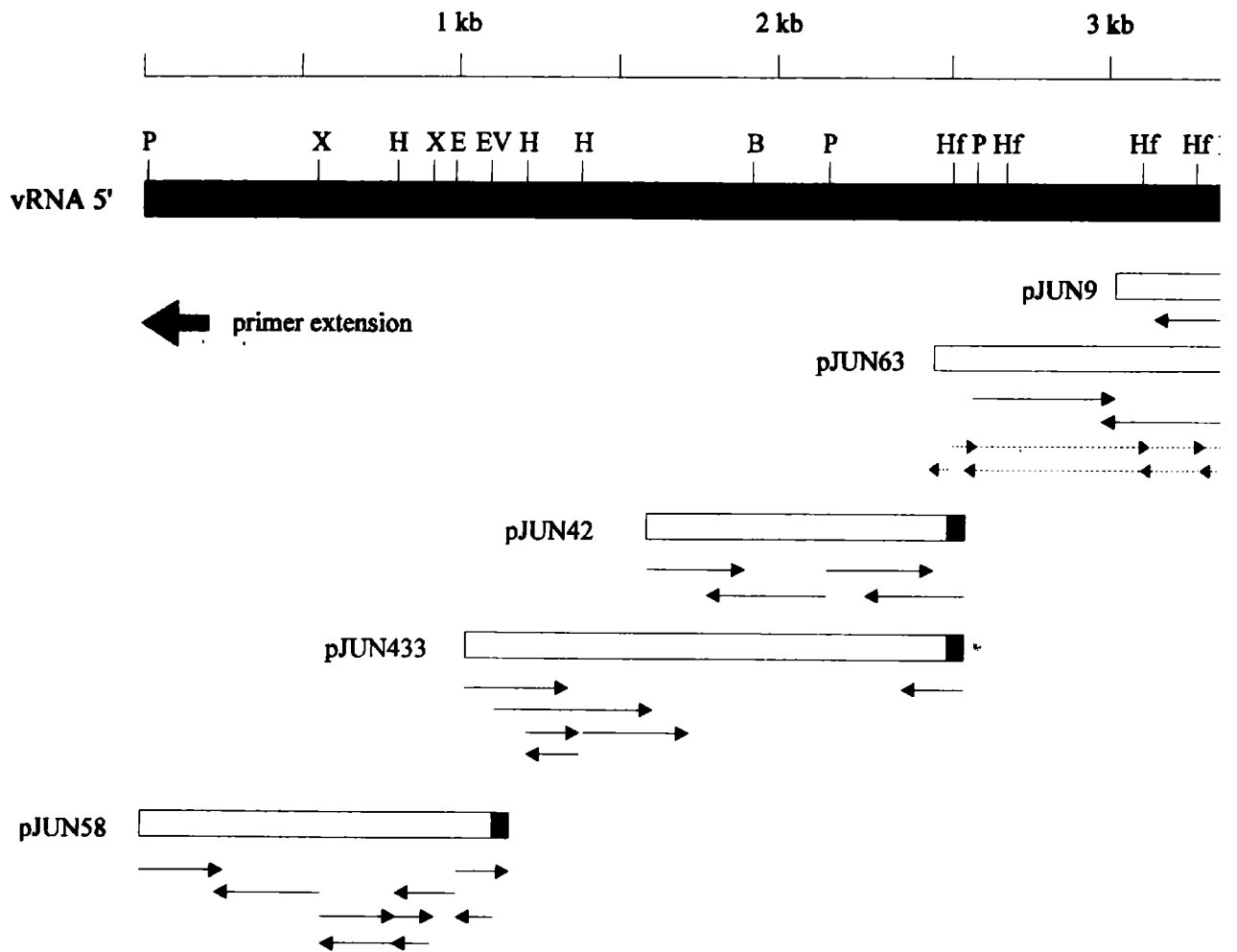


Figura 5.3. Clonado y estrategia de secuenciación del RNA S de la cepa MC2 del virus Junín. La secuencia del RNA S está representada en el sentido viral. Algunos de los clones parcialmente superpuestos se muestran como rectángulos vacíos. Los *primers* utilizados para iniciar la síntesis de cDNA se indican como pequeños rectángulos negros al comienzo de cada clon. Los sitios de restricción utilizados para subclonar en M13 mp18 y mp19 se muestran con abreviaturas (P: PstI; X: XbaI; H: HincII; E: EcoRV; EV: EcoRV; B: BglII; Hf: HinfI; BH: BamHI). Las secuencias obtenidas (Sanger *et al.*, 1977) y la hebra cDNA secuenciada se indican con flechas de línea llena. Con flechas de línea quebrada se señalan fragmentos secuenciados por Maxam & Gilbert (1977). Con una flecha gruesa se muestra la región extremo 5' viral, cuya secuencia fue confirmada por extensión de *primer* utilizando un oligonucleótido complementario a los nucleótidos 144 a 195.

Capítulo 6 Información molecular - Ácidos nucleicos - Cepa MC2

SECUENCIA NUCLEOTIDICA DEL RNA S DE LA CEPA MC2

En base a los datos obtenidos en la etapa de clonado y secuenciamiento y aprovechando la disponibilidad de las secuencias de los RNAs S de otros arenavirus, mediante el análisis comparativo se reconstruyó la secuencia completa del RNA S de la cepa MC2 del virus

Junín (Figura 6.1). Dicha secuencia es de 3400 nucleótidos de longitud y está formada por 903 A, 809 C, 716 G y 972 U (26,6%, 23,8%, 21,0% y 28,6% respectivamente).

La longitud de la secuencia es concordante con la estimada en base a la migración electroforética (Añón *et al.*, 1976) y, además, se corresponde con el rango de tamaños

UGCAGUAAGG	GGAUCCUAGG	CGAUUUUGGU	AACGCUAUA	GUUGUUACUG	CUUUCUAUU	GGACAACAUC	AAACCAUCUA	80
UUUUAACAUG	GGGCAAUUC	UCAGCUUCAU	GCAAGAAUA	CCUACCUUU	UGCAGGAAGC	UCUGAAUAU	GCUCUUGUUG	160
CAGUCAGUCU	CAUUGCCAUC	AUUAAGGGUG	UAGUAAACCU	GUACAAAAGU	GGUUGUCAA	UUUCUGAUCU	AGCAGGAAGA	240
UCCUGCCCAG	GAGCUUUUA	AAUCGGACUG	CACACAGAGG	UCCAGACUG	UGUCCUUCU	CAAUGGUGGG	UCUCUUUUUC	320
CAACAAUCCA	CAUGACCUGC	CUCUGUUGUG	UACCUAAAC	AAGAGCCAUC	UUUACAUAU	GGGGGGCAAU	GCUCUAUUA	400
AGAUCAGCUU	UGAUGACAUC	GCAGUGUUGU	UACCAGAAUA	UGACGUUAU	AUUCAGCAUC	CGGCAGAUAU	GAGCUGGUGU	480
UCUAAAAGUG	AUGAUCAAU	UAGGCUGUCU	CAGUGGUUCA	UGAAUGCUGU	GGGGCAUGAU	UGGUUUCUAG	ACCCACCAU	560
UCUGUGUAGG	AACCGUACAA	AGACAGAAGG	CUUCAUCUU	CAAGUCAUA	CCUCCAGAC	UGGUUCAAU	GAAAACUAUG	640
CCAAGAAGUU	UAAGACUGGU	AUGCACCAU	UAUAUAGAGA	AUACCCCGAC	UCUUGCUUGG	AUGGCAAACU	GUGUUUGAUG	720
AAGGCACAAC	CCACCAGUUG	GCCUCUCCAA	UGUCCACUUG	ACCAUGUCA	CACAUACAU	UUCCUCACAA	GAGGCAAGAA	800
CAUUCAGCUU	CCAAGGAGGU	CUUAAAAGC	AUUCUUUCC	UGGUCUCUGA	CAGACUCAUC	CGGCAAGGAC	ACCCUUGGAG	880
GCUAUUGUCU	AGAAGAGUGG	AUGCUCGUUG	CAGCCAAAUA	GAAGUUAUU	GGCAAUACUG	CUGUAGCAAA	AUGCAAUCUG	960
AAUAUGAGCU	GUGAAUUCUG	UGACAUGCUG	AGGCUUUUG	AUAACAACA	AAAGUCUAUC	AAAACUUUA	AUGCAUAAAC	1040
UAAGAAACAA	CUGAAUUCUGA	UGGGACAGAC	AAUCAUUGCG	CUGAUUUCUG	ACAAUUUAU	GAUGAAAAAC	AAAAUUAGGG	1120
AAUUGAUGAG	UGUCCCUUAC	UGCAAUUACA	CAAAUUUUUG	GUAUGUCAAC	CACACACUU	CAGGACAACA	CUCAUUACCA	1200
AGGUGCUGGU	UAAUAAAAA	CAACAGCUAU	UUGAACAUUU	CUGACUCCG	UAAUGACUGG	AUACUAGAAA	GUGACUUUU	1280
AAUUUCUGAA	AUGCUAAGCA	AAGAGUAUUC	GGACAGGCAG	GGCAAACUC	CCUUGACUU	AGUUGACAUC	UGUUUUUGGA	1360
GCACAGUAUU	CUUCACAGCG	UCCUCUUC	UUACUUGGU	GGGCAUACC	ACCCAUAGG	ACAUCAGAGG	CGAGGCAUGC	1440
CCUCUGCCCC	ACAGGCUAAA	UAGCUUGGGU	GGUUGCAGAU	GUGGUAAGUA	CCCCAAUCU	AAGAAACCA	CAGUUUGGCG	1520
CAGAGGACAC	UAAAGACUCC	CGAAGGUCCC	CACCAGCCCG	GGCAUUGCC	GGGCUUGUGU	GGCCCCCAG	UCCGCGCCU	1600
GGCCGCGGAC	UGGGGAGGCA	CUGCUUA CAG	UGCAUAGGCU	GCCUUCGGGA	GGAACAGCAA	GCUCGGUGGU	AAUAGAGGUG	1680
UAAGUUCUUC	UUCAUAGAGC	UUCCCAUCCA	ACACUGACUG	AAACAUAUUG	CAGUCUAGCA	GAGCAGAGUG	UGGCUACUG	1760
GAGGCCAACU	UAAAGGGAGC	AUCCUUAUCU	CUCUUUUUCU	UGCUGACAAC	CACUCCAUG	UGAUGUUUGC	AUAGGUGGCC	1840
AAAUUCUCC	CAGACCUGUU	GGUCGAACUG	CCUGGCUUGU	UCUGAUGUAA	GCCUAACAUC	AACCAGCUUA	AGAUCUCUUC	1920
UUCCAUGGAG	GUCAACAAC	UUCCUGAUGU	CAUCGGACCC	UUGAGUGGUC	ACAACCAUGU	CCGGAGGACAG	CAAGCCAAUC	2000
ACGUACUAA	GAACUCCUGG	CAUUGCAUCU	UCUAUGUCUU	UCAUUAAGAU	GCCGUGAGAG	UGUCUGCUAC	CAUUUUUAAA	2080
CCUUUCUCA	UCCUUGGUU	UUCUGAAGCA	GUGAAUAUAC	UGUUAACUG	CAGGUCGGAA	CAACGCCAUC	UCAACAGGGU	2160
CAGUAGCUGG	UCCUUCUAAUG	UCGAGCCAAA	GGUUAUUGGU	GGGGUCGAGU	UUCCCCACUG	CCUCUCUGAU	GACAGCUUCU	2240
UGUAUCUCUG	UCAAGUUAGC	CAAUCUCAAA	UUUGACCGU	UCUUUCCGG	UUGUCUAGGU	CCAGCAACUG	GUUUCCUUGU	2320
CAGAUCAAUA	CUUGUGUUGU	CCCAUGACCU	GCCUAUGAUU	UGUGAUCUGG	AACCAAUAU	AGGCCAACCA	UCGCCAGAAA	2400
GGCAAAGUUU	GUACAGAAGG	UUUUAUAAG	GGUUUCUAU	GCCUGGUUUC	UCAUCAUUA	ACAUGCCUUC	UCUUCGUUA	2480
ACCUGAAUGG	UUGAUUUUAU	GAGGGAAGAA	AAGUUUUCUG	GGGUGACUCU	GAUUGUCUCC	AACAUAUUUC	CAUCAUCAAG	2560
AAUGGAUGCA	CCAGCCUUUA	CUGCAGCUGA	AAGACUAAAG	UUGUAGCCAG	AAAUGUUGAU	GGAGCUUUA	UCCUUAGUCA	2640
CAUUCUGGAG	GCAGUCAUGU	UCCUGAGUCA	AUCUGUCAAG	GUCACUCAAG	UUUGGAUACU	UCACAGUGUA	UAGAAGCCCA	2720
AGAGAGGUUA	AAGCCUGUAU	GACACUGUUC	AUUGUCUCAC	CUCCUUGAAC	AGUCAUGCAU	GCAAUUGUCA	AUGCAGGAAC	2800
AGAACCAAAC	UGAUUGUUA	GUUUUGAAGG	AUCUUUAACA	UCCCAUACCC	UCACCACACC	AUUUCCCCA	GUUCCUUGCU	2880
GUUGAAAUCC	CAGUGUUCUC	AAUAUCUCUG	AUCUCUUGGC	CAGUUGUGAC	UGAGACAAGU	UACCCAUUGA	AACCCUUGA	2960
GAGCCUGUCU	CUGCUCUUCU	AAACUUGGUU	UUUAAAUUCC	CAAGGCCAGA	CGCCAACUCC	AUCCGCUCAA	CCUCCCCAG	3040
AUCUCCCGCC	UUGAAAACCG	UGUUUCGUUG	AACACUCCUC	AUGGACAUGA	GUCUGUCAAC	CUCUUUAUUC	AGGUCCCUC	3120
ACUUUAUGAG	GUCUUCUUC	CCCCUUUAG	UCUUUCUGAG	UGCCCGCUGC	ACCUGUGCCA	CUUGGUGAA	GUCAUUGCUG	3200
UCAGCAAUUA	GCUUGGCAUC	CUUCAGAAC	UCCGACUUGA	CAGUCUGAGU	AAAUUGACUC	AAACCUCUCC	UUUAGGACUG	3280
AGUCCAUCUA	AAGCUUGGAA	CCUCUUUGGA	GUGUGGUGG	CCAGAAGAUC	UGGUGGUUUU	GAUCUGAGAA	AAAAUUGCU	3360
AGUGAAAGUG	UUAGACACUA	UGCCUAGGAU	CCACUGUGCG					3400

Figura 6.1. Secuencia nucleotídica del RNA S de la cepa MC2 del virus Junín. Se indican los 3400 nucleótidos del RNA S en sentido viral, numerados a la derecha. En caracteres blancos sobre fondo negro se indican los codones de iniciación y terminación de los genes GPC (polaridad positiva) y N (polaridad negativa, en este caso son los tripletes complementarios). Con flechas apuntando hacia la región del loop se señalan las dos estructuras secundarias intergénicas tipo hairpin loop.

Tabla 6.a. Longitud de los RNAs S de arnavirus

Virus	Cepa	Longitud en nucleótidos	Referencias
Junín	MC2	3400	Ghiringhelli <i>et al.</i> , 1989 Ghiringhelli <i>et al.</i> , 1991
Lassa	GA391 (Nigeria)	3417	Clegg y Oram, 1985 Clegg <i>et al.</i> , 1990
Lassa	Josiah	3402	Auperin and McCormick, 1988
LCM	Armstrong	3376	Salvato <i>et al.</i> , 1988
LCM	WE	3375	Romanowski <i>et al.</i> , 1985 Romanowski and Bishop, 1985
Mopeia	800150	3419	Wilson and Clegg, 1991
Oliveros	---	3535	Bowen <i>et al.</i> , 1996
Pichindé	3739	3419	Auperin <i>et al.</i> , 1984 Bishop and Auperin, 1987
Sabiá	SPH114202	3366	Gonzalez <i>et al.</i> , 1996
Tacaribe	T.RVL.II 573	3421	Franze Fernández <i>et al.</i> , 1987

de los RNAs S de otros arnavirus (Tabla 6.a).

La búsqueda de marcos de lectura abiertos en las seis fases posibles (tres en sentido viral y tres en sentido viral complementario) dio como resultado varios potenciales marcos en ambas orientaciones (Tabla 6.b), entre los cuales fue seleccionado el más largo en cada orientación, en correlación con la información experimental disponible.

Los codones de iniciación y terminación indicados en la Figura 6.1 corresponden a estos marcos seleccionados

REGIONES CODIFICANTES

Gen del precursor de las glicoproteínas

En el primer AUG de polaridad viral, ubicado en los nucleótidos 88-90, comienza un marco de lectura abierto de 1446 nucleótidos de longitud que termina en un codón UAA ubicado en los nucleótidos 1531-1533.

Aunque más allá del AUG₈₈₋₉₀ existen otros codones AUG (Tabla 6.b), el primero es considerado el verdadero codón de iniciación debido a la conservación de algunos de los residuos del entorno definidos por Kozak (1978, 1983, 1984), se conservan la G de +4, la A de -3 y la C de -2 (ver Capítulo 16).

Tabla 6.b. Marcos de lectura abiertos

n°	Ubicación (nt.)	Long. proteína (aa)
Fase 1 (polaridad viral)		
1	88-1533	481
2	109-1533	474
3	469-1533	354
4	520-1533	337
5	661-1533	290
6	718-1533	271
7	901-1533	210
8	919-1533	204
9	985-1533	182
10	1060-1533	157
11	1102-1533	143
12	1126-1533	135
13	1291-1533	80
Fase 2 (polaridad viral)		
1	2093-2344	84
Fase 3 (polaridad viral)		
No hay marcos de lectura mayores de 200 nt.		
Fase 4 (polaridad viral complementaria) *		
1	82-1776	564
2	313-1776	487
3	319-1776	485
4	379-1776	465
5	454-1776	440
6	625-1776	383
7	649-1776	375
8	856-1776	306
9	937-1776	279
10	1252-1776	174
11	1357-1776	139
12	1378-1776	132
13	1423-1776	117
Fase 5 (polaridad viral complementaria) *		
1	1448-1663	72
2	1694-1900	69
Fase 6 (polaridad viral complementaria) *		
No hay marcos de lectura mayores de 200 nt.		

* Se indica la ubicación nucleotídica en sentido VC.

Además, el tamaño del polipéptido resultante (481 aa.) y el peso molecular estimado (55126 Da, sin considerar la glicosilación) se correlacionan con los datos de movilidad electroforética en geles de SDS-PAGE (Rustici, 1984; De Mitri and Martínez Segovia, 1985).

El análisis del uso de codones de este marco de lectura abierto se resume en la **Tabla 6.c.** Como se observa en la misma no todos los aminoácidos presentan preferencias de uso de codón. Considerando la relación entre la

frecuencia particular de un codón y la frecuencia promedio para el mismo, podemos ver que si esta relación es de 1,5 veces o mayor que la frecuencia promedio hay una preferencia neta de uno o más codones para ese aminoácido, mientras que una relación $\geq 1,2$ y $< 1,5$ muestra una preferencia moderada.

En función de lo anterior, se puede destacar el caso de la Arg donde, de los codones iniciados con C dos no son utilizados y otros dos apenas cubren el 15 % de la frecuencia de uso.

Tabla 6.c. Frecuencia de uso de codones en el gen GPC

Aminoácido	Codón	Frecuencia	Aminoácido	Codón	Frecuencia
Ala (A)	GCT	33,3 %	Lys (K)	AAA	51,5 %
	GCC	14,3 %		AAG	48,5 %
	GCA	42,9 %	MET (M)	ATG	100,0 %
	GCG	9,5 %	Phe (F)	TTT	48,0 %
Arg (R)	CGT	10,0 %		TTC	52,0 %
	CGC	5,0 %	Pro (P)	CCT	26,1 %
	CGA	0,0 %		CCC	26,1 %
	CGG	0,0 %		CCA	39,1 %
	AGA	35,0 %		CCG	8,7 %
AGG	50,0 %	Ser (S)	TCT	27,7 %	
Asn (N)	AAT		58,1 %	TCC	16,7 %
	AAC		49,1 %	TCA	13,9 %
Asp (D)	GAT		34,6 %	TCG	2,8 %
	GAC		65,4 %	AGT	16,7 %
Cys (C)	TGT	63,2 %	AGC	22,2 %	
	TGC	36,8 %	Thr (T)	ACT	24,0 %
Gln (Q)	CAA	60,0 %		ACC	28,0 %
	CAG	40,0 %		ACA	48,0 %
Glu (E)	GAA	75,0 %		ACG	0,0 %
	GAG	25,0 %	Trp (W)	TGG	100,0 %
Gly (G)	GGT	26,9 %	Tyr (Y)	TAT	53,3 %
	GGC	38,5 %		TAC	46,7 %
	GGA	23,1 %	Val (V)	GTT	28,6 %
	GGG	11,5 %		GTC	33,3 %
His (H)	CAT	52,9 %		GTA	23,8 %
	CAC	47,1 %	GTG	14,3 %	
Ile (I)	ATT	39,3 %	Term.	TAA	100,0 %
	ATC	39,3 %		TGA	0,0 %
	ATA	21,4 %		TAG	0,0 %
Leu (L)	TTA	19,0 %			
	TTG	19,0 %			
	CTT	17,2 %			
	CTC	8,6 %			
	CTA	12,1 %			
	CTG	24,1 %			

Tabla 6.d. Frecuencia de uso de codones en el gen N

Aminoácido	Codón	Frecuencia	Aminoácido	Codón	Frecuencia
Ala (A)	GCT	32,3 %	Leu (L)	CTC	15,4 %
	GCC	16,1 %		CTA	7,7 %
	GCA	41,9 %		CTG	18,4 %
	GCG	9,7 %	Lys (K)	AAA	47,5 %
Arg (R)	CGT	0,0 %		AAG	52,5 %
	CGC	0,0 %	MET (M)	ATG	100,0 %
	CGA	5,6 %	Phe (F)	TTT	68,4 %
	CGG	5,6 %		TTC	31,6 %
	AGA	58,3 %	Pro (P)	CCT	36,0 %
AGG	30,5 %	CCC		8,0 %	
Asn (N)	AAT	44,0 %		CCA	40,0 %
	AAC	56,0 %		CCG	16,0 %
Asp (D)	GAT	52,6 %	Ser (S)	TCT	20,9 %
	GAC	47,4 %		TCC	20,9 %
Cys (C)	TGT	14,3 %		TCA	20,9 %
	TGC	85,7 %		TCG	2,4 %
Gln (Q)	CAA	46,7 %		AGT	18,6 %
	CAG	53,3 %		AGC	16,3 %
Glu (E)	GAA	48,3 %	Thr (T)	ACT	37,9 %
	GAG	51,7 %		ACC	24,1 %
Gly (G)	GGT	26,2 %			ACA
	GGC	23,8 %		ACG	3,5 %
	GGA	26,2 %	Trp (W)	TGG	100,0 %
	GGG	23,8 %	Tyr (Y)	TAT	54,5 %
His (H)	CAT	36,4 %		TAC	45,5 %
	CAC	63,6 %		GTT	48,6 %
Ile (I)	ATT	50,0 %	Val (V)	GTC	16,2 %
	ATC	23,1 %		GTA	5,4 %
	ATA	26,9 %		GTG	29,8 %
Leu (L)	TTA	7,7 %	Term.	TAA	100,0 %
	TTG	27,7 %		TGA	0,0 %
	CTT	23,1 %		TAG	0,0 %

Gen de la proteína de la nucleocápside

En la fase 4 (Tabla 6.b), de polaridad viral complementaria, el AUG₈₂₋₈₄ comienza un marco de lectura abierto de 1692 nucleótidos de longitud que termina en el UAA₁₇₇₄₋₁₇₇₆.

Aunque más adelante, en la misma fase, existen otros AUGs (Tabla 6.b), el primero es considerado el verdadero codón de iniciación de la proteína N, basado en la conservación de algunos de los nucleótidos del entorno, ya que se conservan la G de +4, la C de -2, la G de -3 y la C de -5 (Kozak, 1978, 1983, 1984; ver Capítulo 16).

Además, el polipéptido resultante tiene un peso molecular calculado de 63033 Da. Este resultado concuerda con el peso molecular estimado para la proteína de la nucleocápside viral a partir de datos de movilidad

electroforética en geles de SDS-PAGE (Martínez Segovia and De Mitri, 1977; Grau *et al.*, 1981; Rustici, 1984).

El análisis del uso de codones de este gen se resume en la Tabla 6.d. Teniendo en cuenta lo ya considerado para el gen GPC, no hay una preferencia generalizada de uso de codones aunque, nuevamente se destaca la Arg, donde dos de los codones iniciados con C no son utilizados, mientras que los otros dos representan apenas el 11,2 % de la frecuencia de uso. En realidad, en este caso, más que hablar de una frecuencia de uso habría que hablar de una frecuencia de desuso, ya que en 7 de 9 aminoácidos (77,8 %) cuyo código está representado por más de dos codones, existe al menos un triplete cuya frecuencia está muy por debajo de la frecuencia promedio.

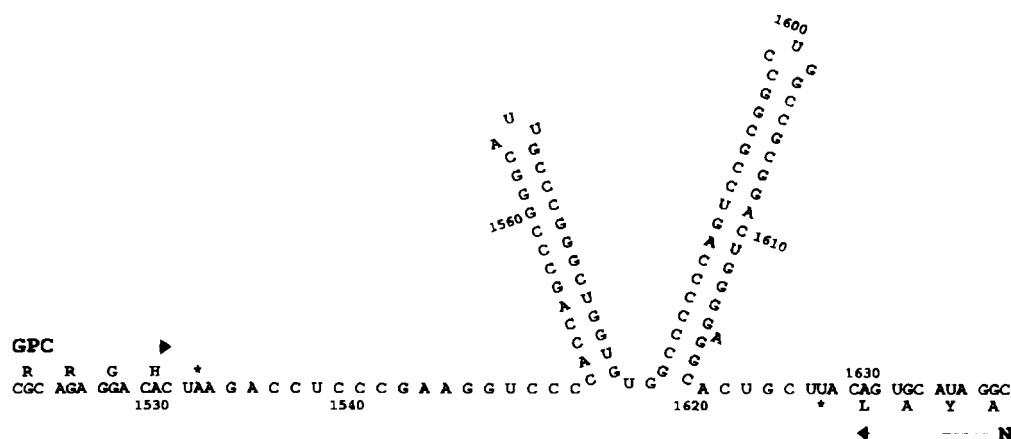


Figura 6.2. Región intergénica del RNA S de la cepa MC2 del virus Junín. Porción del RNA S (nt. 1519-1640) que comprende los extremos de los genes GPC y N (polaridad viral y viral complementaria, respectivamente) y la región intergénica no codificante. Las direcciones de traducción se indican con flechas y los codones de terminación están señalados con asteriscos. Las dos potenciales estructuras tipo *hairpin-loop* están estabilizadas por 13 y 17 pares de bases (ΔG° -39,0 kcal/mol y ΔG° -54,2 kcal/mol, respectivamente).

REGIONES NO CODIFICANTES

Región intergénica

La región intergénica ubicada entre los genes GPC y N, nucleótidos 1534 a 1624 en polaridad viral, contiene dos regiones con autocomplementariedad suficiente como para constituir dos estructuras tipo *hairpin-loop* muy estables (**Figura 6.1**). El *hairpin* cercano al gen GPC (nt. 1551-1579) está formado por 11 pares GC y 2 pares AU (ΔG° -39,0 kcal/mol; Tinoco *et al.*, 1973) mientras que el cercano al gen N (nt. 1582-1618) está formado por 15 pares GC y 2 pares AU (ΔG° -54,2 kcal/mol) (**Figura 6.2**). Estas estructuras, calculadas manualmente al comienzo, fueron confirmadas mediante análisis computacional con el algoritmo de Zuker &

Stiegler (1981).

El virus Junín fue el primer arenavirus en el cual se describió la presencia de un doble *hairpin* intergénico (Ghiringhelli *et al.*, 1987, 1991a), ya que en todas las secuencias publicadas al momento de este hallazgo se había descrito una sola estructura de este tipo (en el **Capítulo 12** se discutirá más sobre este asunto).

Extremos 5' y 3' del genoma viral

Las regiones no traducidas 5' y 3' del RNA viral comprenden 87 y 81 nucleótidos respectivamente. Estas dos regiones presentan una cierta complementariedad y, como ya se mencionó en el capítulo 3, pueden ser alineadas para formar una estructura tipo *panhandle* (ΔG° -58,0 kcal/mol; **Figura 6.3**).

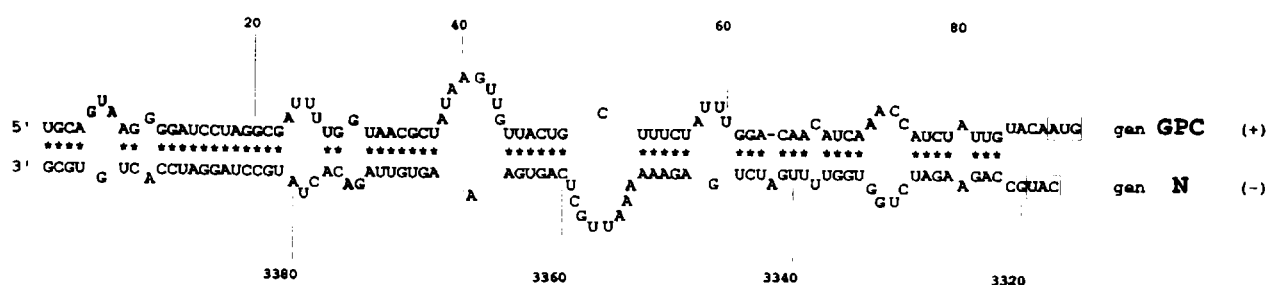


Figura 6.3. Extremos 5' y 3' del RNA S de la cepa MC2 del virus Junín. Estructura *panhandle* predicha para los extremos del RNA S. La complementariedad de nucleótidos entre los extremos no codificantes se indica con asteriscos. Se han incorporado también los pares no convencionales G-U. La energía de estabilización de esta estructura secundaria es de -58,0 kcal/mol (Tinoco *et al.*, 1973). Con recuadros se señalan los codones de iniciación de los genes GPC y N (el triplete complementario al codón AUG en este caso).

RECONSTRUCCION DE GENES

Gen GPC

Por digestión de los clones pJUN58 y pJUN433 con endonucleasas de restricción se obtuvieron los fragmentos necesarios para la reconstrucción. El clon pJUN58 (1125 bp) fue digerido con PstI y EcoRV para generar un fragmento de 1083 bp, que contiene el extremo 5' viral y la mayor parte del gen GPC; el clon pJUN433 (1565 bp) fue digerido con EcoRV y BglII para generar un fragmento de 829 bp, que

contiene el extremo 3' del gen GPC, la región intergénica y la porción 3' del gen N.

Ambos fragmentos fueron purificados a partir de un gel de agarosa y ligados al vector BlueScript™ KS⁺ (Stratagene, La Jolla) digerido previamente con PstI y BamHI (sitio compatible con el extremo BglII), utilizando T4 DNA ligasa. Se transformaron bacterias DH5αF' competentes y se recuperaron las colonias con el plásmido recombinante. La reconstrucción se caracterizó con endonucleasas de restricción (Figura 6.4).

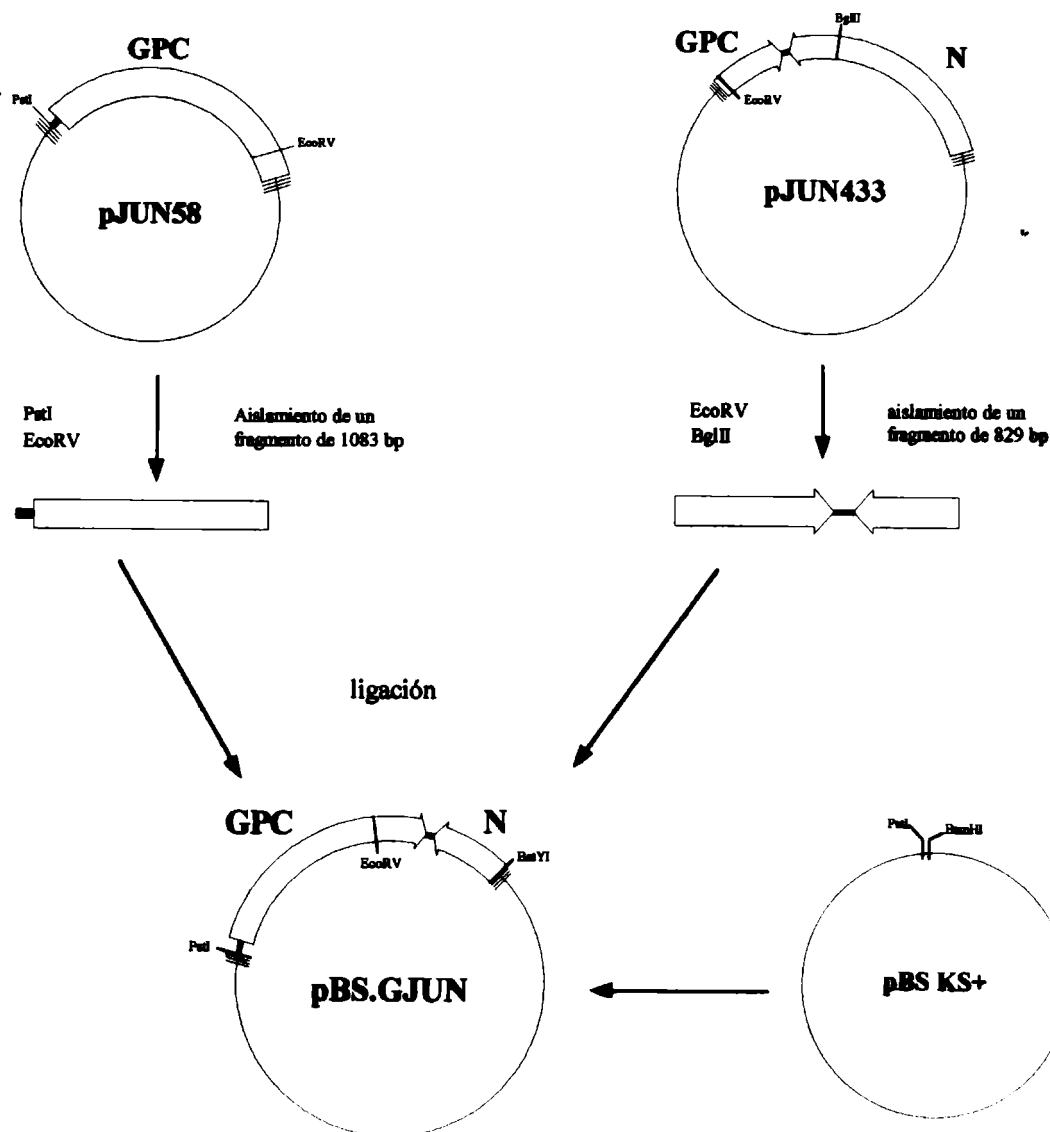


Figura 6.4. Reconstrucción del gen GPC de la cepa MC2 del virus Junín. Se indican los sitios de restricción utilizados en la reconstrucción del gen GPC. Las regiones codificantes aparecen como flechas o rectángulos vacíos, que indican la polaridad del marco de lectura. En trazo grueso aparecen las secuencias virales no codificantes.

Gen N

Por digestión de los clones pJUN63 y pJUN42 con endonucleasas de restricción se obtuvieron los fragmentos necesarios para la reconstrucción. El clon pJUN63 (947 bp) fue digerido con BamHI, PstI y HinfI para generar dos fragmentos de restricción, de 811 bp y 53 bp respectivamente. El fragmento de 811 bp contiene el extremo 3' viral y la mitad del gen N. El clon pJUN42 fue digerido con HinfI y EcoRI para generar un fragmento de 980 bp que contiene la mitad 3' del gen N y la mayor parte

de la región intergénica.

Todos los fragmentos fueron purificados a partir de geles de agarosa. La reconstrucción se realizó en dos etapas. Primero se clonó el fragmento de 811 bp en el vector pBS KS⁺. Una vez recuperado y caracterizado este plásmido recombinante, se realizó la ligación conjunta de los fragmentos de 53 bp y 980 bp al mismo.

Todas las transformaciones se hicieron, alternativamente, en XL1 Blue o DH5 α F'. La construcción final se controló con endonucleasas de restricción (Figura 6.5).

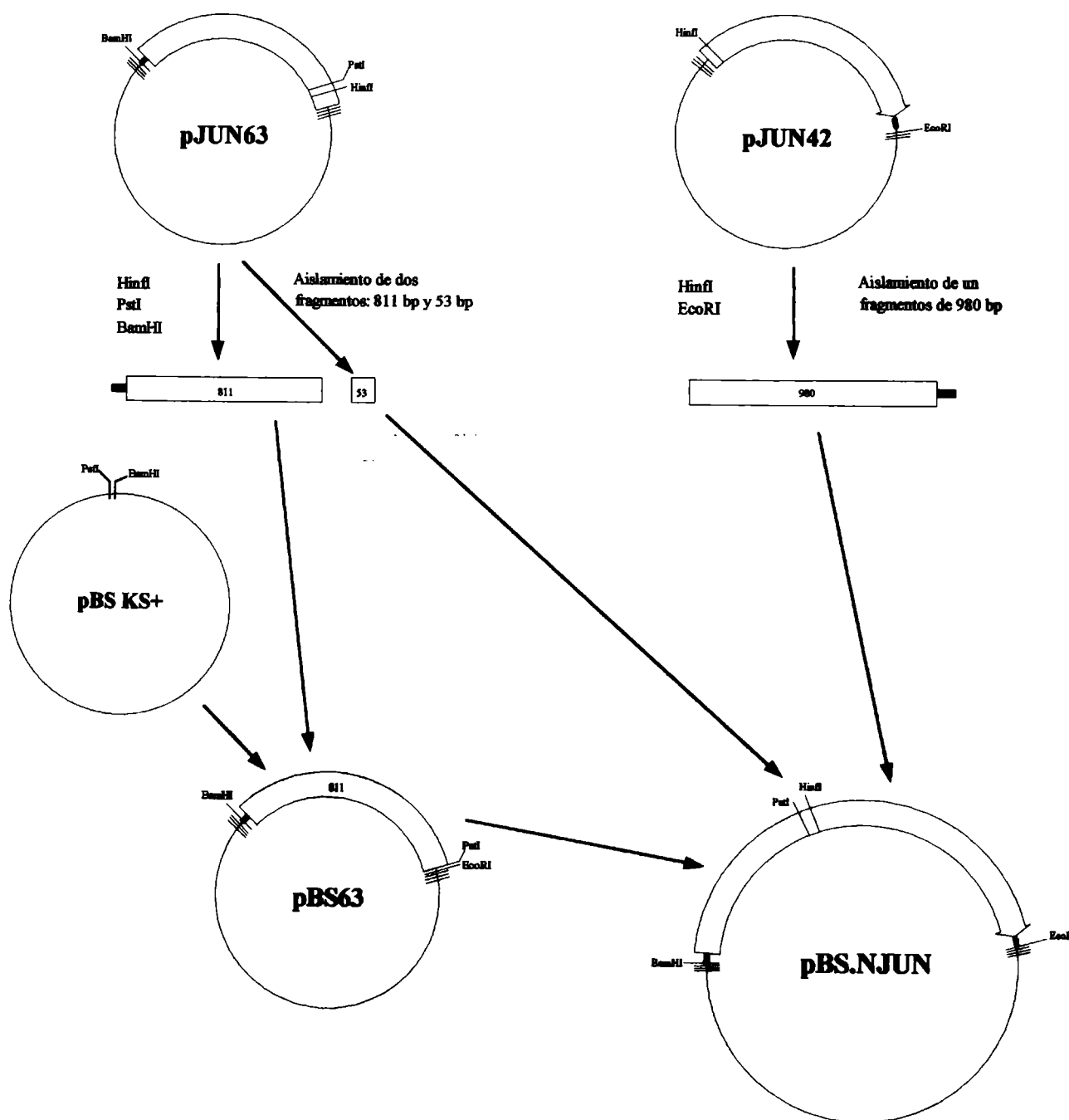


Figura 6.5. Reconstrucción del gen N de la cepa MC2 del virus Junín. Se indican los sitios de restricción utilizados en la reconstrucción del gen N. Las regiones codificantes aparecen como flechas o rectángulos vacíos. En trazo grueso aparecen las secuencias virales no codificantes (realizado por la Lic. Eva Manzella).

Capítulo 7 Información molecular - Proteínas - Cepa MC2

GLICOPROTEINAS VIRALES

Estructura primaria del precursor de las glicoproteínas (GPC)

El marco de lectura abierto en la mitad 5' del RNA S (polaridad viral), delimitado por los codones AUG₈₈₋₉₀-UAA₁₅₃₁₋₁₅₃₃, posibilita la traducción de un polipéptido de 481 aminoácidos (Figura 7.1), que es glicosilado y procesado postraduccionalmente.

En la proteína GPC pueden definirse tres

dominios: el péptido señal, la glicoproteína G1 y la glicoproteína G2.

Estructura del péptido señal

El péptido señal de MC2 tiene una longitud de 54 aminoácidos, concordando con los datos de Buchmeier *et al.* (1987) y Burns & Buchmeier (1993).

El mismo está constituido por 42,6 % de aminoácidos hidrofóbicos, 20,3 % de aminoácidos hidrofílicos y 37,1 % de aminoácidos ambivalentes (Tabla 7.a).

MGQFISFMQE	IPTFLQEALN	<u>I</u> ALVAVSLIA	I	I	K	G	V	V	N	L	Y	K	S	G	C	S	I	L	D	L	A	G	50								
RSCPRAFKIG	LHTEVPDCVL	LQWWVSFSNN	PHDLPLLCTL																				100								
<u>N</u> ASF	KISFDD	IAVLLPEYDV	IIQHPADMSW	CSKSDDQIRL	SQWFMNAVGH																		150								
DWYLDPPFLC	<u>R</u> NRT	KTEGFI	FQV	<u>N</u> T	S	K	T	G	I	NENYAKKFKT	GMHHLREYYP												200								
DSCLDGKLC	L	MKAQPTSWPL	QCPLDHVNTL	HFLTRGKNIQ	L	P	R	R	S	L	K	A	F	F									250								
SWSLTDSSGK	DTPGGYCLEE	WMLVAAKMKC	FGNTAVAKCN	LNHDSEFCDM																			300								
LRLFDYNKNA	IKTLNDETKK	QVNLMGQTIN	ALISDNLLMK	NKIRELMSVP																			350								
Y		KFWYV	<u>N</u> H	T	L	S	G	Q	H	S	L	P	R	C	W	L	I	K						400							
EMLSKEYSDR	QGKTPLTLVD	<u>I</u> C	F	W	S	T	V	F	F	E	T	<u>A</u> S	L	F	L	H	L	V	G	I	P	T	H	R	H	I	R	G	E	A	450
CPLPHRLNSL	GGCRCGKYPN	LKKPTVWRRG	H																												481

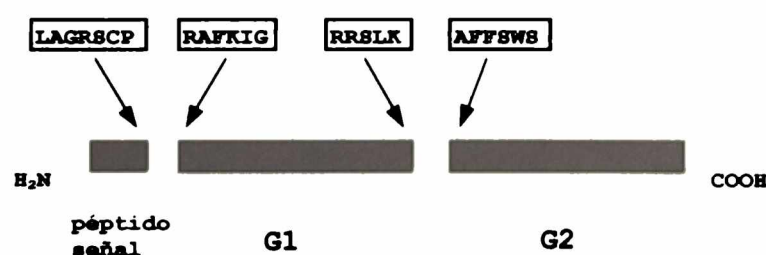


Figura 7.1. Precursor de las glicoproteínas (GPC) de la cepa MC2 del virus Junín. La secuencia aminoacídica fue deducida a partir de la secuencia nucleotídica mostrada en el capítulo 6 (Figura 6.1). Con recuadros, se indican los sitios potenciales de N-glicosilación (NX{T/S}, Bause, 1983), en grisado aquellos que están conservados (en posición) en todos los arenavirus (ver Capítulo 12). Con una barra se señala el sitio de reconocimiento de la proteasa de origen celular. Con un doble subrayado se indican las dos hélices transmembrana predichas (en el péptido señal y en G2). En el esquema se muestran los tres productos proteicos que se generan por procesamiento postraduccional de la proteína GPC. Los sitios de clivaje proteolítico y las secuencias correspondientes han sido determinados por homología con LCM (Buchmeier *et al.*, 1987; Burns and Buchmeier, 1993).

Tabla 7.a. Composición aminoacídica del péptido señal y las glicoproteínas G1 y G2

Aminoácido	Péptido señal		G1		G2	
	Cant.	Frecuencia	Cant.	Frecuencia	Cant.	Frecuencia
Ala (A)	5	9,3 %	7	3,6 %	9	3,8 %
Arg (R)	1	1,9 %	8	4,1 %	11	4,7 %
Asn (N)	2	3,7 %	11	5,7 %	18	7,7 %
Asp (D)	1	1,9 %	13	6,7 %	12	5,1 %
Cys (C)	2	3,7 %	7	3,6 %	10	4,3 %
Gln (Q)	3	5,6 %	8	4,1 %	4	1,7 %
Glu (E)	2	3,7 %	5	2,6 %	9	3,8 %
Gly (G)	4	7,4 %	9	4,7 %	13	5,6 %
His (H)	0	0,0 %	9	4,7 %	8	3,4 %
Ile (I)	7	13,0 %	10	5,2 %	11	4,7 %
Leu (L)	7	13,0 %	23	11,9 %	28	12,0 %
Lys (K)	2	3,7 %	14	7,2 %	17	7,3 %
Met (M)	2	3,7 %	4	2,1 %	7	3,0 %
Phe (F)	3	5,6 %	10	5,2 %	12	5,1 %
Pro (P)	2	3,8 %	12	6,2 %	9	3,8 %
Ser (S)	5	9,3 %	13	6,7 %	18	7,7 %
Thr (T)	1	1,9 %	10	5,2 %	14	6,0 %
Trp (W)	0	0,0 %	6	3,1 %	7	2,9 %
Tyr (Y)	1	1,9 %	6	3,1 %	8	3,4 %
Val (V)	4	7,4 %	8	4,1 %	9	3,8 %

Agrupamiento según Karlin *et al.* (1990): en rojo los hidrofóbicos, en negro los hidrofílicos y en fucsia los ambivalentes (no son ni fuertemente hidrofílicos ni fuertemente hidrofóbicos)

El estudio de los perfiles de hidropatía, accesibilidad promedio y accesibilidad individual (PHDacc, Rost and Sander, 1994c) permite encontrar una mayor concentración de aminoácidos hidrofóbicos en la región delimitada por los residuos 20 a 40 (Figura 7.2).

El análisis por computadora de la potencial estructura secundaria de la secuencia aminoacídica particular, realizado con diferentes algoritmos, permite efectuar predicciones que resultan coincidentes con los datos experimentales conocidos para otros arenavirus.

De acuerdo al algoritmo de Klein *et al.* (1985) el dominio en cuestión corresponde a un polipéptido integral con un segmento transmembrana delimitado por los aminoácidos 20 a 36. En función del algoritmo de Rao & Argos (1986) la predicción correspondiente determina la existencia de una α hélice transmembrana entre los aminoácidos 18 a 36. Por último, algoritmos más actualizados y refinados, como el PHDhtm (Rost *et al.*, 1995) y el PHDsec (Rost and Sander, 1993a,b,c, 1994a,b,d), predicen la existencia de una α hélice transmembrana entre los aminoácidos 19

a 47 (PHDhtm) o dos α hélices distanciadas un aminoácido (12 a 41 y 43 a 48) (PHDsec).

Como se puede ver, los datos de predicción son coherentes entre si y, además, coinciden con los requerimientos estructurales mínimos de un péptido señal.

Estructura de las glicoproteínas G1 y G2

Considerando el sitio de reconocimiento para la proteasa de origen celular, caracterizado en LCM (Buchmeier *et al.*, 1987, Burns & Buchmeier, 1993) y en el virus Junín (Padula *et al.* (1996), que da origen a las glicoproteínas G1 y G2, se pueden definir los dominios correspondientes en MC2, con 193 y 234 aminoácidos respectivamente.

La glicoproteína G1 está constituida por 28,6 % de aminoácidos hidrofóbicos, 35,1 % de aminoácidos hidrofílicos y 36,3% de aminoácidos ambivalentes, mientras que la glicoproteína G2 está constituida por 28,6 % de aminoácidos hidrofóbicos, 33,8 % de aminoácidos hidrofílicos y 37,6 % de aminoácidos ambivalentes. La comparación numérica anterior, o un análisis más detallado (Tabla 7.a), no muestran diferencias significativas entre G1 y G2.

El estudio de la distribución de aminoácidos hidrofóbicos, mediante perfiles de hidropatía y accesibilidad muestra diferencias concretas entre las glicoproteínas G1 y G2.

relativamente regular de aminoácidos hidrofóbicos y no hidrofóbicos, lo cual redonda en una mayor probabilidad promedio de exposición al solvente y sugiere que podría ser

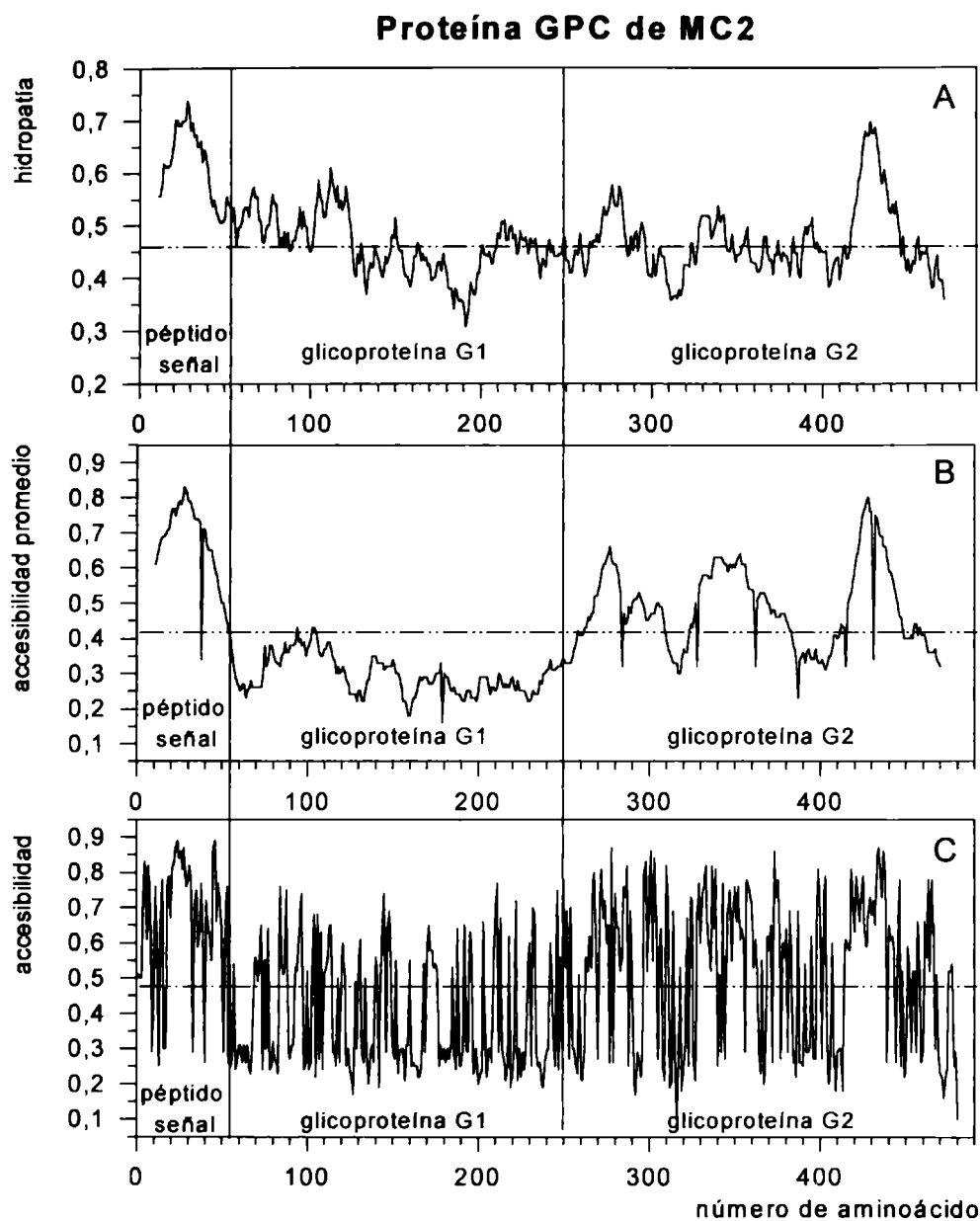


Figura 7.2. Análisis de hidropatía y accesibilidad de la proteína GPC de MC2. Las líneas verticales marcan las regiones de cada gráfico correspondientes a los dominios: péptido señal, glicoproteína G1 y glicoproteína G2. Las líneas quebradas corresponden a los valores promedio para la proteína completa. **A:** perfil de hidropatía calculado con el programa AASCALE (PC/Gene™, IntelliGenetics Inc.) en base a la tabla de hidrofobicidades de Kyte y Doolittle (1982), se utilizó una ventana de 21 residuos y los datos finales fueron normalizados a una escala de 0 a 1. **B y C:** perfiles de accesibilidad al solvente. Los datos de accesibilidad fueron calculados con el programa PHDacc (Rost and Sander, 1994c) y procesados posteriormente con el programa DATMANIP (Ghiringhelli, P.D., 2002, esta Tesis) para confeccionar ambos gráficos (ver Métodos y Miscelánea). Los valores fueron normalizados a una escala de 0 a 1, en donde a mayor valor menor accesibilidad. En **B** se representa la accesibilidad promedio calculada utilizando una ventana de 21 residuos. En **C** se representa la accesibilidad particular de cada residuo.

La proteína G1 tiene una distribución una proteína soluble (Figura 7.2).

En la proteína G2, en cambio, la distribución es menos regular, existiendo grupos de aminoácidos hidrofóbicos. En particular, es necesario destacar la región delimitada por los residuos 168 a 198 (415-445 en GPC), rica en aminoácidos hidrofóbicos y con muy escasa probabilidad de exposición al solvente, las cuales son características típicas de una región transmembrana. Este conjunto de datos sugiere que G2 podría ser una proteína asociada a membranas (Figura 7.2).

Otra diferencia significativa entre las glicoproteínas G1 y G2 radica en la distribución de prolinas y glicinas (Figura 7.3), lo cual sugiere que el plegamiento de ambas proteínas podría ser distinto. En G1, la distribución de prolinas es aproximadamente uniforme, con sólo 3 regiones de rigidez relativa baja (residuos 32-61, 72-101 y 104-145), mientras que la distribución de glicinas es asimétrica, con una mayor concentración en la mitad carboxilo terminal. En G2, en cambio, la distribución de prolinas es muy asimétrica y la de glicinas es aproximadamente uniforme, definiendo 3 regiones de menor rigidez (residuos 17-102, 125-167 y 169-193)

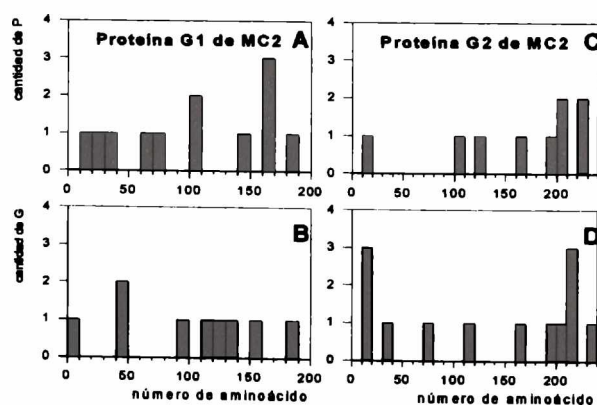


Figura 7.3. Distribución de prolinas y glicinas en las glicoproteínas G1 y G2. Se graficó la cantidad del residuo particular cada 10 aminoácidos. **A:** distribución de prolinas en G1. **B:** distribución de glicinas en G1. **C:** distribución de prolinas en G2. **D:** distribución de glicinas en G2.

El análisis por computadora de la potencial estructura secundaria de la glicoproteína G1 señala a ésta como una proteína periférica (Klein *et al.*, 1985), con 3 región de α hélice, 8 regiones β hoja plegada y 12 regiones definidas como *loops*, donde podrían existir segmentos con estructura *random coil*, β -turn y/o γ -turn (PHDsec, Rost and Sander, 1993a,b,c, 1994a,b,d) (Tabla 7.b).

El análisis por computadora de la potencial estructura secundaria de la glicoproteína G2 señala a esta como una proteína integral multimérica, con un segmento transmembrana delimitado por los aminoácidos 171 a 188 (Klein *et al.*, 1985), 175 a 195 (Eisenberg *et al.*, 1984a), 169 a 193 (Rao and Argos, 1986) o 175 a 190 (PHDhtm, Rost *et al.*, 1995). Las predicciones del PHDsec (Rost and Sander, 1993 a, b c, 1994 a, b, d) revelan la existencia de 4 regiones de α hélice, 4 regiones β hoja plegada y 8 regiones definidas como *loops*, donde podrían existir segmentos con estructura *random coil*, β -turn y/o γ -turn (Tabla 7.c).

La búsqueda en la base de datos PROSITE (*PROtein SITES*, Fuchs, 1991; Bairoch, 1993) predice la existencia de varios sitios potenciales de glicosilación, fosforilación, etc., tanto en G1 como en G2 (Tabla 7.d). Hasta el momento, no existen datos experimentales que demuestren la fosforilación o miristoilación en G1 y/o G2, o la sulfatación en G2, por lo cual se desconoce si estos sitios son biológicamente significativos. Con respecto a los sitios de N-glicosilación, se conoce que son utilizados, pero no se sabe cuántos ni cuáles de los 8 sitios potenciales están glicosilados *in vivo*.

Tabla 7.b. Estructuras secundarias predichas para la glicoproteína G1 de MC2

α hélices		
n°	posición	longitud
$\alpha 0$	13-23	11
$\alpha 1$	82-94	13
$\alpha 2$	127-136	10
β hojas plegadas		
n°	Posición	longitud
$\beta 2-0$	33-36	4
$\beta 2-1$	40-44	5
$\beta 2-2$	49-53	5
$\beta 2-3$	57-61	5
$\beta 2-4$	66-69	4
$\beta 3-1$	113-122	10
$\beta 4-1$	155-159	5
$\beta 4-2$	176-180	5
loops		
n°	Posición	longitud
L1	1-12	12
L2	24-32	10
L3	37-39	3
L4	45-48	4
L5	54-56	3
L6	62-65	4
L7	70-81	12
L8	95-112	18
L9	123-126	4
L10	137-154	18
L11	160-175	16
L12	181-193	11

La numeración de las estructuras α y β está basada en la comparación entre las G1 de todos los arenavirus (ver Capítulo 13)

Tabla 7.c. Estructuras secundarias predichas para la glicoproteína G2 de MC2

α hélices		
n°	posición	longitud
$\alpha 1$	19-33	15
$\alpha 2$	50-100	51
$\alpha 3$	142-163	22
$\alpha 4$	170-191	22
β hojas plegadas		
n°	posición	longitud
$\beta 1$	3-6	4
$\beta 2$	37-41	5
$\beta 3$	108-114	7
$\beta 4$	125-129	5
loops		
n°	posición	longitud
L1	7-18	12
L2	34-36	3
L3	42-49	8
L4	101-107	7
L5	116-124	9
L6	130-141	12
L7	164-169	6
L8	192-234	43

La numeración de las estructuras α y β está basada en la comparación entre las G2 de todos los arenavirus (ver Capítulo 13)

Tabla 7.d. Sitios potenciales identificados en las glicoproteínas G1 y G2

Sitio	G1		G2		Código Prosite
	Cant.	Posición	Cant.	Posición	
N-glicosilación	4	91, 101, 162, 174	4	353, 361, 378, 383	PDOC00001
{S/T} fosforilación (proteínquinasas y caseínquinasas)	5	103, 107, 132, 164, 175	10	202, 245, 253, 248, 313, 318, 408, 417, 442, 475	PDOC00004, PDOC00005 y PDOC00006
N-miristoilación	2	99, 179	3	326, 461, 462	PDOC00008

PROTEINA DE LA NUCLEOCAPSIDE (N)

Estructura de la proteína N

El marco de lectura abierto de polaridad viral complementaria existente en la mitad 3' del RNA S, delimitado por los codones AUG₈₂₋₈₄-UAA₁₇₇₄₋₁₇₇₆ (numeración VC), codifica para un polipéptido de 564 aminoácidos (Figura 7.4) y un peso molecular estimado de 63033 Da.

La proteína N está constituida por 28,3 % de aminoácidos hidrofóbicos, 36,7 % de aminoácidos hidrofílicos y 35,0 % de

aminoácidos ambivalentes, un análisis más detallado se presenta en la Tabla 7.e.

La carga neta estimada es de +14,5 a pH neutro, en clara dependencia de la relativa riqueza en Arg y Lys, aminoácidos que se hallan dispersos en toda la proteína N, con varios grupos de dos, tres y hasta cuatro residuos básicos (Figura 7.4). Esta característica está en relación con la función estructural de la proteína N de asociarse con el RNA viral para formar las nucleocápsides.

MAHSKEVPSF	RWTQSL	GL	SQFTQTVKSD	VLKDAKLIAD	SIDFNQVAQV	50			
QRAL	ET	EG	EEDLNKLRDL	NKEVDRLMSM	RSVQRNTVFK	AGDLGRVERM	100		
ELASGLGNLK	TKF	AETGS	QGVYMGNLSQ	SQLA	SEIL	RTLGFQQQGT	150		
GGNGVVRVWD	VKDPSKLNQ	FGSVPALTIA	CMTVQGGETM	NSVIQALTSL		200			
GLLYTVKYPN	LSDLDRLTQE	HDCLQIVTKD	ESSINISGYN	FSLSAAVKAG		250			
ASILDDGNML	ETIRVTPDNF	SSLIKSTIQV	EGMFIDE	KPGNRNPYEN		300			
LLYKLCLSGD	GWPYIGSR	SR	LI	GR	SWDNTS	IDLT	PVAG	PRQPEKNGQN	350
LRLANLTEIQ	EAVIREAVGK	LDPTNTLWLD	IEGPATDPVE	MALFQPAGSK		400			
YIHCF	PHD	EKGFKNGSRH	SHGILMKDIE	DAMPGVLSYV	IGLLPPDMVV	450			
TTQGSDDI	LFDLHG	RD	KLVDVRLTSE	QARQFDQQVW	EKFGHLCKHH	500			
NGVVVS	DKDAPFKLAS	SEPHCALLDC	IMFQSVLDGK	LYEEELTPLL		550			
PPSLLFLPKA	AYAL					564			

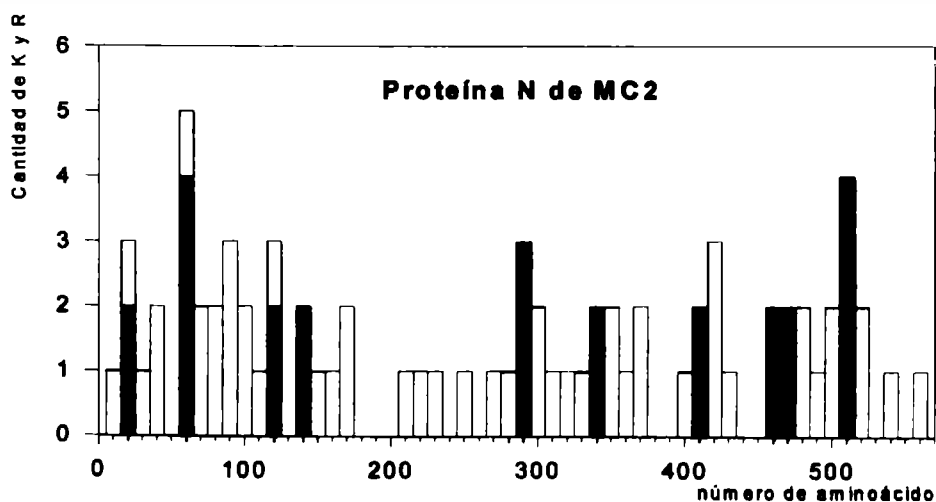


Figura 7.4. Proteína de la nucleocápside (N) de la cepa MC2 del virus Junín. La secuencia aminoacídica fue deducida a partir de la secuencia nucleotídica mostrada en el capítulo 6 (Figura 6.1). Con recuadros grisados se destacan los conjuntos de aminoácidos básicos Arg (R) y Lys (K). El doble subrayado indica la señal de dimerización predicha. El gráfico de barras muestra la cantidad de R y K cada 10 aminoácidos. Las barras blancas indican el total de los mismos. Las barras grisadas superpuestas corresponden a la fracción representada por uno o más conjuntos de 2, 3 o 4 aminoácidos básicos.

Tabla 7. e. Composición aminoacídica de la proteína N

Aminoácido	Cantidad	Frecuencia	Aminoácido	Cantidad	Frecuencia
Ala (A)	31	5,4 (%)	Leu (L)	65	11,5 (%)
Arg (R)	36	6,3 (%)	Lys (K)	40	7,0 (%)
Asn (N)	25	4,4 (%)	Met (M)	14	2,4 (%)
Asp (D)	38	6,7 (%)	Phe (F)	19	3,3 (%)
Cys (C)	7	1,2 (%)	Pro (P)	25	4,4 (%)
Gln (Q)	30	5,3 (%)	Ser (S)	43	7,6 (%)
Glu (E)	29	5,1 (%)	Thr (T)	29	5,1 (%)
Gly (G)	42	7,4 (%)	Trp (W)	6	1,0 (%)
His (H)	11	1,9 (%)	Tyr (Y)	11	1,9 (%)
Ile (I)	26	4,6 (%)	Val (V)	37	6,5 (%)

Agrupamiento según Karlin *et al.* (1990): en rojo los hidrofóbicos, en negro los hidrofílicos y en fucsia los ambivalentes (no son ni fuertemente hidrofílicos ni fuertemente hidrofóbicos)

Proteína N de MC2

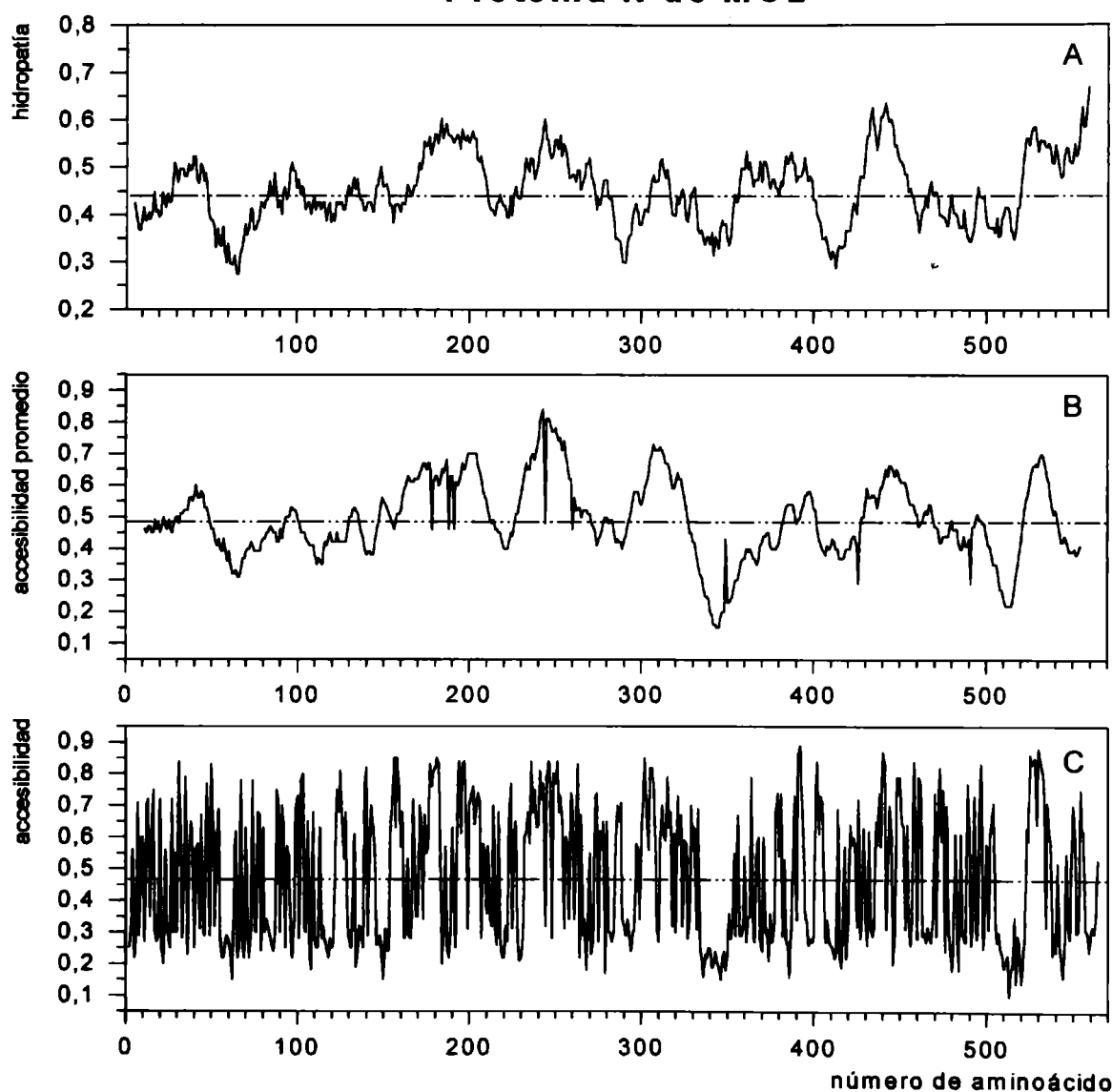


Figura 7.5. Análisis de hidropatía y accesibilidad de la proteína N de MC2. Las líneas quebradas corresponden a los valores promedio para la proteína completa. **A:** perfil de hidropatía calculado con el programa AASCALE (PC/Gene™, IntelliGenetics Inc.) en base a la tabla de hidrofobicidades de Kyte y Doolittle (1982), se utilizó una ventana de 21 residuos y los datos finales fueron normalizados a una escala de 0 a 1. **B y C:** perfiles de accesibilidad al solvente. Los datos de accesibilidad fueron calculados con el programa PHDacc (Rost and Sander, 1994c) y procesados posteriormente con el programa DATMANIP (Ghiringhelli, P.D., esta Tesis) para confeccionar ambos gráficos (ver Métodos y Miscelánea). Los valores fueron normalizados a una escala de 0 a 1, en donde a mayor valor menor accesibilidad. En **B** se representa la accesibilidad promedio calculada utilizando una ventana de 21 residuos. En **C** se representa la accesibilidad particular de cada residuo.

El análisis del perfil de hidropatía (**Figura 7.5**) muestra varias regiones de alta hidrofobicidad (residuos 28-48, 164-210, 231-273, 357-399, 426-456 y 520-564) y varias regiones de baja hidrofobicidad (residuos 49-83, 281-304, 332-353, 400-425, 472-494 y 496-519). Los perfiles de accesibilidad promedio y accesibilidad individual (**Figura 7.5**) permiten obtener resultados coincidentes con los datos de hidrofobicidad.

El estudio de la distribución de aminoácidos que pueden afectar directamente el plegamiento de la proteína, prolina (induce torsiones) y glicina (incrementa la flexibilidad), permite observar detalles estructurales interesantes (**Figura 7.6**). El análisis de la distribución de prolina señala la existencia de tres regiones de menor rigidez (residuos 9-163, 210-266 y 447-514), donde la ausencia de este residuo particular dificulta la formación de estructuras de tipo β -turn. La observación de la cantidad y distribución de glicinas sugiere que la flexibilidad de la cadena polipeptídica es grande.

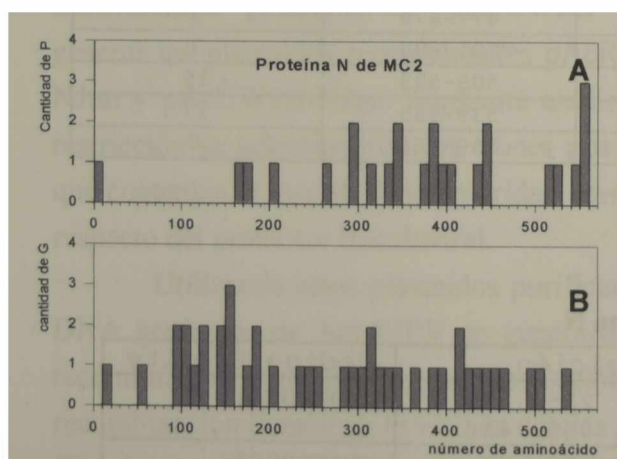


Figura 7.6. Distribución de prolina y glicinas en la proteína N. Se graficó la cantidad del residuo particular cada 10 aminoácidos. **A:** distribución de prolina. **B:** distribución de glicinas.

El estudio computacional de la potencial estructura secundaria de la proteína N indica la presencia de 17 regiones de α hélice, 17 regiones β hoja plegada y 35 regiones definidas como *loops*, donde podrían existir segmentos con estructura *random coil*, β -turn y/o γ -turn (PHDsec; Rost and Sander, 1993 a, b, c, 1994a, b, d) (**Tabla 7.f**).

La búsqueda en la base de datos PROSITE (Fuchs, 1991; Bairoch, 1993) predice la existencia de varios sitios potenciales de glicosilación, fosforilación, etc. (**Tabla 7.g**). Hasta el momento, no existen datos experimentales que demuestren la miristilación, la sulfatación o la amidación en N, por lo cual se desconoce si estos sitios son biológicamente significativos. Respecto a los sitios de N-glicosilación, se conoce que no son utilizados. Por otra parte, la señal de dimerización podría estar asociada con las interacciones proteína N-proteína N en la nucleocápside y/o con la interacción proteína N-proteína X (viral o celular) para el ejercicio de funciones regulatorias. Por último, en referencia a los sitios de fosforilación, no hay resultados experimentales concluyentes acerca de su utilización *in vivo*, pero existen evidencias que sugieren la presencia intracelular de la proteína N en dos estados, fosforilado (Bruns *et al.*, 1986) y no fosforilado. Esto último, podría estar en relación con posibles funciones regulatorias de la proteína.

Tabla 7.f. Estructuras secundarias predichas para la proteína N de MC2

α hélices			loops		
n°	posición	longitud	n°	posición	longitud
$\alpha 1$	11-19	9	L1	1-10	10
$\alpha 2$	27-39	13	L2	20-26	7
$\alpha 3$	45-55	11	L3	40-44	5
$\alpha 4$	64-79	16	L4	56-63	8
$\alpha 5$	96-113	18	L5	80-86	7
$\alpha 6$	129-143	15	L6	92-95	4
$\alpha 7$	176-185	10	L7	114-122	9
$\alpha 8$	193-200	8	L8	126-128	3
$\alpha 9$	241-254	14	L9	144-155	12
$\alpha 10$	271-282	12	L10	162-175	14
$\alpha 11$	356-370	12	L11	186-192	7
$\alpha 12$	426-431	6	L12	201-203	3
$\alpha 13$	437-444	8	L13	208-225	18
$\alpha 14$	457-465	9	L14	230-240	11
$\alpha 15$	480-500	21	L15	255-260	6
$\alpha 16$	524-538	14	L16	267-270	4
$\alpha 17$	556-558	3	L17	283-288	6
β hojas plegadas			L18	291-301	11
n°	posición	longitud	L19	309-313	5
$\beta 1$	87-91	5	L20	317-320	4
$\beta 2$	123-125	3	L21	324-330	7
$\beta 3$	156-161	6	L22	334-355	22
$\beta 4$	204-207	4	L23	371-377	7
$\beta 5$	226-229	4	L24	383-389	7
$\beta 6$	261-266	6	L25	396-401	6
$\beta 7$	288-290	3	L26	408-425	18
$\beta 8$	302-308	7	L27	432-436	5
$\beta 9$	314-316	3	L28	445-449	5
$\beta 10$	321-323	3	L29	454-456	3
$\beta 11$	331-333	3	L30	466-471	6
$\beta 12$	378-382	5	L31	477-479	3
$\beta 13$	390-395	6	L32	501-503	3
$\beta 14$	402-407	6	L33	509-523	15
$\beta 15$	450-453	4	L34	539-555	17
$\beta 16$	472-476	5	L35	559-564	6
$\beta 17$	504-508	5			

Tabla 7.g. Sitios potenciales identificados en la proteína N

Sitio	Cantidad	Posición	Código Prosite
N-glicosilación	8	127, 210, 235, 240, 269, 328, 355, 416	PDOC00001
{S/T} fosforilación (proteínquinasas y caseínquinasas)	14	9, 15, 26, 57, 79, 205, 212, 228, 252, 262, 329, 334, 506, 535	PDOC00005 y PDOC00006
N-miristilación	7	107, 122, 149, 172, 201, 417, 502	PDOC00008
Amidación	1	465	PDOC00009
Señal de dimerización (Helix-loop-helix)	1	317	PDOC00038
Y-fosforilación Tirosinquinasa	1	295	PDOC00007

Capítulo 8

Expresión del gen N de la cepa MC2 en el sistema baculovirus-células de insecto

Clonado del gen N en un vector de transferencia, generación de baculovirus recombinantes y análisis del producto de expresión

Como se muestra en la **Figura 8.1**, el gen N de la cepa MC2 del virus Junín fue subclonado a partir del plásmido recombinante pBS-NJUN (ver **Capítulo 6, Figura 6.5**), mediante digestión con las endonucleasas Bam HI y Sac II. El fragmento recuperado de un gel de agarosa fue tratado con Nucleasa S1 para generar extremos romos. Posteriormente, se le adicionaron *linkers* Bam HI en ambos extremos y el fragmento digerido con Bam HI fue ligado al vector pUC18. A partir de un clón positivo se recuperó el inserto mediante digestión con BamHI y se lo ligó a los vectores de transferencia pAcRP25 y pAcUW2B para generar los plásmidos recombinantes pAcRP25-NJun y pAcUW2B-NJun. Mediante análisis de restricción se seleccionaron los clones positivos que contenían el inserto en la polaridad correcta, respecto del promotor baculoviral.

Utilizando estos plásmidos purificados y DNA genómico de AcMNPV se generaron los recombinantes mediante cotransfección y recombinación homóloga *in vivo* en células *Sf*21 (*Spodoptera frugiperda* 21). Los recombinantes poliedrina negativos fueron aislados a partir de placas individuales y clonados mediante tres rondas sucesivas de purificación por plaqueo. Los diferentes recombinantes obtenidos fueron amplificados mediante infección de monocapas de células *Sf* 21, para generar los stocks correspondientes.

Utilizando dos de los stocks de AcNPV obtenidos por recombinación homóloga con la construcción pAcUW2B.N: pAcUW2B.N811 y

pAcUW2B.N812, se infectaron células de *Sf* 21 en monocapa, con una multiplicidad de infección $\sim 0,1$ ufp/ml. Los cultivos desarrollaron a 28°C en medio TC-100 (GIBCO BRL, Gaithersburg, MD, USA) - 10% de suero fetal bovino. A los 5 días postinfección fueron recolectados los sobrenadantes, centrifugados 10 min. a 10000 rpm en una centrifuga Sorvall RC5 para eliminar restos celulares, y titulados sobre células *Sf*21.

Con ambos recombinantes amplificados se infectaron células *Sf* 21 en monocapa con distintas multiplicidades de infección. A las 20, 40 y 60 horas postinfección se recolectaron las células infectadas se procesaron de acuerdo a lo descrito en **Métodos (Capítulo 16)** y se resolvieron mediante electroforesis en geles de poliacrilamida desnaturalizantes (**Figura 8.2**).

También se realizó un análisis de las variaciones en rendimiento según la metodología de cultivo. Para ello, se infectaron cultivos en suspensión y cultivos en monocapa con una m.d.i.= 10 (multiplicidad de infección=10), y se incubaron a 28°C durante 60 hs. El análisis y cuantificación se realizó por SDS-PAGE (datos no mostrados). Los rendimientos obtenidos en la producción de proteína N oscilaron entre 20-50 $\mu\text{g/ml}$ a partir de 10^6 células, según la metodología utilizada para el cultivo (suspensión o monocapa, respectivamente).

La identidad de la proteína expresada fue confirmada mediante un ensayo de *Western blot* (**Figura 8.3**), revelado con una mezcla de anticuerpos monoclonales anti-N de Junín (Sánchez *et al.*, 1989).

Por último, se efectuó un análisis de la solubilidad de la proteína producida, mediante interrupción por sonicación de las células

infectadas y separación de las fracciones soluble e insoluble. Como se puede observar en la **Figura 8.4**, la mayor parte de la proteína N se acumula en forma insoluble.

Empleo del producto de expresión en ensayos de ELISA

El lisado celular semipurificado de cultivos de células *Sf* 21 infectadas con el recombinante Ac-N811 fue utilizado en el desarrollo de un test de ELISA para la detección de anticuerpos circulantes. El sistema utilizado para el revelado fue el de IgG conjugada a peroxidasa y *O*-fenilendiamina (OPD) como sustrato. Las condiciones del ensayo fueron optimizadas en todos sus parámetros: elección de buffer de bloqueo, determinación de la concentración de antígeno, dilución de suero óptima, dilución del conjugado, tiempo de reacción y concentración de H₂O₂ en el buffer de revelado, etc. (datos no mostrados).

En la **Figura 8.5** se pueden observar los resultados obtenidos en un ensayo de ELISA utilizando el antígeno recombinante y suero de cobayos, infectados y no infectados.

Por último, en colaboración con la Lic. Eva María Manzella, se realizaron ensayos de ELISA con sueros humanos, normales y de pacientes. Los datos obtenidos sobre 45 sueros normales y 49 sueros de pacientes (con diagnóstico de FHA) (datos no mostrados), permiten afirmar que se cuenta con una herramienta recombinante específica, la proteína N. La misma, se puede utilizar de manera equivalente al antígeno viral *wild type*, y no implica el riesgo de trabajar con material patógeno para obtenerlo y utilizarlo en los ensayos de ELISA (García Franco *et al.*, 1988).

Empleo del producto de expresión en ensayos de replegamiento y recuperación de inmunoreactividad

Como se mostró en la **Figura 8.4** la mayor parte de la proteína N recombinante se encuentra en la fracción insoluble de los lisados

de células infectadas. Este es un fenómeno relativamente común en la producción de proteínas recombinantes, especialmente cuando se sobreproducen en forma significativa.

Los agregados están constituidos por múltiples moléculas de la proteína N interactuando entre sí; además, puede haber también, alguna proteína celular o baculoviral.

Los agregados de proteína N sólo se pueden desarmar en condiciones altamente desnaturizantes: 2% SDS, 8 M urea o 6 M clorhidrato de guanidinio. En estas condiciones, la proteína N permanece soluble a temperatura ambiente durante bastante tiempo. Sin embargo, es imposible pensar en realizar ningún ensayo de funcionalidad de la proteína N recombinante en estas condiciones.

Para superar este inconveniente, se desarrolló un método que permite desarmar los agregados y recuperar, aunque sea parcialmente, el plegamiento original. Como se describe en **Métodos (Capítulo 16)** el procedimiento desarrollado, sobre proteínas inmovilizadas en membrana, permite recuperar la reactividad con anticuerpos *Western* -, que reconocen epitopes de tipo conformacional (**Figura 8.6**). De esta manera, se puede asumir que la proteína N recombinante ha recuperado, al menos en parte, la conformación original.

Empleo del producto de expresión en ensayos de unión a RNA

Asumiendo que, durante el ciclo de infección viral, la proteína N puede tener, al menos, dos funciones: una estructural (formación de las nucleocápsides) y otra regulatoria (antiterminación de la transcripción), debería existir una afinidad diferencial de la proteína N, tanto por la región *panhandle* (extremos 5' y 3' virales en estructura doble cadena), como por la región intergénica con las estructuras tipo *hairpin loop* formadas.

Para comprobar la capacidad de unión de la proteína N a estas regiones del RNA viral de la cepa MC2 del virus Junín, se generaron las

sondas correspondientes mediante transcripción *in vitro* de plásmidos recombinantes preparados *ad hoc*. Como se muestra en las Figuras 8.7 y 8.8, la proteína N recombinante posee capacidad de unión a las diferentes sondas de RNA utilizadas. Es importante destacar que, si las membranas no se someten al proceso de replegamiento, la capacidad de unión a las sondas de RNA es nula (datos no mostrados).

Empleo del producto de expresión en ensayos de unión a Zinc

En 1989, Salvato y Shimomaye, observan, indirectamente, que la proteína N del virus LCM tiene capacidad de unión a zinc en un ensayo de *zinc blot*. Un análisis comparativo de las proteínas N (ver Capítulo 14, Figuras 14.9 y 14.10), permite predecir que las proteínas N de todos los arnavirus descritos podrían tener la capacidad de unir zinc *in vivo* e *in vitro*.

Para comprobar esto, se realizó un ensayo de *zinc blot* con extractos proteicos de viriones de 4 arnavirus (LCM, Junín MC2, Tacaribe y Pichindé) (Figura 8.9). Los resultados obtenidos confirman la predicción teórica. Dado que en este experimento, no se observa si las proteínas recombinantes también presentan la capacidad de unión a zinc, se repitió el ensayo sólo con las proteínas recombinantes, sin o con tratamiento de replegamiento. Los resultados obtenidos (Figura 8.10) muestran que la proteína N del virus Junín MC2 retiene la capacidad de unión a zinc, con o sin tratamiento de replegamiento, mientras que la proteína N del virus LCM no parece unir zinc, a menos que se realice el tratamiento de replegamiento. Esta diferencia puede deberse a la presencia de algunas histidinas adicionales en el virus Junín, previas al sitio predicho de unión de zinc, no existentes en LCM (ver Apéndice 7, Figura Ap7.4). Estos últimos ensayos se repitieron varias veces, obteniendo resultados equivalentes (datos no mostrados).

Empleo del producto de expresión en ensayos de unión a RNA en presencia o ausencia de Zinc

Teniendo en cuenta que la proteína N tiene una carga neta positiva a pH neutro (+14,5, Capítulo 7, página 57) y que una de las posibilidades para la interacción con el RNA puede radicar en interacciones de tipo polar, en el Capítulo 7 (Figura 7.4) se analiza la distribución de los aminoácidos cargados positivamente. En esta distribución, hay dos grupos muy importantes, tanto en el extremo amino- como en el extremo carboxilo-terminal.

Por otra parte, un análisis detallado de la distribución de cisteínas e histidinas capaces de coordinar el zinc en un motivo tipo *zinc finger* muestra que los únicos residuos C y H con una disposición capaz de coordinar zinc de esta forma, se encuentran flanqueando el conjunto de aminoácidos cargados de la región carboxilo-terminal (ver Capítulo 14, Figura 14.11).

En función de esto, se realizó un ensayo de interacción entre la proteína N recombinante y una sonda de RNA correspondiente a la región *panhandle*, incluyendo (o no) 20 μM ZnCl_2 durante el tratamiento de replegamiento. Los resultados obtenidos (Figura 8.11), claramente muestran que la capacidad de unión de la proteína N a la sonda de RNA se ve afectada por la presencia de zinc durante el tratamiento de replegamiento. Mientras que las cantidades de proteína inmovilizadas en las membranas corresponden a diluciones seriadas con un factor de 5, el porcentaje de cpm retenidas es muy diferente en ambos paneles. En la Figura 8.11 (panel A), se puede observar que la cantidad de sonda retenida es directamente proporcional a la cantidad de proteína inmovilizada; mientras que en la Figura 8.11 (panel B), se puede observar que la cantidad de sonda retenida no es proporcional. Esto sugiere que el replegamiento obtenido en presencia o ausencia de zinc es equivalente, pero no idéntico.

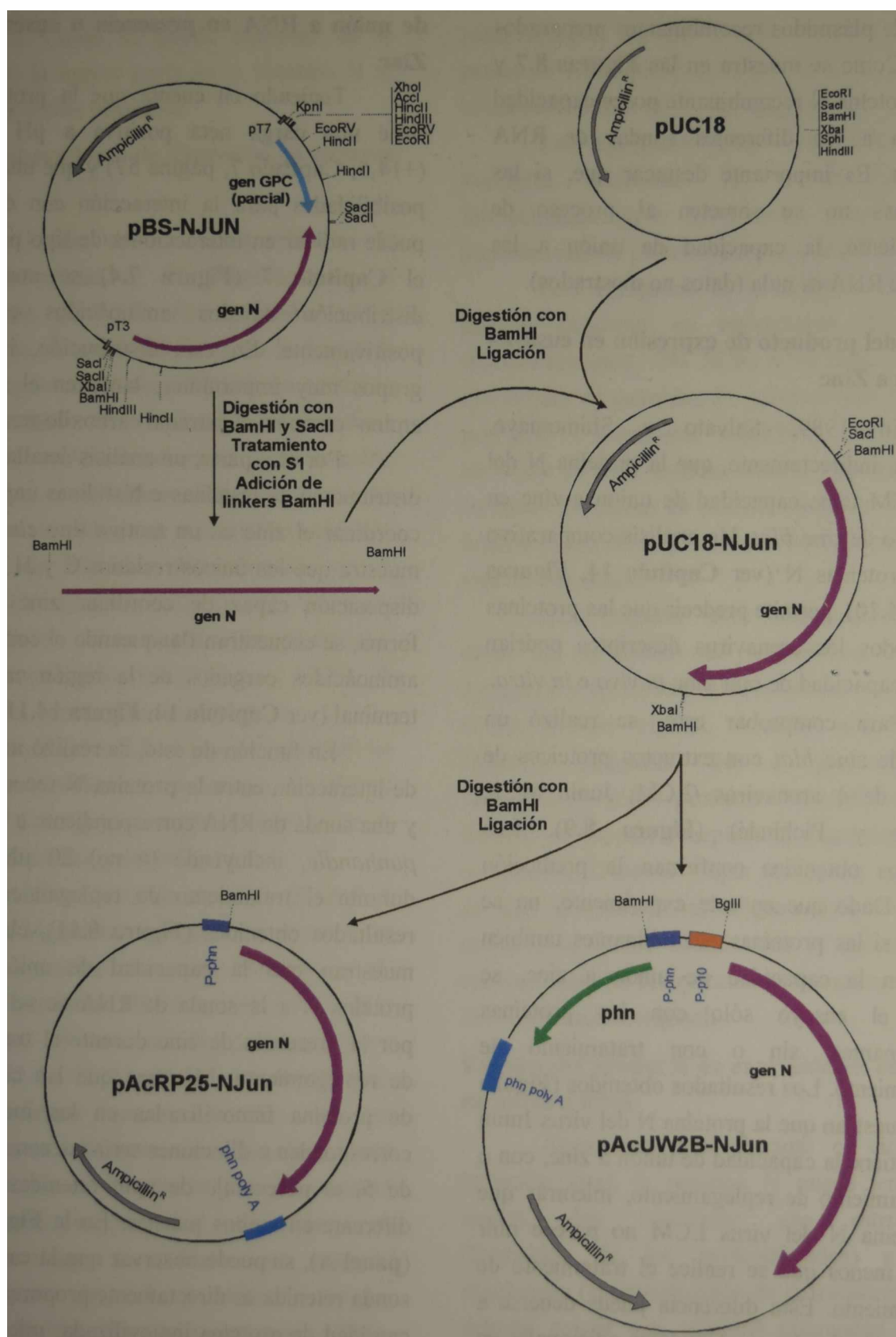


Figura 8.1. Construcción de los plásmidos de transferencia pAcRP25-Njun y pAcUW2B-Njun. En el esquema se indican sólo los sitios de restricción relevantes para la estrategia de construcción de los plásmidos recombinantes. El gen *N* está representado como una flecha de color violeta y el gen *phn* está representado con una flecha verde. Los promotores de *poliedrina* y *p10* se muestran con rectángulos, azul y naranja, respectivamente.

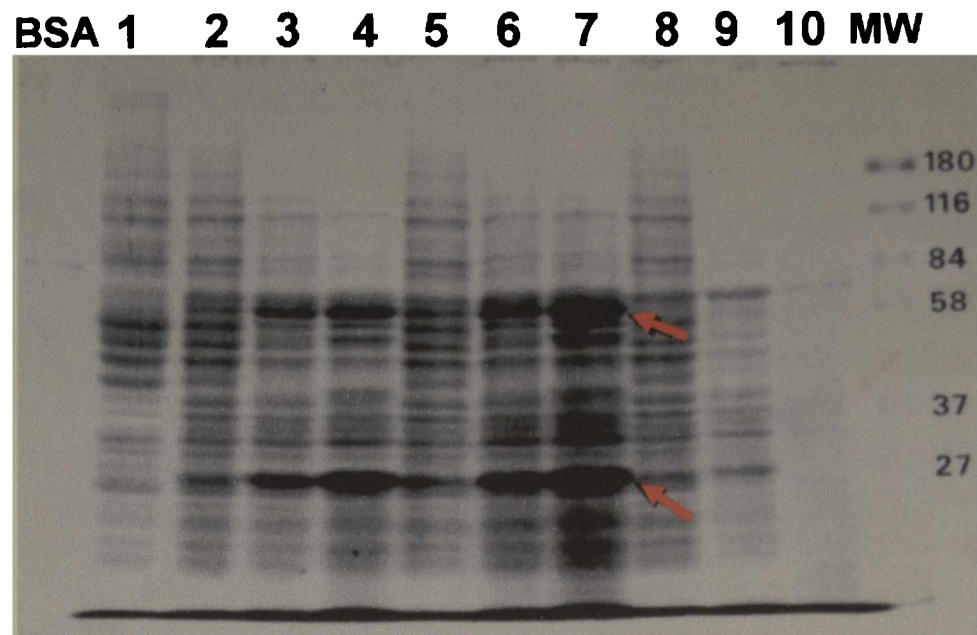


Figura 8.2. Cinética de expresión de la proteína N del virus Junín MC2 en células de insecto. Con flechas rojas se indican la proteína N (~63 kDa) y la poliedrina (~29 kDa). MW: Se indican los pesos moleculares de las distintas bandas. **Calle 1:** Proteínas totales de un lisado de células Sf 21 no infectadas. **Calles 2, 3 y 4:** Proteínas totales de un lisado de células Sf 21 infectadas con el recombinante Ac.N811. Las alícuotas fueron tomadas a las 20, 40 y 60 hs. postinfección, respectivamente. **Calles 5, 6 y 7:** Proteínas totales de un lisado de células Sf 21 infectadas con el recombinante Ac.N812. Las alícuotas fueron tomadas a las 20, 40 y 60 hs. postinfección, respectivamente. **Calles 8, 9 y 10:** Proteínas totales de un lisado de células Sf 21 infectadas con AcMNPV *wild type*. Las alícuotas fueron tomadas a las 20, 40 y 60 hs. postinfección, respectivamente.

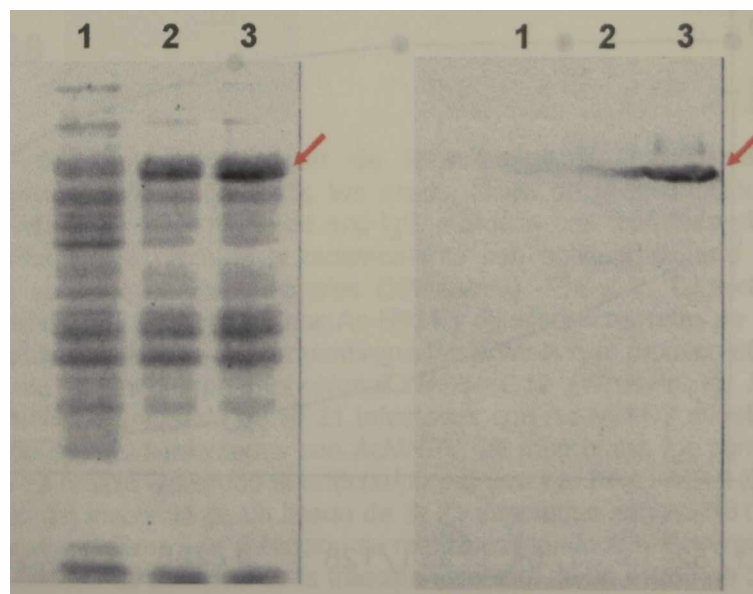


Figura 8.3. Confirmación mediante *Western blot* de la identidad de la proteína N del virus Junín MC2 expresada en células de insecto infectadas con el recombinante Ac-N811. Con una flecha roja se indica la proteína N (~63 kDa). Proteínas totales de un lisado de células Sf 21 infectadas con el recombinante Ac.N811. Alícuotas tomadas a las 24 hs (calle 1), 48 hs (calle 2) y 65 hs (calle 3) postinfección. **Panel izquierdo:** SDS-PAGE. **Panel derecho:** *Western blot* revelado con un anticuerpo monoclonal específico.

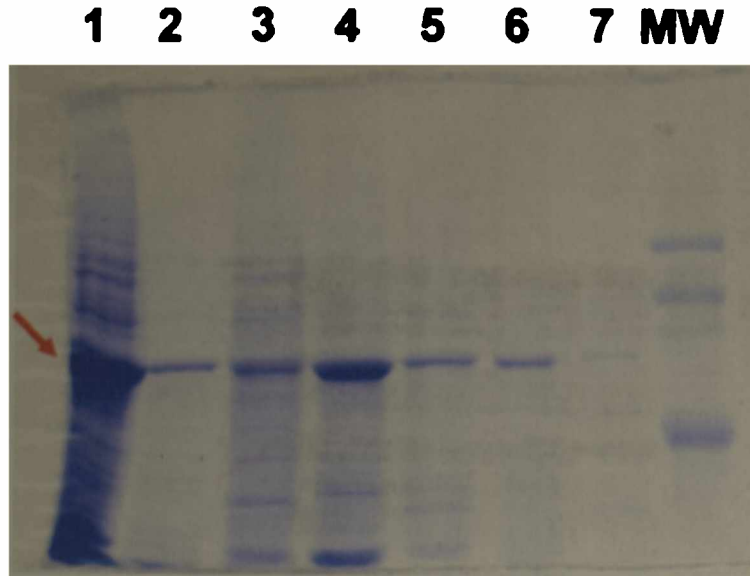


Figura 8.4. Fraccionamiento del lisado de células infectadas con el recombinante Ac-N811. Con una flecha roja se indica la proteína N (~63 kDa). MW: marcador de peso molecular (fosforilasa b, albúmina bovina y albúmina de huevo). Calle 1: Proteínas totales de un lisado de células Sf 21 infectadas con el recombinante Ac.N811. Calles 2 y 7: Diluciones 10⁻² y 10⁻³, respectivamente, del lisado completo. Calle 3: Fracción soluble (sobrenadante) después de la lisis de las células por sonicación. Calle 4: Fracción insoluble (precipitado) luego de la ruptura celular. Calles 5 y 6: Diluciones 10⁻¹ y 10⁻², respectivamente, de las muestras de 3 y 4.

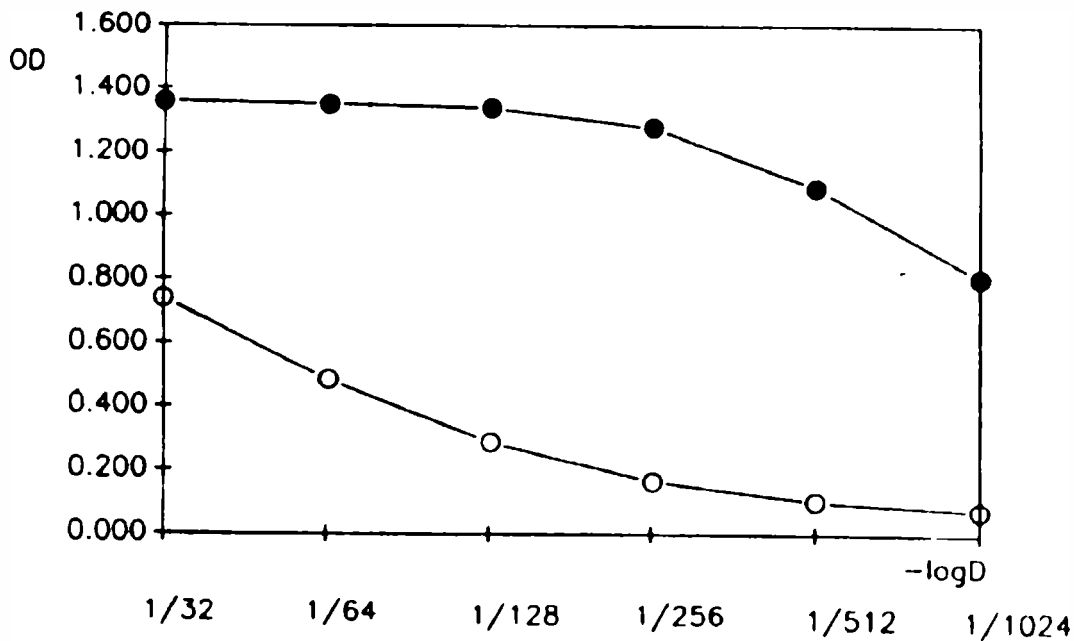


Figura 8.5. Ensayo de ELISA utilizando el antígeno N expresado en células de insecto infectadas con el recombinante Ac-N811. Las células infectadas fueron lisadas por sonicación. La fracción insoluble fue resuspendida en 50 mM Tris-HCl pH 7,4 + 0,9 % NaCl y utilizada para el ensayo a una concentración de 0,8 µg/ml. Luego de la incubación con la muestra de suero, los anticuerpos N específicos fueron detectados utilizando anti-IgG marcada con peroxidasa y O-fenilendiamina + H₂O₂ como sustratos.

● Suero de cobayo infectado con virus Junín. ○ Suero de cobayo no infectado.

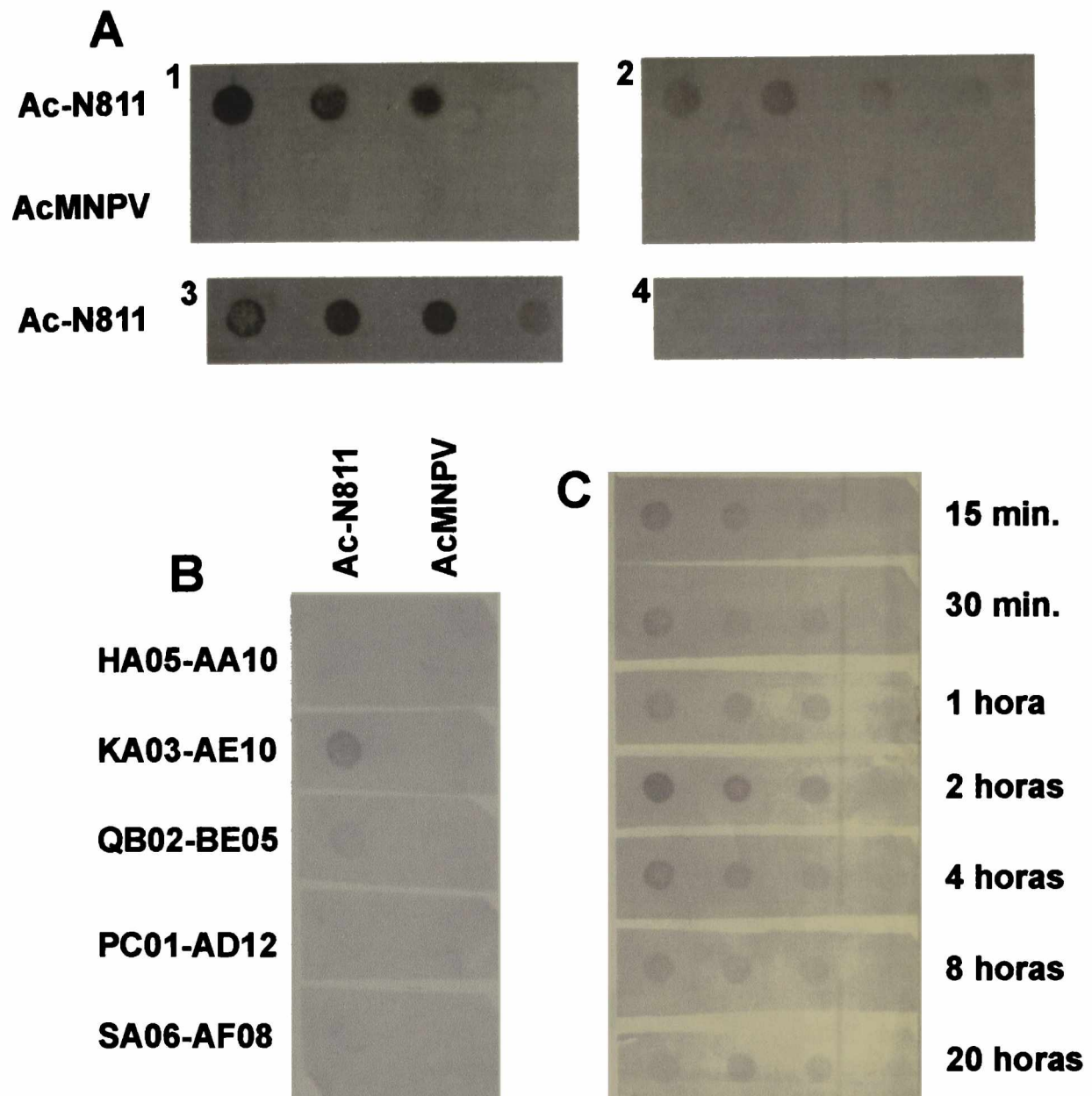


Figura 8.6. Recuperación de inmunoreactividad de la proteína N recombinante mediante un tratamiento de replegamiento *in vitro*. En todos los casos, luego de la incubación con el anticuerpo monoclonal específico, el revelado se hizo utilizando anti-IgG marcada con fosfatasa alcalina y NBT+BCIP como sustratos. **A.** Reactividad de la proteína N recombinante con anticuerpos que reconocen epitopes secuenciales (*Western +*) y epitopes conformacionales (*Western -*). *Panel 1.* Diluciones seriadas de la fracción insoluble de un lisado de *Sf 21* infectadas con Ac-N811 y diluciones seriadas de la fracción insoluble de un lisado de *Sf 21* infectadas con AcMNPV. La membrana fue sometida al proceso de replegamiento y la detección se realizó utilizando el anticuerpo monoclonal NA05-AG12 (*Western +*). *Panel 2.* Diluciones seriadas de la fracción insoluble de un lisado de *Sf 21* infectadas con Ac-N811 y diluciones seriadas de la fracción insoluble de un lisado de *Sf 21* infectadas con AcMNPV. La membrana fue sometida al proceso de replegamiento y la detección se realizó utilizando el anticuerpo monoclonal KA03-AE10 (*Western -*). *Panel 3.* Diluciones seriadas de la fracción insoluble de un lisado de *Sf 21* infectadas con Ac-N811. La membrana no fue sometida al proceso de replegamiento y la detección se realizó utilizando el anticuerpo monoclonal NA05-AG12 (*Western +*). *Panel 4.* Diluciones seriadas de la fracción insoluble de un lisado de *Sf 21* infectadas con Ac-N811. La membrana no fue sometida al proceso de replegamiento y la detección se realizó utilizando el anticuerpo monoclonal KA03-AE10 (*Western -*). **B.** Reactividad de distintos anticuerpos monoclonales anti-N (*Western -*) con la fracción insoluble de lisados de células infectadas. Las membranas fueron sometidas al proceso de replegamiento. **C.** Cinética de replegamiento de la proteína N recombinante utilizando el anticuerpo monoclonal KA03-AE10 (*Western -*). Las membranas fueron sometidas al proceso de replegamiento e incubadas a 4°C por los periodos indicados.

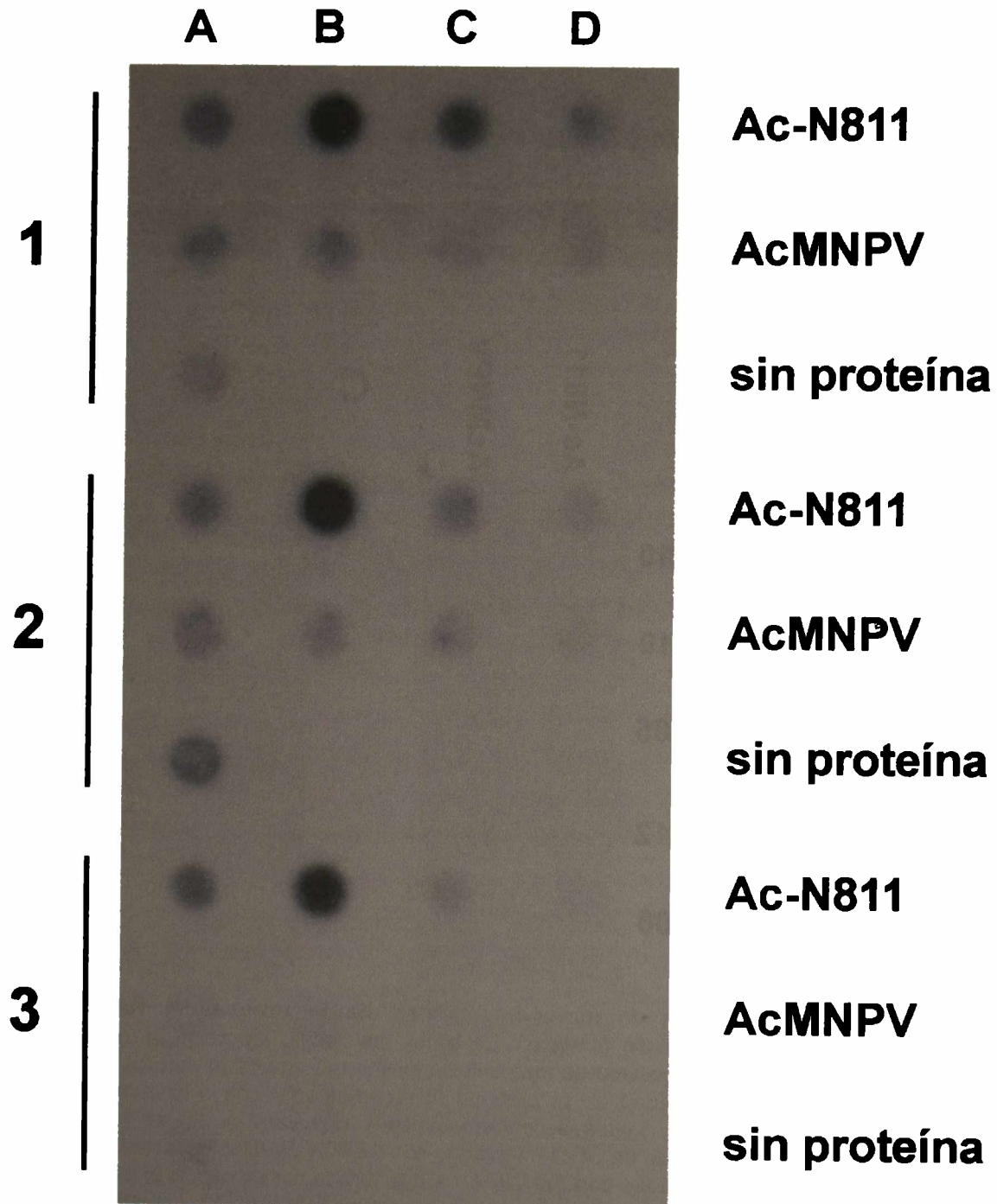


Figura 8.7. Ensayos de unión en dot de RNA a la proteína N recombinante. Las columnas A, B, C y D corresponden a 16 μg , 3,2 μg , 0,64 μg y 0,128 μg , respectivamente, de la fracción insoluble de lisados proteicos de células Sf 21 infectadas con el recombinante Ac-N811 o con AcMNPV. La membrana fue sometida al proceso de replegamiento descrito en **Métodos (Capítulo 16)**. **Panel 1:** sonda ssRNA de 110 nt de longitud, correspondientes al extremo 5' viral, marcada con $\alpha(^{32}\text{P})$ ATP. **Panel 2:** sonda dsRNA, correspondiente a la estructura tipo *panhandle* (extremo 5' + 3' viral), marcada con $\alpha(^{32}\text{P})$ ATP. **Panel 3:** sonda ssRNA de 105 nt de longitud, correspondientes al extremo 3' viral, marcada con $\alpha(^{32}\text{P})$ ATP.

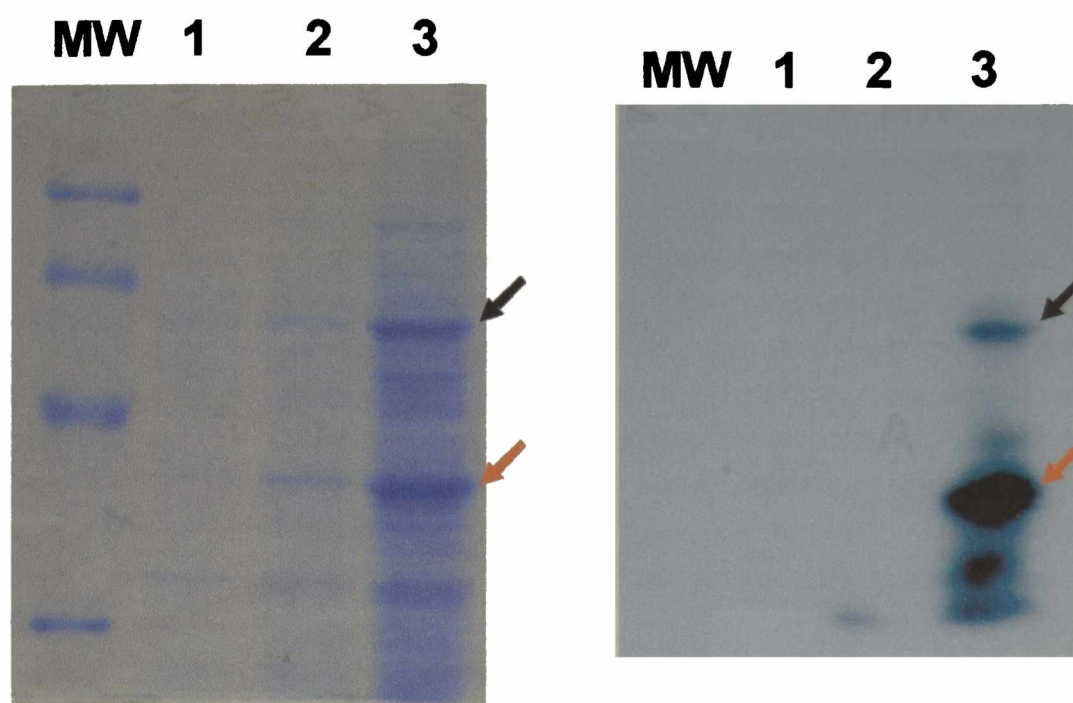


Figura 8.8. Unión de RNA a la proteína N recombinante en un ensayo de *Northwestern*. Diferentes cantidades de la fracción insoluble de un lisado proteico de células *Sf 21* infectadas con el recombinante Ac-N811 fueron corridas en un SDS-PAGE, transferidas a una membrana de nylon, sometidas al proceso de renaturalización e incubadas con una sonda de RNA viral, correspondiente a la región intergénica, marcada con $\alpha(^{32}\text{P})$ ATP. **Panel izquierdo:** SDS-PAGE. **Panel derecho:** autorradiografía del *Northwestern*. MW: Marcadores de peso molecular. Calle 1: 0,8 μg de proteínas totales. Calle 2: 1,6 μg de proteínas totales. Calle 3: 8 μg de proteínas totales. Con flechas negras se indica la proteína N del virus Junín MC2 y con flechas naranjas un producto de degradación detectado mediante *Western blot* (datos no mostrados).

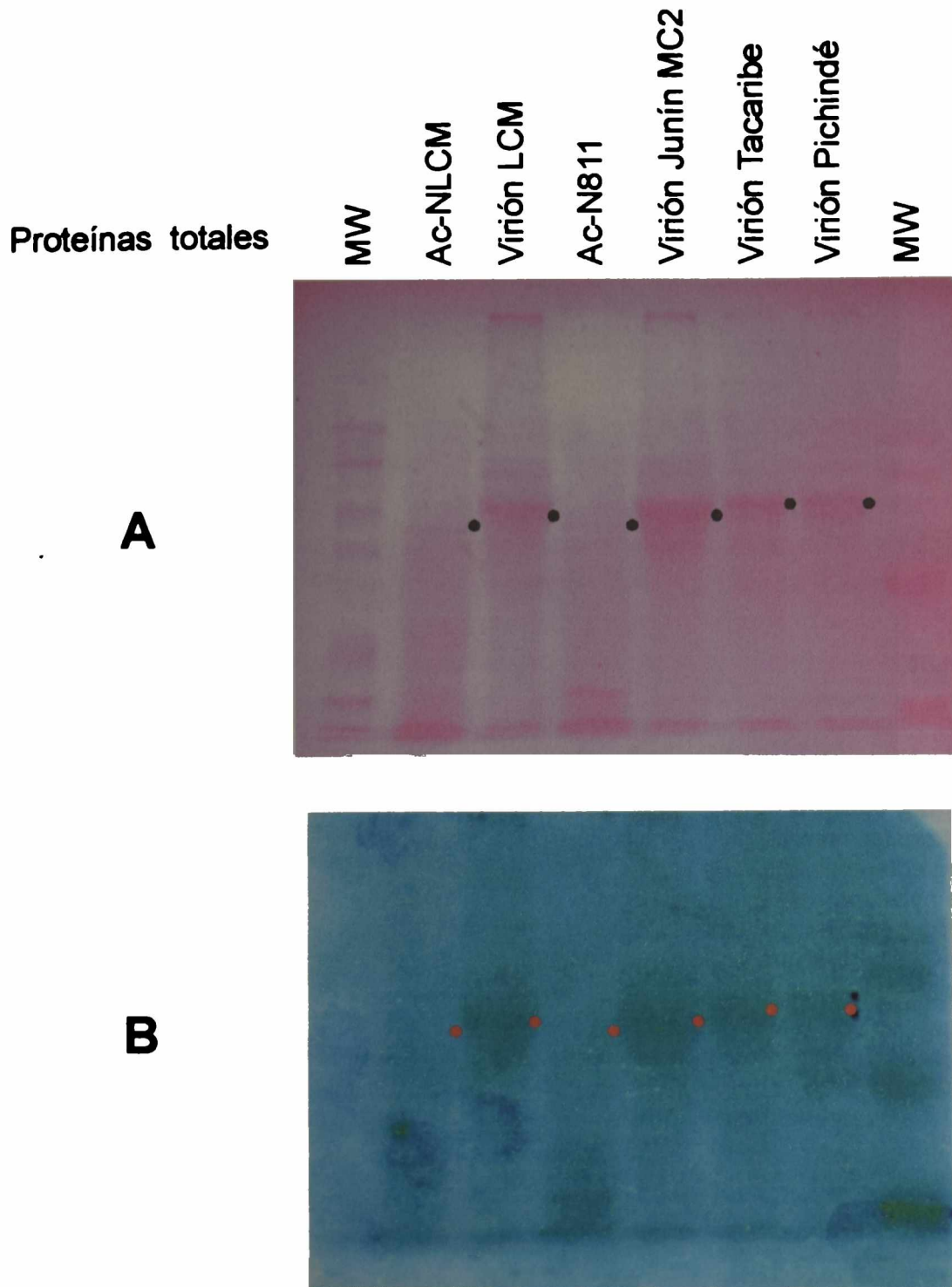


Figura 8.9. Ensayos de zinc blot con proteínas Ns *wild type* de arenavirus. Proteínas totales de lisados proteicos de células Sf 21 infectadas con el recombinante Ac-NLCM, con el recombinante Ac-N811, y de viriones purificados de LCM, Junín MC2, Tacaribe y Pichindé fueron corridas en un SDS-PAGE y transferidas a una membrana de nylon. Posteriormente, las membranas fueron sometidas al proceso de replegamiento e incubadas con $^{65}\text{ZnCl}_2$ durante 2 horas a temperatura ambiente y reveladas mediante autorradiografía. **Panel A:** Fotografía de la membrana de nylon teñida con rojo Ponceau luego de la transferencia. **Panel B:** Autorradiografía después de la incubación con $^{65}\text{ZnCl}_2$. MW: Marcadores de peso molecular. Con puntos negros o rojos se indican las proteínas N de los distintos virus.

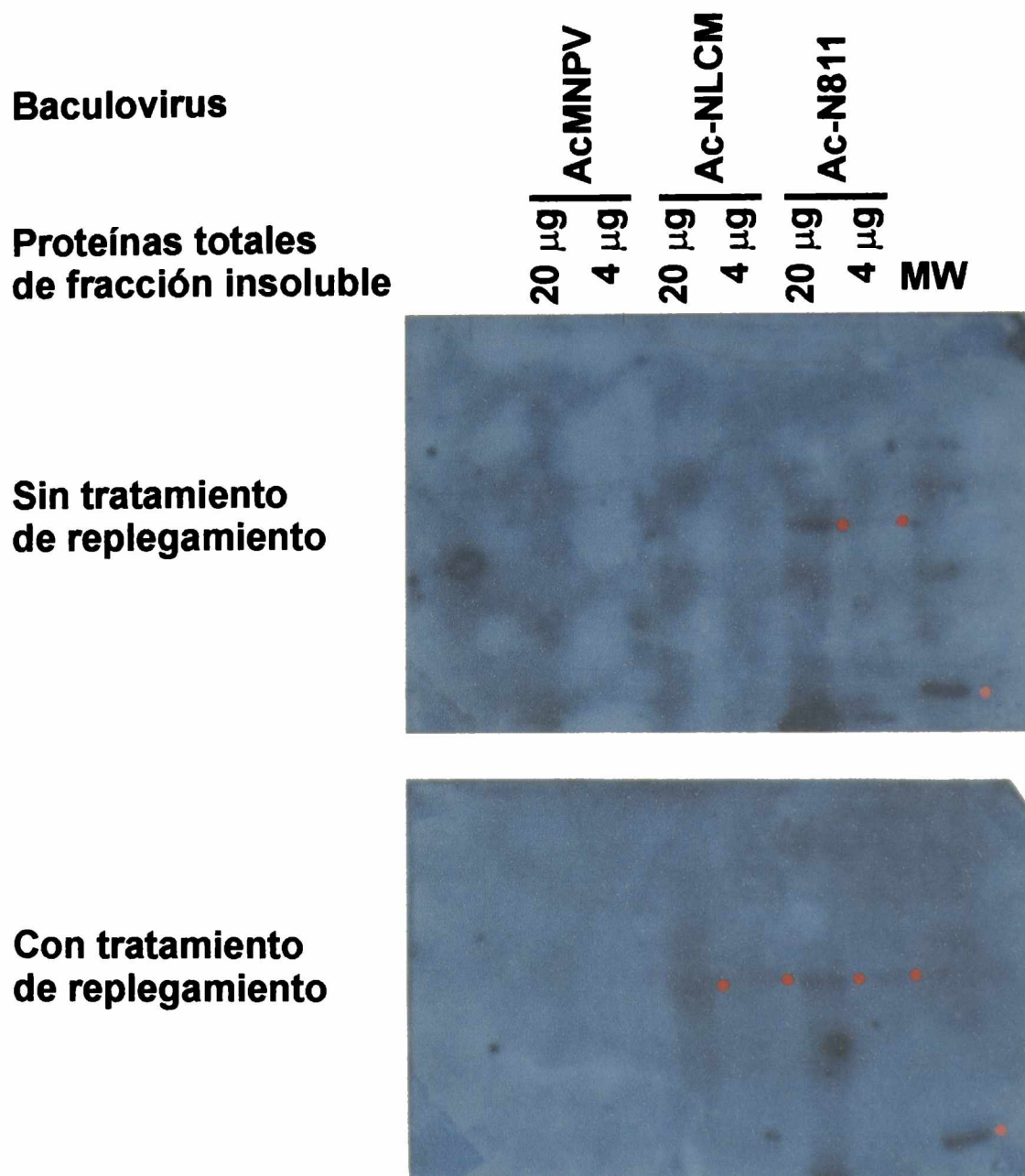


Figura 8.10. Ensayos de zinc blot con la proteína N recombinante. Diferentes cantidades de la fracción insoluble de lisados proteicos de células Sf 21 infectadas con AcMNPV, con el recombinante Ac-NLCM, o con el recombinante Ac-N811 fueron corridas en un SDS-PAGE y transferidas a una membrana de nylon. Posteriormente, las membranas fueron incubadas con $^{65}\text{ZnCl}_2$ durante 2 horas a temperatura ambiente y reveladas mediante autorradiografía. **Panel superior:** Membrana de nylon sin tratamiento de replegamiento. **Panel inferior:** Membrana de nylon con tratamiento de replegamiento. MW: Marcadores de peso molecular. Con puntos rojos se indican las proteínas N de los virus LCM y Junín MC2 y con un punto rosa se indica la anhidrasa carbónica, proteína presente en el marcador de peso molecular y con actividad de unión de zinc descripta.

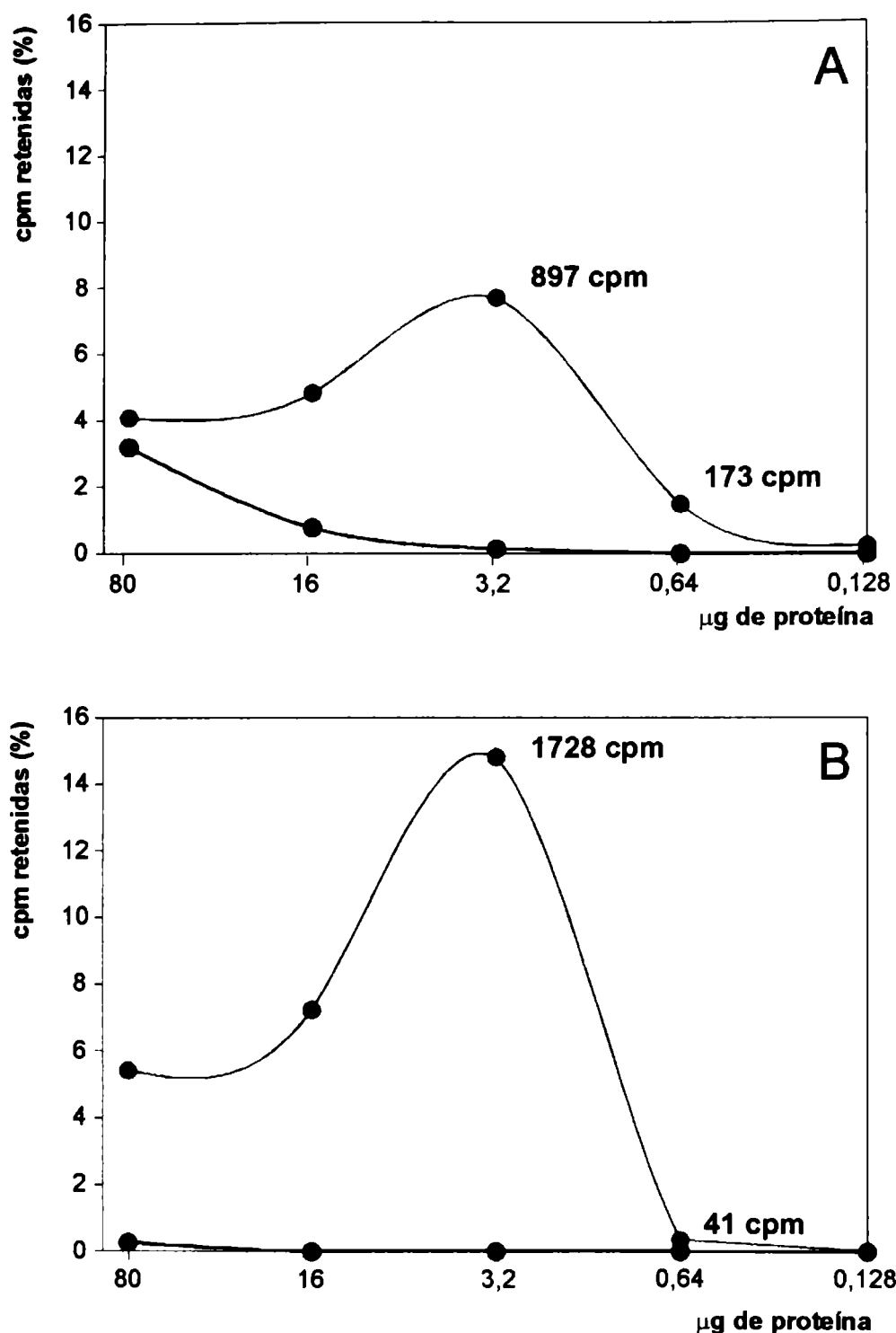


Figura 8.11. Ensayos de unión de RNA en dot en presencia o ausencia de zinc. Diluciones seriadas de proteínas totales de lisados proteicos de células Sf 21 infectadas con el recombinante Ac-N811 (rojo), con AcMNPV (negro) fueron inmovilizadas en membranas de nylon. Las membranas fueron sometidas al proceso de plegamiento en presencia o ausencia de 20 μM ZnCl₂. Posteriormente, fueron incubadas con una sonda de dsRNA, correspondiente a la estructura tipo *panhandle* (extremo 5' + 3' viral), marcada con α(³²P) ATP y reveladas mediante autorradiografía. **Panel A:** Membrana de nylon sometida al proceso de plegamiento en presencia de zinc. **Panel B:** Membrana de nylon sometida al proceso de plegamiento en ausencia de zinc.

PARTE III

RESULTADOS:

VIRUS JUNIN CEPA VACUNAL CANDID #1

Capítulo 9 Clonado y secuenciamiento del RNA S - Cepa Candid #1

CARACTERISTICAS DE LA CEPA CANDID #1

En 1958, en la ciudad de Junín, se aisló una nueva cepa de virus Junín muy virulenta, XJ (Parodi), a partir de un paciente con Fiebre Hemorrágica Argentina (Parodi *et al.*, 1958). Este aislamiento fue pasado 2 veces por cobayo y 11 veces por cerebro de ratón en el laboratorio del Dr. Parodi (Buenos Aires, BA). En 1959, la cepa XJ fue enviada al Dr. J. Casals (Rockefeller Foundation, New York, NY) y, posteriormente, fueron realizados varios pasajes por cerebro de ratón en el laboratorio Yale Arbovirus Research Unit (New Haven, CT). El pasaje 43 fue recibido en el laboratorio de USAMRIID (United States Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, MD) en 1979, y amplificado mediante un pasaje por cerebro de ratón, dando origen a la cepa XJ #44. Este homogenado de cerebro fue utilizado para infectar una línea celular diploide, derivada de pulmón de feto de mono Rhesus (FRhL-2, fetal rhesus lung diploid cells), realizando en ellas varios pasajes (Figura 9.1).

Como resultado de esta colaboración entre instituciones argentinas y estadounidenses, se obtuvo una cepa de virus Junín, llamada Candid #1, con alto grado de atenuación (Barrera Oro and Eddy, 1982; Peters *et al.*, 1987).

Luego de rigurosos controles biológicos en monos Rhesus (Mc Kee *et al.*, 1984, 1985), la cepa Candid #1 fue utilizada en voluntarios humanos y, a continuación, se realizó un ensayo clínico extensivo en el área endémica de FHA (Maiztegui *et al.*, 1987, 1998).

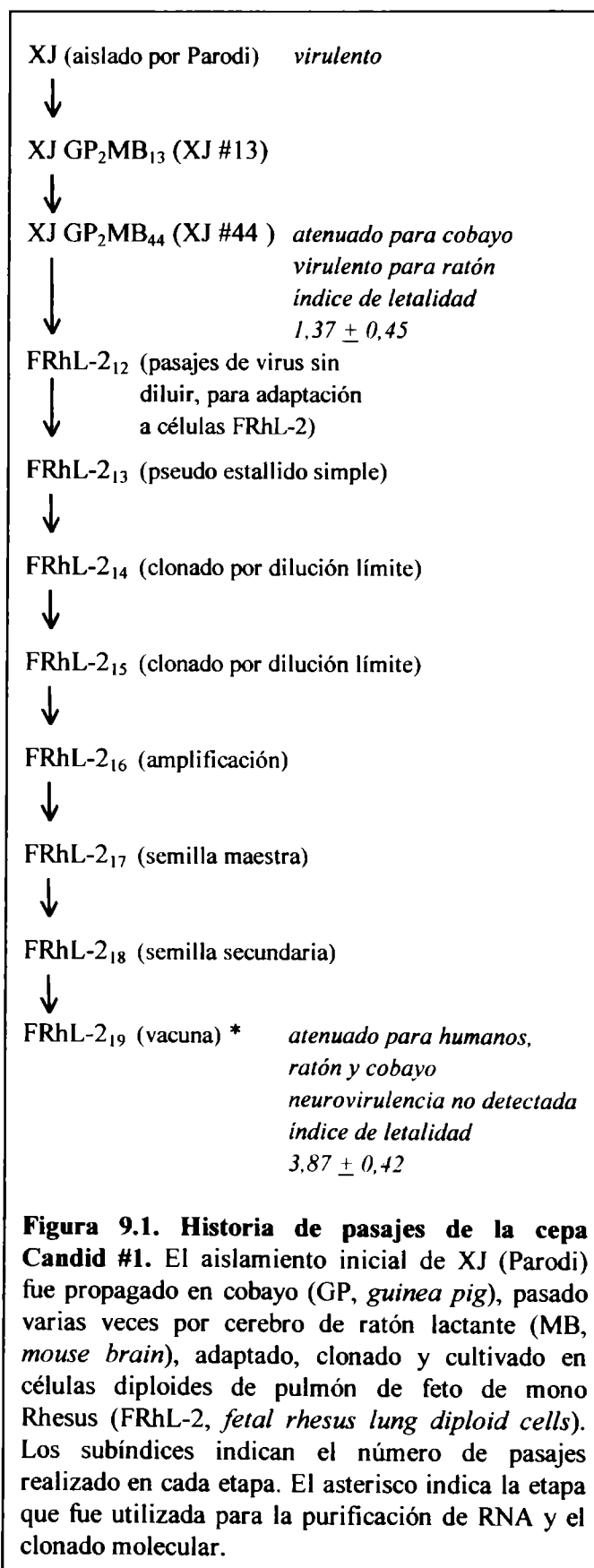


Figura 9.1. Historia de pasajes de la cepa Candid #1. El aislamiento inicial de XJ (Parodi) fue propagado en cobayo (GP, guinea pig), pasado varias veces por cerebro de ratón lactante (MB, mouse brain), adaptado, clonado y cultivado en células diploides de pulmón de feto de mono Rhesus (FRhL-2, fetal rhesus lung diploid cells). Los subíndices indican el número de pasajes realizado en cada etapa. El asterisco indica la etapa que fue utilizada para la purificación de RNA y el clonado molecular.

CRECIMIENTO EN CELULAS FRhL-2

Preparación de un stock de trabajo

La cepa vacunal Candid #1 fue gentilmente provista por los Dres. J.I. Maiztegui y A.M. Ambrosio, del INEVH (Pergamino), y propagada en células FRhL-2 certificadas.

Para mantener la similitud de pasajes entre los viriones de las vacunas y los utilizados en los estudios bioquímicos, se preparó un stock de trabajo mediante infección de monocapas de células FRhL-2 con la semilla maestra (Barrera Oro ARG 8/009, 14.8.81), con una multiplicidad de infección de 0,01 ufp. Al cuarto día post-infección, se colectó el sobrenadante y se lo dividió en alícuotas, que fueron almacenadas a -70°C .

Preparación de viriones para estudios bioquímicos

Con el fin de obtener viriones para purificar el RNA, se infectaron monocapas de células FRhL-2 (50% de confluencia) con las alícuotas obtenidas a partir de la semilla maestra. El inóculo, diluido en medio de cultivo, se usó para infectar con una m.d.i. de 0,1 ufp/cél.

Al cuarto día post-infección se colectaron los sobrenadantes para la posterior purificación de viriones (Grau *et al.*, 1981, Rosas *et al.*, 1988).

PURIFICACION DE RNA VIRAL

Los viriones presentes en los sobrenadantes de cultivos infectados, fueron concentrados por ultracentrifugación o precipitados con PEG 8000. El RNA viral fue purificado utilizando tiocianato de guanidinio (Chomczynski and Sacchi, 1987) y los precipitados finales de RNA fueron resuspendidos en agua y almacenados en alícuotas a -70°C .

El RNA obtenido fue analizado por electroforesis en gel de agarosa en presencia de 10 mM hidróxido de metilmercurio(II) como

agente desnaturalizante (Bayley and Davidson, 1976).

SINTESIS DE cDNA, CLONADO Y SECUENCIAMIENTO

Las síntesis de cDNA se realizaron utilizando la técnica descrita por Gubler & Hoffman (1983; Gubler, 1988), que emplea una transcriptasa reversa, RNasa H y DNA polimerasa I.

En la síntesis inicial de cDNA se utilizó una mezcla de tres oligonucleótidos sintéticos fosforilados, complementarios a los nucleótidos 1-19, 847-865 y 2255-2271 (numeración viral complementaria) de la cepa MC2 del virus Junín (Ghiringhelli *et al.*, 1989, 1991) (**Figura 9.2**).

El cDNA sintetizado fue reparado con la DNA polimerasa del bacteriofago T4, fraccionado por tamaño mediante electroforesis en gel de agarosa y recuperado mediante elución con polvo de vidrio o electroelución.

El cDNA eluido en el rango de 1,0 - 4,0 kpb fue ligado en el sitio SmaI del plásmido pBS SK⁺ (Bluescript SK⁺, Stratagene, La Jolla, CA), clonado en DH5 α F' y seleccionado inicialmente en placas de agar-LB con ampicilina, IPTG y X-gal.

Una metodología alternativa utilizada fue ligar adaptadores BstXI al cDNA, previo al fraccionamiento por tamaño, y el cDNA purificado fue ligado al sitio BstXI del plásmido pcDNA II (Invitrogen, San Diego, CA).

La detección de clones recombinantes fue realizada mediante hibridación de colonias sobre papel de filtro Whatman 541 (adaptado de Grunstein and Hogness, 1975), utilizando sondas radioactivas derivadas de clones de la cepa MC2.

Aunque se utilizaron simultáneamente los tres oligonucleótidos (*primers* A, B y C, **Figura 9.2**), los clones positivos detectados correspondieron a moléculas iniciadas en los *primers* B o C.

De esta manera, en hibridaciones sucesivas, se obtuvieron varios clones

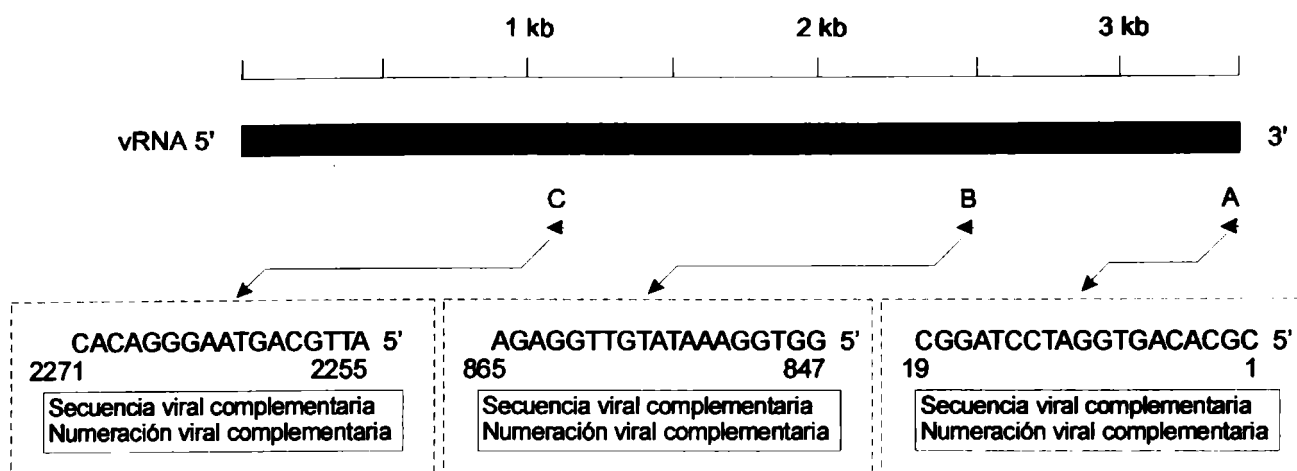


Figura 9.2. Síntesis de cDNA correspondiente al RNA S de la cepa Candid #1 del virus Junín. La secuencia del RNA S está representada en el sentido viral. Los *primers* utilizados para iniciar la síntesis de cDNA se indican como pequeñas flechas negras ubicadas en las posiciones relativas complementarias al RNA S. En los recuadros de líneas quebradas se muestra el detalle de las secuencias de los oligonucleótidos utilizados y la numeración viral complementaria derivada de la cepa MC2 (Ghiringhelli *et al.*, 1989, 1991).

independientes, de diferentes longitudes. Iniciados con el *primer* B, se obtuvieron varios clones, de diferentes longitudes y parcialmente solapados. Los clones más largos que se obtuvieron fueron: el pB220 (1779 pb), el G20 (1192 pb) y el F220 (2260 pb). Mientras que iniciados con el *primer* C, se obtuvieron los clones pC6 y pC295, de igual longitud (1094 pb), pero derivados de eventos de ligación, transformación e hibridación independientes. El análisis y la confirmación de los clones positivos se efectuó mediante digestión con enzimas de restricción y *Northern blot* (Maniatis *et al.*, 1982) (datos no mostrados).

Este conjunto de clones corresponde a la región 5' no codificante, el gen del precursor de las glicoproteínas (GPC), la región intergénica y el extremo final del gen de la nucleoproteína (N).

Para completar la información correspondiente al gen de la nucleoproteína, y con la finalidad de incrementar la eficiencia del clonado de cDNA y la obtención de copias completas del gen N, la región correspondiente fue amplificada mediante RT-PCR (*reverse transcriptase-polymerase chain reaction*) (Doherty *et al.*, 1989).

El producto de la síntesis de la primer cadena del cDNA fue empleado como molde en

una reacción de PCR con la DNA polimerasa de *Thermus aquaticus* (*Taq* DNA pol). Como *primers* se utilizaron dos oligonucleótidos sintéticos: N1 y N2, que permiten obtener como producto de amplificación el gen N completo sin las regiones no codificantes previa y posterior. Además de la secuencia nucleotídica específica, en ambos *primers* fue incorporada la secuencia nucleotídica correspondiente a los sitios de reconocimiento de las endonucleasas de restricción BglII y BamHI (Figura 9.3).

Se realizó un total de 35 ciclos con el siguiente perfil: 30 seg. a 92°C, 2 min. a 54°C y 3 min. a 72°C. Para la reacción se utilizaron las condiciones generales (buffer, concentración de Mg⁺², etc.) recomendadas por los fabricantes (Promega, Madison, WI).

El producto de la amplificación fue resuelto por electroforesis en gel de agarosa 1% en buffer TAE con 0,25 µg/ml de bromuro de etidio. El fragmento de DNA correspondiente al gen N (~1,7 kb) fue recuperado utilizando polvo de vidrio y NaI (GenClean, BIO 101, La Jolla, CA). Se repararon los extremos con el fragmento de Klenow de la DNA polimerasa I de *E. coli* en presencia de los 4 dNTPs, se digirió con la endonucleasa de restricción Bam HI, se ligó al sitio correspondiente del plásmido

pUC19 (Yanish-Perron *et al.*, 1985) y se clonó en DH5 α F', realizando la selección inicial en placas de agar-LB con ampicilina, IPTG y X-gal.

Mediante hibridación de colonias (Grunstein and Hogness, 1975), utilizando sondas radioactivas derivadas de clones de la cepa MC2, se detectaron varios clones recombinantes, entre ellos pN14 y pN175 (ambos de 1723 pb de longitud). El análisis y la confirmación de los clones positivos se efectuó mediante digestión con enzimas de restricción (Sambrook *et al.*, 1989).

En base a los patrones de restricción individuales, se fragmentaron los insertos de los diferentes clones recombinantes y se subclonaron en M13 mp18 y mp19 (Yanish-Perron *et al.*, 1985).

Los extremos no codificantes 5' y 3' fueron generados en forma independiente mediante extensión de *primer* y amplificación por PCR, utilizando metodologías alternativas.

Dada la alta conservación de los 19 nucleótidos del extremo 3' del RNA S (está presente en todos los arenavirus), el cDNA correspondiente a esta región se generó mediante RT-PCR (Figura 9.3).

El producto de la síntesis de la primer cadena del cDNA fue empleado como molde en una reacción de PCR con la DNA polimerasa de *Thermus aquaticus*. Como *primers* se utilizaron dos oligonucleótidos sintéticos: A (*primer* universal de los arenavirus) y JS186, que permiten obtener un producto de amplificación de 186 pb.

Se realizó un total de 35 ciclos con el siguiente perfil: 30 seg. a 92°C, 40 seg. a 54°C y 1 min. a 72°C. Para la reacción se utilizaron las condiciones generales (buffer, concentración de Mg⁺², etc.) recomendadas por los fabricantes (Promega, Madison, WI).

El producto de la amplificación fue resuelto por electroforesis en gel de agarosa

1,5 % en buffer TAE con 0,25 μ g/ml de bromuro de etidio. El fragmento de DNA correspondiente fue recuperado utilizando polvo de vidrio y NaI (GenClean, BIO 101, La Jolla, CA). Se repararon los extremos con la DNA polimerasa del bacteriofago T4 en presencia de los 4 dNTPs, se ligó al sitio Hinc II del plásmido pUC19 (Yanish-Perron *et al.*, 1985) y se clonó en DH5 α F', realizando la selección inicial en placas de agar-LB con ampicilina, IPTG y X-gal. Mediante digestión con endonucleasas de restricción se confirmó la presencia de insertos del tamaño correcto.

Por otra parte, como los primeros 19 nt. del extremo no codificante 5' del RNA S de la cepa MC2 del virus Junín no presentan una total complementariedad con los 19 nt. del extremo 3', se decidió utilizar un esquema diferente para determinar con precisión los primeros nucleótidos.

Se realizó una extensión de *primer* utilizando el *primer* JUNGE. Al extremo 3' del cDNA en el híbrido RNA/cDNA se le ligó un oligonucleótido sintético de secuencia definida (JunNA, ver Capítulo 17). Luego se amplificó por PCR en condiciones estándar, utilizando el juego de *primers* JUNGE/JunNB (parcialmente complementario al JunNA) (Figura 9.3).

El producto de la amplificación fue resuelto por electroforesis en gel de agarosa 1,5 % en buffer TAE con 0,25 μ g/ml de bromuro de etidio. El fragmento de DNA correspondiente fue recuperado utilizando polvo de vidrio y NaI (GenClean, BIO 101, La Jolla, CA). Se repararon los extremos con la DNA polimerasa del bacteriofago T4 en presencia de los 4 dNTPs, se ligó al sitio Hinc II del plásmido pUC19 (Yanish-Perron *et al.*, 1985) y se clonó en DH5 α F', realizando la selección inicial en placas de agar-LB con ampicilina, IPTG y X-gal. Mediante digestión con endonucleasas de restricción se confirmó la presencia de insertos del tamaño correcto.

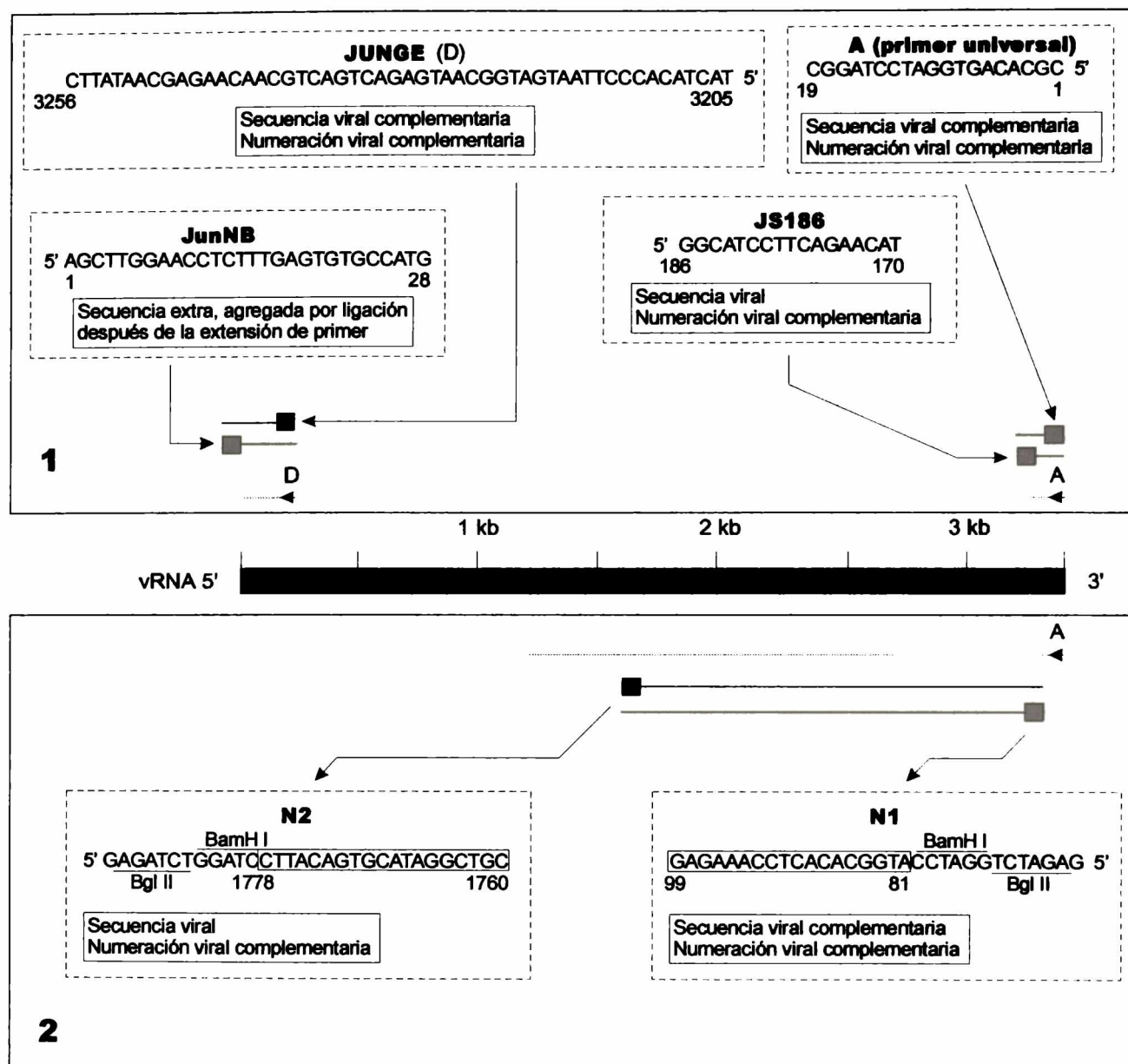


Figura 9.3. Síntesis de cDNA y amplificación por PCR de regiones del RNA S de la cepa *Candid #1* del virus Junín. La secuencia del RNA S está representada en el sentido viral entre los dos paneles. Los *primers* utilizados para iniciar las síntesis de cDNA se indican como pequeñas flechas negras ubicadas en las posiciones relativas complementarias al RNA S y el cDNA sintetizado se muestra como una línea de puntos. En ambos paneles, en los recuadros de líneas quebradas se muestra el detalle de las secuencias de los oligonucleótidos utilizados y la numeración viral complementaria derivada de la cepa MC2 (Ghiringhelli *et al.*, 1989, 1991). **Panel 1: Extensiones de primer y PCRs para definir los extremos no codificantes 5'y 3'.** Para el extremo 3', se realizó la síntesis de cDNA utilizando el *primer* A (*primer* universal de los arenavirus). Luego se amplificó mediante PCR en condiciones estándar, utilizando el juego de *primers* A/JS186. Para el extremo 5', se realizó la síntesis de cDNA utilizando el *primer* JUNGE. Al extremo 3' del cDNA en el híbrido RNA/cDNA se le ligó un oligonucleótido sintético de secuencia definida (JunNA, ver Capítulo 17). Luego se amplificó por PCR en condiciones estándar, utilizando el juego de *primers* JUNGE/JunNB (parcialmente complementario al JunNA). **Panel 2: Síntesis de cDNA correspondiente a la mitad 3' del RNA S y amplificación por PCR del gen N.** La síntesis de cDNA se realizó utilizando el *primer* A. Luego se amplificó mediante PCR en condiciones estándar, utilizando el juego de *primers* N1/N2. Las secuencias nucleotídicas delimitadas con recuadros de líneas delgadas corresponden a regiones homólogas en la cepa MC2, en secuencia viral complementaria (N1, extremo 5' del gen N) o en secuencia viral (N2, extremo 3' del gen N). Con líneas delgadas y el nombre correspondiente se indican sitios de restricción agregados para facilitar el clonado.

Todas las secuencias nucleotídicas fueron determinadas por el método de Sanger *et al.* (1977) utilizando una DNA polimerasa de T7 modificada (Sequenase™, USB, Cleveland, Ohio, USA). Las zonas ricas en pares GC se secuenciaron sustituyendo el dGTP por dITP obteniéndose lecturas claras y precisas. La información obtenida fue procesada en una computadora PS2-50Z, utilizando el programa DNASIS (Hitachi Software Engineering Co.), y en una MicroVAX 3100, utilizando el conjunto

de programas del *Genetics Computer Group* (GCG, Madison, WI, USA).

En resumen, a partir de cuatro clones parciales, con diferente grado de superposición (pC6, pF220, pN175 y p3-Ext-Cd1), se determinó la secuencia completa del RNA S de la cepa Candid #1 del virus Junín (3413 nt). La mayor parte de la secuencia nucleotídica fue confirmada mediante el secuenciamiento de varios clones independientes: pC295, pB220, pG220, pN14 y p5-Ext-Cd1 (Figura 9.4).

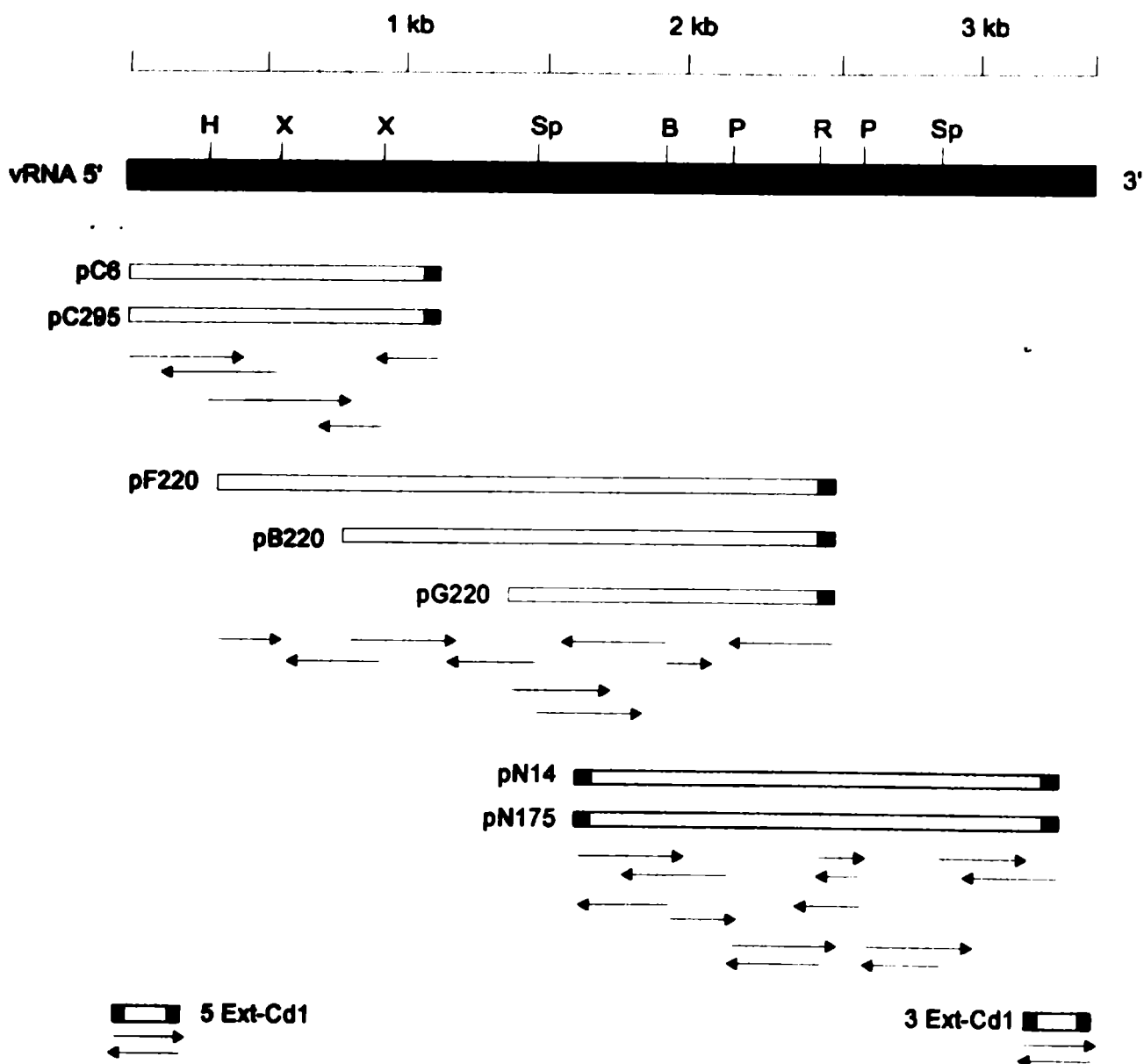


Figura 9.4. Clonado y estrategia de secuenciamiento del RNA S de la cepa Candid #1 del virus Junín. La secuencia del RNA S está representada en el sentido viral. Algunos de los clones parcialmente superpuestos se muestran como rectángulos vacíos. Los primers utilizados para iniciar la síntesis de cDNA se indican como pequeños rectángulos negros al comienzo de cada clon. Los primers utilizados para realizar las amplificaciones mediante PCR se indican como pequeños rectángulos grisados en los extremos de cada clon. Los sitios de restricción utilizados para subclonar en M13 mp18 y mp19 se muestran con abreviaturas (P: PstI; X: XbaI; H: HindIII; R: Rsa I; Sp: Sph I; B: BamHI). Las secuencias obtenidas (Sanger *et al.*, 1977) y las hebras de cDNA secuenciadas se indican con flechas de línea llena.

Capítulo 10

Información molecular - Ácidos nucleicos - Cepa Candid #1

SECUENCIA NUCLEOTIDICA DEL RNA S DE LA CEPA CANDID #1 (CD1)

En base a los datos obtenidos en la etapa de clonado y secuenciamiento y aprovechando la disponibilidad de las secuencias de los RNAs S de otros arenavirus, mediante el análisis comparativo se reconstruyó la secuencia completa del RNA S de la cepa CD1 del virus

Junín (Figura 10.1). Dicha secuencia es de 3413 nucleótidos de longitud y está formada por 905 A, 805 C, 717 G y 986 U (26,5%, 23,6%, 21,0% y 28,9% respectivamente).

La longitud de la secuencia es concordante con la estimada en base a la migración electroforética (Añón *et al.*, 1976) y, además, se corresponde con el rango de tamaños de los RNAs S de otros arenavirus (Tabla 10.a).

CUGCAGUAAG	GGGAUCCUAG	GCGAUUUUGG	UAACGCUGUA	AGUUGUUACU	GCUUUCUAUU	UGGACAACAU	CAAACCAUCU	80
AUUGUACA	GGGGCAAUUC	AUCAGCUUCA	UGCAAGAAAU	ACCUACCUUU	UUGCAGGAAG	CUCUGAAUUU	UGCUCUUGUU	160
GCAGUCAGUC	UCAUUGCCAU	CAUUAAGGGU	GUAGUAAACC	UGUACAAAAG	UGGUUUGUUC	CAAUUCUUUG	UAUUCUAGC	240
ACUCGCAGGA	AGAUCUCGCA	CAGAAGAAGC	UUUUAAAAUC	GGACUCGACA	CAGAGUCCA	GACUGUGUCC	UUCUCAUUGG	320
UGGGUCUCUU	UCCAACAAU	CCACAUGACC	UGCCUCUGUU	GUGUACCUUA	AACAAGAGCC	AUCUUUACAU	UAAGGGGGGC	400
AAUGCUUCAU	UCAAGAUCAG	CUUUGAUGAC	AUCGCAGUGU	UGUUACCAGA	AUAUGACGUU	AUAAUUCAGC	AUCCGGCAGA	480
UAUGAGCUGG	UGUUCUAAAA	GUGAUGAUCA	AAUUUGGCUG	UCUCAGUGGU	UCAUGAAUGC	UGUGGGGCAU	GAUUGGUUUC	560
UAGACCCACC	AUUUCUGUGU	AGGAACCGUA	CAAAGACAGA	AGGCUUCAUC	UUUCAAGUCA	AUACCUCCAA	GACUGGUUUC	640
AAUGAAAACU	AUGCCAAGAA	GUUUAAGACU	GGUAUGCACC	AUUUAUAUAG	AGAAUACCCC	GACUCUUGCU	UGGAUGGCAA	720
ACUGUGUUUG	AUGAAGGCAC	AACCCACCAG	UUGGCCUCUC	CAAUGUCCAC	UUGACCAUGU	CAACACAUUA	CAUUUCCUCA	800
CAAGAGGCAA	GAACAUUCAG	CUUCCAAGGA	GGUCUUUAAA	AGCAUUCUUU	UCCUGGUCUC	UGACAGACUC	AUCCGGCAAAG	880
GACACCCUCG	GAGGCUAUUG	UCUAGAAGAG	UGGAUGCUCG	UUGCAGCCAA	AAUGAAGUGU	UUUGGCAAUA	CUGCUGUAGC	960
AAAAUGCAAU	CUGAAUCAUG	ACUCUGAAUU	CUGUGACAUG	CUGAGGCUUU	UUGAUUACAA	CAAAAUGCU	AUCAAACCU	1040
UAAAUGAUGA	AACUAAGAAA	CAAGUAAAUC	UGAUGGGACA	GACAAUCAAU	GCGCUGAUUU	CUGACAAUUU	AUUGAUGAAA	1120
AACAAAUAUA	GGGAACUGAU	GAGUGUCCCU	UACUGCAAUU	ACACAAAUAU	UUGGUUUGUC	AACCACACAC	UUUCAGGACA	1200
ACACUCAUUA	CCAAGGUGCU	GGUUAUAAA	AAACAACAGC	UAUUUGAACA	UCUCUGACUU	CCGUAAUGAC	UGGAUUAUUG	1280
AAAGUGACUU	CUUAAUUUCU	GAAAUGCUAA	GCAAAGAGUA	UUCGGACAGG	CAGGGUAAA	CUCUUUGAC	UUUAGUUGAC	1360
AUCUGUAUUU	GGAGCACAGU	AUUCUUCACA	GCGUCACUCU	UCCUUCACUU	GGUGGGUUAU	CCUCCCCACA	GACACAUCAG	1440
GGGCGAAGCA	UGCCCUUUGC	CACACAGGUU	GAACAGCUUG	GGUGGUUGCA	GAUGUGGUAA	GUACCCCAAU	CUAAAGAAAC	1520
CAACAGUUUG	GCGUAGAGGA	CAC	JAA	ACC	UCCUGAGGGU	CCCCACCAGC	CCGGGCACUG	1600
CAGUCCGCGG	CCUGGCGCGG	GACUGGGGAG	GCACUGC	JUA	CAGUGCAUAG	GCUGCCUUCG	GGAGGAACAG	1680
GGUAAUAGAG	GUGUAAGUUC	UUUCUCAUAG	AGCUUCCCAU	CCAACACUGA	CUGAAACAUU	AUGCAGUCUA	GCAGAGCACA	1760
GUGUGGCUCA	CUGGAGGCCA	ACUUAAAGGG	AGCAUCUUA	UCUCUCUUU	UCUUGCUGAC	AACCACUCCA	UUGAUGAUUU	1840
UGCAUAGGUG	GCCAAAUUUC	UCCAGACCU	GUUGGUCAA	UCGCCUGGCU	UGUUCUGAUG	UAAGCCUAAC	AUCAACCAGC	1920
UUAAGAUCUC	UUUCUCCAUG	GAGGUCAAAC	AACUUCUGA	UGUCAUCGGA	CCUUGAGUG	GUCACAACCA	UGUCCGGAGG	2000
CAGCAAGCCA	AUCACGUAAAC	UAAGAACUCC	UGGCAUUGCA	UCUUCUAUGU	CUUUCAUUAA	GAUGCCGUGA	GAGUGUCUGC	2080
UACCAUUUUU	AAACCCUUUC	UCAUCAUGUG	GUUUUCUGAA	GCAGUGAAUA	UACUGCUUAC	CUGCAGGCUG	GAACAACGCC	2160
AUCUCAACAG	GGUCAGUAGC	UGGUCCUUCU	AUGUCGAGCC	AAAGGGUUAU	GGUGGGGUGC	AGUUUCCCCA	CUGCCUCUCU	2240
GAUGACAGCU	UCUUGUAUCU	CUGUCAAGUU	AGCCAAUCUC	AAAUUCUGAC	CGUUCUUUUC	CGGUUGUCUA	GGUCCAGCAA	2320
CUGGUUUCCU	UGUCAGAUCA	AUACUUGUGU	UGUCCCAUGA	CCUGCCUUAU	AUUUGUGAUC	UGGAACCAAU	AUAAGGCCAA	2400
CCAUCGCCAG	AAAGGCAAAG	UUUGUACAGA	AGGUUUUCAU	AAGGGUUUCU	AUUGCCUGGU	UUCUCAUCA	UAAACAUGCC	2480
UUCUCUUCGU	UUAACCUGAA	UGGUUGAUUU	UAUGAGGGAA	GAAAAGUUUU	CUGGGGUGAC	UCUGAUUGUC	UCCAACAUAU	2560
UUCACCAUC	AAGAAUGGAU	GCACCAGCCU	UUACUGCAGC	UGAAAGACUA	AAGUUGUAGC	CAGAAAUGUU	GAUGGAGCUU	2640
UCAUCCUUAU	UCACAAUCUG	GAGGCAGUCA	UGUUCUGAG	UCAUUCUGUC	AAGGUCACUC	AAGUUUGGAU	ACUUCACAGU	2720
GUAUAGAAGC	CCAAGAGAGG	UUAAAGCCUG	UAUGACACUG	UUCAUUGUCU	CACCUCCUUG	AACAGUCAUG	CAUGCAAUUG	2800
UCAAUGCAGG	AACAGAACCA	AACUGAUUGU	UAAGUUUUGA	AGGAUCUUUA	ACAUCCCAUA	CCUCACCAC	ACCAUUUCCC	2880
CCAGUUCCUU	GCUGUUGAAA	UCCAGUGUU	UCUAAUUCU	CUGAUUCUUU	GGCCAGUUGU	GAUUGAGACA	AGUUACCCAU	2960
GUAAACCCUU	UGAGAGCCUG	UCUCUGUCU	UCUAAUCUUG	GUUUUUAAUU	UCUCAAGGUC	AGACGCCAAC	UCCAUCAGCU	3040
CAUCCUCCC	CAGAUUCUCC	GCCUUGAAA	CUGUGUUUCG	UUGAACACUC	UCUAUGGACA	UGAGUCUGUC	AACCUCUUUA	3120
UUCAGGUCCC	UCAACUUUUU	GAGGUCUUUC	UCCCCCUUU	UAGUCUUUCU	GAGUGCCCGC	UGCACCUGUG	CCUCUUGGUU	3200
GAAGUCAUUG	CUGUCAGCAA	UUAGCUUGGC	AUCCUUCAGA	ACAUCGGACU	UGACAGUCUG	AGUAAAUUGA	CUCAAACCUC	3280
UCCUUAAAGGA	CUGAGUCCAU	CUAAAGCUUG	GAACCUCUUU	GGAGUGUGCC	AUCCAGAAAG	AUCUGGGUGU	UUUGAUCUGA	3360
GAAAAAAUUG	CUCAGUGAAA	GUGUAGACA	CUAUGCCUAG	GAUCCACUGU	GCG			3413

Figura 10.1. Secuencia nucleotídica del RNA S de la cepa CD1 del virus Junín. Se indican los 3413 nucleótidos del RNA S en sentido viral, numerados a la derecha. En caracteres blancos sobre fondo negro se indican los codones de iniciación y terminación de los genes GPC (polaridad positiva) y N (polaridad negativa, en este caso son los tripletes complementarios). Con flechas apuntando hacia la región del *loop* se señalan las dos estructuras intergénicas tipo *hairpin-loop*.

Tabla 10.a. Longitud de los RNAs S de arenavirus

Virus	Cepa	Longitud en nucleótidos	Referencias
Junín	MC2	3400	Ghiringhelli <i>et al.</i> , 1989 Ghiringhelli <i>et al.</i> , 1991
Junín	CD1	3413	Ghiringhelli <i>et al.</i> , 1997 Albariño <i>et al.</i> , 1997
Lassa	GA391 (Nigeria)	3417	Clegg and Oram, 1985 Clegg <i>et al.</i> , 1990
Lassa	Josiah	3402	Auperin and McCormick, 1988
LCM	Armstrong	3376	Salvato <i>et al.</i> , 1988
LCM	WE	3375	Romanowski <i>et al.</i> , 1985 Romanowski and Bishop, 1985
Mopeia	800150	3419	Wilson <i>et al.</i> , 1991
Olivero	----	3534	Bowen <i>et al.</i> , 1996
Pichindé	3739	3418	Auperin <i>et al.</i> , 1984 Bishop and Auperin, 1987
Sabiá	SPH114202	3366	Gonzalez <i>et al.</i> , 1996
Tacaribe	T.RVL.II 573	3421	Franze Fernández <i>et al.</i> , 1987

La búsqueda de marcos de lectura abiertos en las seis fases posibles (tres en sentido viral y tres en sentido viral complementario) dio como resultado varios potenciales marcos en ambas orientaciones (**Tabla 10.b**), entre los cuales fue seleccionado el más largo en cada orientación, en correlación con la información experimental disponible.

Los codones de iniciación y terminación indicados en la **Figura 10.1** corresponden a estos marcos seleccionados

REGIONES CODIFICANTES

Gen del precursor de las glicoproteínas

El primer AUG de polaridad viral, ubicado en los nucleótidos 89-91, comienza un marco de lectura abierto de 1458 nucleótidos de longitud que termina en un codón UAA ubicado en los nucleótidos 1556-1558.

Aunque más allá del AUG₈₉₋₉₁ existen otros codones AUG (**Tabla 10.b**), el primero se considera el verdadero codón de iniciación debido a la conservación de algunos de los residuos del entorno definidos por Kozak (1978, 1983, 1984), se conservan la G de +4, la A de -3 y la C de -2 (ver **Capítulo 16**).

Tabla 10.b. Marcos de lectura abiertos

n°	Ubicación (nt.)	Long. proteína (aa)
Posición 1 (polaridad viral)		
No hay marcos de lectura mayores de 200 nt.		
Posición 2 (polaridad viral)		
1	89-1546	485
2	110-1546	478
3	317-1546	409
4	482-1546	354
5	533-1546	337
6	674-1546	290
7	731-1546	271
8	914-1546	210
9	932-1546	204
10	998-1546	182
11	1073-1546	157
12	1115-1546	143
13	1139-1546	135
14	1304-1546	80
Posición 3 (polaridad viral)		
1	2106-2357	84
Posición 4 (polaridad viral complementaria) *		
1	82-1776	564
2	313-1776	487
3	319-1776	485
4	379-1776	465
5	454-1776	440
6	625-1776	383
7	649-1776	375
8	856-1776	306
9	937-1776	279
10	1252-1776	174
11	1357-1776	139
12	1378-1776	132
13	1423-1776	117
Posición 5 (polaridad viral complementaria) *		
1	1448-1663	72
2	1694-1900	69
Posición 6 (polaridad viral complementaria) *		
No hay marcos de lectura mayores de 200 nt.		

* Se indica la ubicación nucleotídica en sentido VC.

Además, el tamaño del polipéptido resultante (485 aa.) y el peso molecular estimado (55559 Da, sin considerar la glicosilación) se correlacionan con los datos obtenidos para MC2 (Rustici, 1984; De Mitri and Martínez Segovia, 1985) y otros .

El análisis del uso de codones de este marco de lectura abierto se resume en la **Tabla 10.c.** Como se observa en la misma no todos los aminoácidos presentan preferencias de uso de codón. Considerando la relación entre la

frecuencia particular de un codón y la frecuencia promedio para el mismo, podemos ver que si esta relación es 1,5 veces o mayor que la frecuencia promedio hay una preferencia neta de uno o más codones para ese aminoácido, mientras que una relación $\geq 1,2$ y $< 1,5$ muestra una preferencia moderada.

En función de lo anterior, se puede destacar el caso de la Arg, donde, de los codones iniciados con C tres no son utilizados y otro apenas cubre el 16,7 % de la frecuencia de uso.

Tabla 10.c. Frecuencia de uso de codones en el gen GPC

Aminoácido	Codón	Frecuencia	Aminoácido	Codón	Frecuencia
Ala (A)	GCT	31,8 %	Lys (K)	AAA	51,5 %
	GCC	13,6 %		AAG	48,5 %
	GCA	45,5 %	MET (M)	ATG	100,0 %
	GCG	9,1 %	Phe (F)	TTT	40,0 %
Arg (R)	CGT	16,7 %		TTC	60,0 %
	CGC	0,0 %	Pro (P)	CCT	33,3 %
	CGA	0,0 %		CCC	19,0 %
	CGG	0,0 %		CCA	42,9 %
	AGA	33,3 %		CCG	4,8 %
	AGG	50,0 %	Ser (S)	TCT	24,3 %
Asn (N)	AAT	54,8 %		TCC	18,9 %
	AAC	45,2 %		TCA	16,2 %
Asp (D)	GAT	33,3 %		TCG	2,7 %
	GAC	66,7 %		AGT	16,2 %
Cys (C)	TGT	58,8 %		AGC	21,6 %
	TGC	41,2 %	ACT	26,9 %	
Gln (Q)	CAA	56,3 %	Thr (T)	ACC	23,1 %
	CAG	43,7 %		ACA	50,0 %
Glu (E)	GAA	83,3 %		ACG	0,0 %
	GAG	16,7 %	Trp (W)	TGG	100,0 %
Gly (G)	GGT	37,1 %	Tyr (Y)	TAT	53,3 %
	GGC	29,6 %		TAC	46,7 %
	GGA	22,2 %	Val (V)	GTT	23,8 %
	GGG	11,1 %		GTC	23,8 %
His (H)	CAT	47,0 %		GTA	28,6 %
	CAC	53,0 %	GTG	23,8 %	
Ile (I)	ATT	35,7 %	Term.	TAA	100,0 %
	ATC	42,9 %		TGA	0,0 %
	ATA	21,4 %		TAG	0,0 %
Leu (L)	TTA	20,7 %			
	TTG	22,4 %			
	CTT	12,1 %			
	CTC	12,1 %			
	CTA	8,6 %			
	CTG	24,1 %			

Tabla 10.d. Frecuencia de uso de codones en el gen N

Aminoácido	Codón	Frecuencia	Aminoácido	Codón	Frecuencia
Ala (A)	GCT	35,3 %	Leu (L)	CTC	15,1 %
	GCC	16,1 %		CTA	7,6 %
	GCA	41,9 %		CTG	19,7 %
	GCG	9,7 %	Lys (K)	AAA	48,8 %
Arg (R)	CGT	0,0 %	AAG	51,2 %	
	CGC	0,0 %	MET (M)	ATG	100,0 %
	CGA	5,7 %	Phe (F)	TTT	72,2 %
	CGG	2,9 %	TTC	27,8 %	
	AGA	60,0 %	Pro (P)	CCT	36,0 %
	AGG	31,4 %		CCC	8,0 %
Asn (N)	AAT	41,7 %		CCA	40,0 %
AAC	58,3 %	CCG		16,0 %	
Asp (D)	GAT	51,3 %	Ser (S)	TCT	21,4 %
	GAC	48,7 %		TCC	21,4 %
Cys (C)	TGT	14,3 %		TCA	21,4 %
	TGC	85,7 %		TCG	2,4 %
Gln (Q)	CAA	48,4 %		AGT	19,1 %
	CAG	51,6 %		AGC	14,3 %
Glu (E)	GAA	45,2 %	Thr (T)	ACT	37,9 %
	GAG	54,8 %		ACC	24,2 %
Gly (G)	GGT	29,3 %		ACA	37,9 %
	GGC	22,0 %	ACG	0,0 %	
	GGA	26,8 %	Trp (W)	TGG	100,0 %
	GGG	22,0 %	Tyr (Y)	TAT	54,5 %
His (H)	CAT	36,4 %		TAC	45,5 %
	CAC	63,6 %	GTT	48,6 %	
Ile (I)	ATT	51,9 %	Val (V)	GTC	17,1 %
	ATC	22,2 %		GTA	5,7 %
	ATA	25,9 %		GTG	28,6 %
Leu (L)	TTA	7,6 %	Term.	TAA	100,0 %
	TTG	27,3 %		TGA	0,0 %
	CTT	22,7 %		TAG	0,0 %

Gen de la proteína de la nucleocápside

En la fase 4 (Tabla 10.b), de polaridad viral complementaria, el AUG₈₂₋₈₄ comienza un marco de lectura abierto de 1695 nucleótidos de longitud que termina en el UAA₁₇₇₄₋₁₇₇₆.

Aunque más adelante, en la misma fase, existen otros AUGs (Tabla 10.b), el primero es considerado el verdadero codón de iniciación de la proteína N, basado en la conservación de algunos de los nucleótidos del entorno, se conservan la G de +4, la C de -2, la G de -3 y la C de -5 (Kozak, 1978, 1983, 1984; ver Capítulo 16).

Además, el polipéptido resultante tiene un peso molecular calculado de 63123 Da. Este resultado concuerda con el peso molecular estimado para la proteína de la nucleocápside viral de MC2 (Martínez Segovia and De Mitri,

1977; Grau *et al.*, 1981; Rustici, 1984), y los de otros arenavirus.

El análisis del uso de codones de este gen se resume en la Tabla 10.d. Teniendo en cuenta lo ya considerado para el gen GPC, no hay una preferencia generalizada de uso de codones, aunque, nuevamente se destaca la Arg, donde dos de los codones iniciados con C no son utilizados, mientras que otros dos representan apenas el 8,6 % de la frecuencia de uso. En realidad, en este caso, más que hablar de una frecuencia de uso habría que hablar de una frecuencia de desuso, ya que en 7 de 9 aminoácidos (77,8 %) cuyo código está representado por más de dos codones, existe al menos un triplete cuya frecuencia está muy por debajo de la frecuencia promedio.

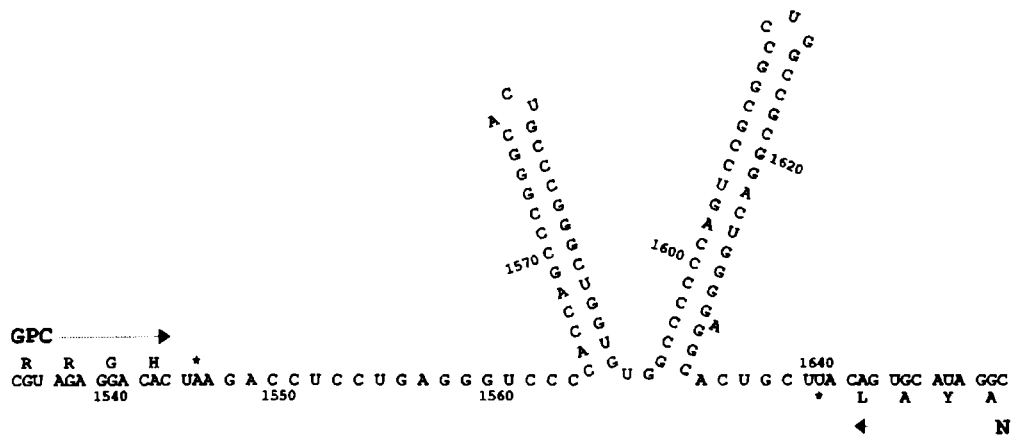


Figura 10.2. Región intergénica del RNA S de la cepa CD1 del virus Junín. Porción del RNA S (nt. 1532-1652) que comprende los extremos de los genes GPC y N (polaridad viral y viral complementaria, respectivamente) y la región intergénica no codificante. Las direcciones de traducción se indican con flechas y los codones de terminación están señalados con asteriscos. Las dos potenciales estructuras tipo *hairpin-loop* están estabilizadas por 13 y 17 pares de bases (ΔG° -39,0 kcal/mol y ΔG° -54,2 kcal/mol, respectivamente).

REGIONES NO CODIFICANTES

Región intergénica

La región intergénica ubicada entre los genes GPC y N, nucleótidos 1547 a 1637 en polaridad viral, contiene dos regiones con autocomplementariedad suficiente como para constituir dos estructuras tipo *hairpin-loop* muy estables (Figura 10.1). El *hairpin* cercano al gen GPC (nt. 1567-1592) está formado por 11 pares GC y 2 pares AU (ΔG° -39,0 kcal/mol; Tinoco *et al.*, 1973) mientras que el cercano al gen N (nt. 1595-1632) está formado por 15 pares GC y 2 pares AU (ΔG° -54,2 kcal/mol) (Figura 10.2). Estas estructuras, calculadas manualmente al comienzo, fueron confirmadas mediante análisis computacional con el algoritmo de Zuker & Stiegler (1981).

La secuencia de la región intergénica de CD1 permitió confirmar la presencia del doble *hairpin* descrito en MC2 (Ghiringhelli *et al.*, 1987, 1991). Además, previamente fue re-secuenciada la región intergénica de Tacaribe (Franze-Fernández *et al.*, comunicación personal), demostrando que este arenavirus también posee un doble *hairpin*.

Extremos 5' y 3' del genoma viral

Las regiones no traducidas 5' y 3' del RNA viral comprenden 88 y 81 nucleótidos respectivamente. Estas dos regiones presentan una cierta complementariedad y, como ya se mencionó en el capítulo 3, pueden ser alineadas para formar una estructura tipo *panhandle* (ΔG° -58,0 kcal/mol) (Figura 10.3).

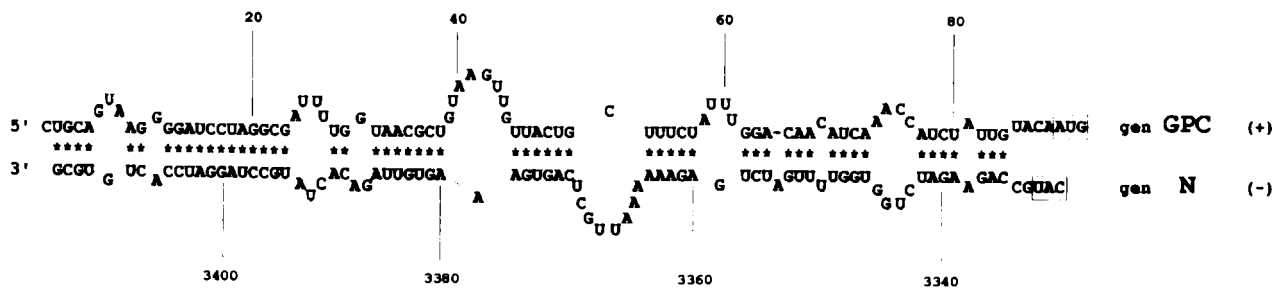


Figura 10.3 Extremos 5' y 3' del RNA S de la cepa CD1 del virus Junín. Estructura *panhandle* predicha para los extremos del RNA S. La complementariedad de nucleótidos entre los extremos no codificantes se indica con asteriscos. Se han incorporado también los pares no convencionales G-U. La energía de estabilización de esta estructura secundaria es de -58,0 kcal/mol (Tinoco *et al.*, 1973). Con recuadros se señalan los codones de iniciación de los genes GPC y N.

Capítulo 11 Información molecular - Proteínas - Cepa Candid #1

GLICOPROTEINAS VIRALES

Estructura primaria del precursor de las glicoproteínas (GPC)

El marco de lectura abierto en la mitad 5' del RNA S (polaridad viral), delimitado por los codones AUG₈₉₋₉₁-UAA₁₅₄₆₋₁₅₄₈, posibilita la traducción de un polipéptido de 485 aminoácidos (Figura 11.1), que es glicosilado y procesado postraduccionalmente.

En la proteína GPC pueden ser definidos

tres dominios: el péptido señal, la glicoproteína G1 y la glicoproteína G2.

Estructura del péptido señal

El péptido señal de CDI tiene una longitud de 58 aminoácidos, concordando con los datos de Buchmeier *et al.* (1987) y Burns & Buchmeier (1993).

El mismo está constituido por 48,2 % de aminoácidos hidrofóbicos, 18,9 % de aminoácidos hidrofílicos y 32,9 % de aminoácidos ambivalentes (Tabla 11.a).

MGQFISFMQE	IPTFLOEALN	<u>I</u> ALVAVSLIA	<u>I</u> IKGVVNLK	SGLFOFFVEL	50
ALAGRSCTEE	AFKIGLHTEF	QTVSFSMVGL	FSNNPHDLPL	LCTL██████HLY	100
IKGGNASFKI	SFDDIAVLLP	EYDVIIQHPA	DMSWCSKSD	QIWLSQWFMN	150
AVGHDWYLDP	PFLCRNRTKT	EGFIFQVNTS	KTGINENYAK	KFKTGMHHL	200
REYPDSCLDG	KLCLMKAQPT	SWPLQCPLDH	VNTLHFLTRG	KNIQLPRRSL	250
KAFFSWSLTD	SSGKDTPGGY	CLEEWMLVAA	KMKCFGNTAV	AKCNLNHDSE	300
FCDMLRLFYD	NKNAIKTLND	ETKKQVNLMG	QTINALISDN	LLMKNKIREL	350
MSVPYC██████K	FWYVNHNTLSG	QHSLPRCWLI	K██████YL██████D	FRNDWILESD	400
FLISEMLSKE	YSDRQGKTPL	TLVDICIWST	<u>V</u> FEETASLEFLH	<u>L</u> VGIPSHRHI	450
RGEACPLPHR	LNSLGGCRCG	KYPNLKKPTV	WRRGH		485

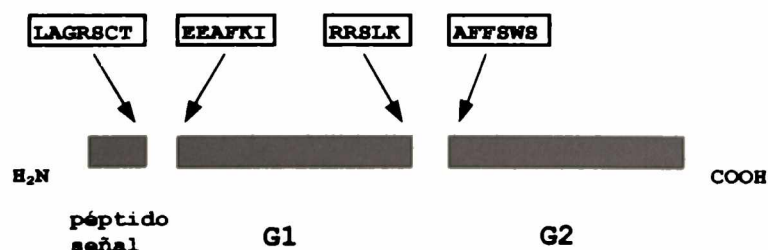


Figura 11.1. Precursor de las glicoproteínas (GPC) de la cepa Candid #1 del virus Junín. La secuencia aminoacídica fue deducida a partir de la secuencia nucleotídica mostrada en el capítulo 10 (Figura 10.1). Con recuadros, se indican los sitios potenciales de N-glicosilación (NX{T/S}, Bause, 1983), en grisado aquellos que están conservados (en posición) en todos los arnavirus (ver capítulo 12). Con una barra se señala el sitio de reconocimiento de la proteasa de origen celular. Con un doble subrayado se indican las dos hélices transmembrana predichas (en el péptido señal y en G2). En el esquema se muestran los tres productos proteicos que se generan por procesamiento postraduccional de la proteína GPC. Los sitios de clivaje proteolítico y las secuencias correspondientes han sido determinados por homología con LCM (Buchmeier *et al.*, 1987; Burns and Buchmeier, 1993).

Tabla 11.a. Composición aminoacídica del péptido señal y las glicoproteínas G1 y G2

Aminoácido	Péptido señal		G1		G2	
	Cant.	Frecuencia	Cant.	Frecuencia	Cant.	Frecuencia
Ala (A)	6	10,3 %	7	3,7 %	9	3,8 %
Arg (R)	1	1,8 %	6	3,1 %	11	4,7 %
Asn (N)	2	3,4 %	11	5,8 %	18	7,6 %
Asp (D)	0	0,0 %	12	6,3 %	12	5,1 %
Cys (C)	1	1,8 %	6	3,1 %	10	4,2 %
Gln (Q)	4	6,9 %	8	4,2 %	4	1,7 %
Glu (E)	2	3,4 %	7	3,7 %	9	3,8 %
Gly (G)	4	6,9 %	10	5,2 %	13	5,5 %
His (H)	0	0,0 %	9	4,7 %	8	3,4 %
Ile (I)	6	10,3 %	10	5,2 %	12	5,1 %
Leu (L)	8	13,8 %	21	11,0 %	29	12,3 %
Lys (K)	2	3,4 %	13	6,8 %	18	7,6 %
Met (M)	2	3,4 %	5	2,6 %	7	3,0 %
Phe (F)	7	12,1 %	12	6,3 %	11	4,7 %
Pro (P)	1	1,8 %	11	5,8 %	9	3,8 %
Ser (S)	4	6,9 %	14	7,3 %	19	8,1 %
Thr (T)	2	3,4 %	11	5,8 %	13	5,5 %
Trp (W)	0	0,0 %	5	2,6 %	7	3,0 %
Tyr (Y)	1	1,8 %	6	3,1 %	8	3,4 %
Val (V)	5	8,6 %	7	3,7 %	9	3,8 %

Agrupamiento según Karlin *et al.* (1990): en rojo los hidrofóbicos, en negro los hidrofílicos y en fucsia los ambivalentes (no son ni fuertemente hidrofílicos ni fuertemente hidrofóbicos)

El estudio de los perfiles de hidropatía, accesibilidad promedio y accesibilidad individual (PHDacc, Rost and Sander, 1994c) permite encontrar una mayor concentración de aminoácidos hidrofóbicos en la región delimitada por los residuos 20 a 40 (Figura 11.2).

El análisis por computadora de la potencial estructura secundaria de la secuencia aminoacídica particular, realizado con diferentes algoritmos, permite efectuar predicciones que resultan coincidentes con los datos experimentales conocidos para otros arenavirus.

De acuerdo al algoritmo de Klein *et al.* (1985) el dominio en cuestión corresponde a un polipéptido integral con un segmento transmembrana delimitado por los aminoácidos 20 a 39. En función del algoritmo de Rao & Argos (1986) la predicción correspondiente determina la existencia de una α hélice transmembrana entre los aminoácidos 18 a 39. Por último, algoritmos más actualizados y refinados, como el PHDhtm (Rost *et al.*, 1995) y el PHDsec (Rost and Sander, 1993a,b,c, 1994a,b,d), predicen la existencia de una α hélice transmembrana entre los aminoácidos 19

a 52 (PHDhtm) o dos α hélices distanciadas un aminoácido (12 a 41 y 43 a 52, PHDsec).

Como se puede ver, los datos de predicción son coherentes entre sí y, además, coinciden con los requerimientos estructurales mínimos de un péptido señal.

Estructura de las glicoproteínas G1 y G2

Considerando el sitio de reconocimiento para la proteasa de origen celular, caracterizado en LCM (Buchmeier *et al.*, 1987, Burns & Buchmeier, 1993) y en el virus Junín (Padula *et al.* (1996), que da origen a las glicoproteínas G1 y G2, se pueden definir los dominios correspondientes en Candid #1, con 193 y 234 aminoácidos respectivamente.

La glicoproteína G1 está constituida por 28,8 % de aminoácidos hidrofóbicos, 34,6 % de aminoácidos hidrofílicos y 36,6% de aminoácidos ambivalentes, mientras que la glicoproteína G2 está constituida por 28,9 % de aminoácidos hidrofóbicos, 33,9 % de aminoácidos hidrofílicos y 37,2 % de aminoácidos ambivalentes. La comparación numérica anterior, o un análisis más detallado (Tabla 11.a), no muestran diferencias significativas entre G1 y G2.

El estudio de la distribución de aminoácidos hidrofóbicos, mediante perfiles de hidropatía y accesibilidad muestra diferencias concretas entre las glicoproteínas G1 y G2.

La proteína G1 tiene una distribución

relativamente regular de aminoácidos hidrofóbicos y no hidrofóbicos, lo cual redunda en una mayor probabilidad promedio de exposición al solvente y sugiere que podría ser una proteína soluble (Figura 11.2).

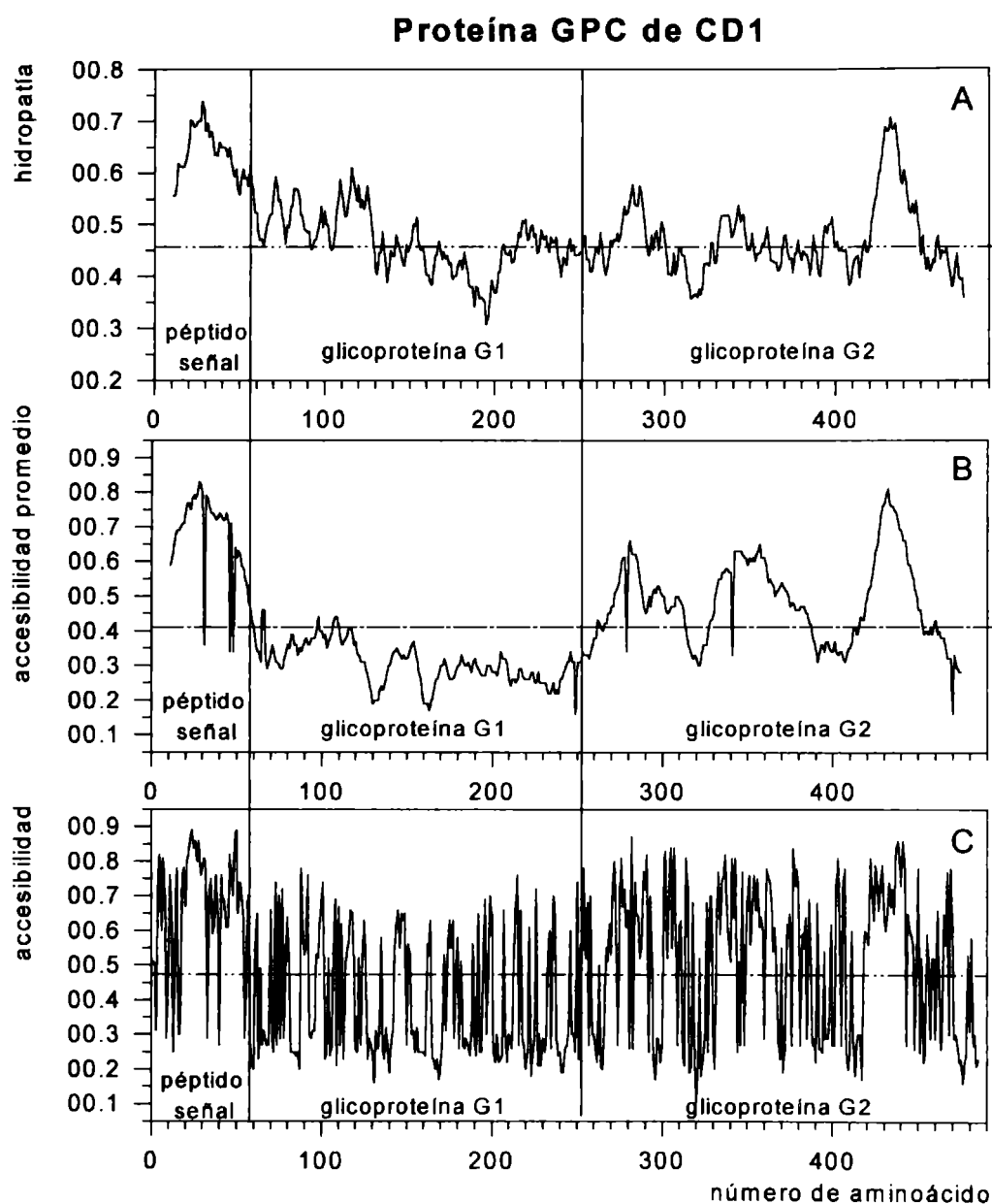


Figura 11.2. Análisis de hidropatía y accesibilidad de la proteína GPC de CD1. Las líneas verticales marcan las regiones de cada gráfico correspondientes a los dominios: péptido señal, glicoproteína G1 y glicoproteína G2. Las líneas quebradas corresponden a los valores promedio para la proteína completa. **A:** perfil de hidropatía calculado con el programa AASCALE (PC/Gené™, IntelliGenetics Inc.) en base a la tabla de hidrofobicidades de Kyte & Doolittle (1982), se utilizó una ventana de 21 residuos y los datos finales fueron normalizados a una escala de 0 a 1. **B** y **C:** perfiles de accesibilidad al solvente. Los datos de accesibilidad fueron calculados con el programa PHDacc (Rost & Sander, 1994c) y procesados posteriormente con el programa DATMANIP (Ghiringhelli, P.D., 2002, esta Tesis) para confeccionar ambos gráficos (ver Métodos y Miscelánea). Los valores fueron normalizados a una escala de 0 a 1, en donde a mayor valor menor accesibilidad. En **B** se representa la accesibilidad promedio calculada utilizando una ventana de 21 residuos. En **C** se representa la accesibilidad particular de cada residuo.

En la proteína G2, en cambio, la distribución es menos regular, existiendo grupos de aminoácidos hidrofóbicos. En particular, es necesario destacar la región delimitada por los residuos 168 a 198 (419-449 en GPC), rica en aminoácidos hidrofóbicos y con muy escasa probabilidad de exposición al solvente, las cuales son características típicas de una región transmembrana. Este conjunto de datos sugiere que G2 podría ser una proteína asociada a membranas (Figura 11.2).

Otra diferencia significativa entre las glicoproteínas G1 y G2 radica en la distribución de prolina y glicinas (Figura 11.3), lo cual sugiere que el plegamiento de ambas proteínas podría ser distinto. En G1, la distribución de prolina es aproximadamente uniforme, con sólo 3 regiones de rigidez relativamente baja (residuos 32-61, 72-101 y 104-145), mientras que la distribución de glicinas es asimétrica, con una mayor concentración en la mitad carboxilo terminal. En G2, en cambio, la distribución de prolina es muy asimétrica y la de glicinas es aproximadamente uniforme, definiendo 3 regiones de menor rigidez (residuos 17-102, 125-167 y 169-193)

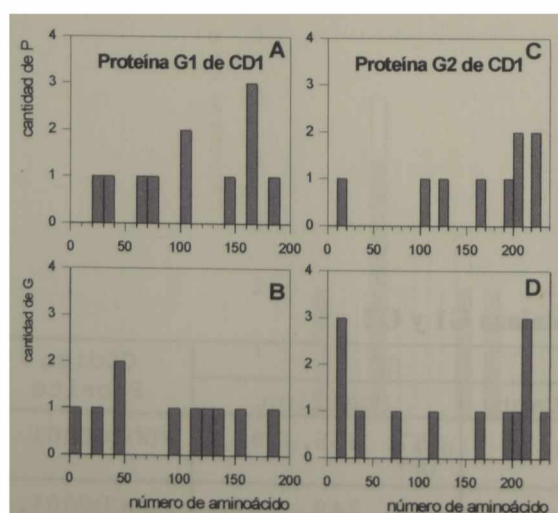


Figura 11.3. Distribución de prolina y glicinas en las glicoproteínas G1 y G2. Se graficó la cantidad del residuo particular cada 10 aminoácidos. **A:** distribución de prolina en G1. **B:** distribución de glicinas en G1. **C:** distribución de prolina en G2. **D:** distribución de glicinas en G2.

El análisis por computadora de la potencial estructura secundaria de la glicoproteína G1 señala a ésta como una proteína periférica (Klein *et al.*, 1985), con 3 regiones de α hélice, 8 regiones β hoja plegada y 12 regiones definidas como *loops*, donde podrían existir segmentos con estructura *random coil*, β -turn y/o γ -turn (PHDsec, Rost and Sander, 1993 a, b, c, 1994a, b, d) (Tabla 11.b).

El análisis por computadora de la potencial estructura secundaria de la glicoproteína G2 señala a ésta como una proteína integral multimérica, con un segmento transmembrana delimitado por los aminoácidos 171 a 188 (Klein *et al.*, 1985), 175 a 195 (Eisenberg *et al.*, 1984a), 169 a 193 (Rao & Argos, 1986) o 175 a 190 (PHDhtm, Rost *et al.*, 1995). Las predicciones del PHDsec (Rost and Sander, 1993 a, b c, 1994 a, b, d) revelan la existencia de 4 regiones de α hélice, 4 regiones β hoja plegada y 8 regiones definidas como *loops*, donde podrían existir segmentos con estructura *random coil*, β -turn y/o γ -turn (Tabla 11.c).

La búsqueda en la base de datos PROSITE (*PROtein SITEs*, Fuchs, 1991; Bairoch, 1993) predice la existencia de varios sitios potenciales de glicosilación, fosforilación, etc., tanto en G1 como en G2 (Tabla 11.d). Hasta el momento, no existen datos experimentales que demuestren la fosforilación o miristilación en G1 y/o G2, por lo cual se desconoce si estos sitios son biológicamente significativos. Con respecto a los sitios de N-glicosilación, se conoce que son utilizados, pero no se sabe cuántos ni cuáles de los 8 sitios potenciales están glicosilados *in vivo*.

Tabla 11.b. Estructuras secundarias predichas para la glicoproteína G1 de CD1

α hélices		
n°	posición	longitud
α0	15-22	8
α1	82-94	13
α2	127-136	10
β hojas plegadas		
n°	posición	longitud
β2-0	33-36	4
β2-1	40-44	5
β2-2	49-53	5
β2-3	57-61	5
β2-4	66-69	4
β3-1	113-122	10
β4-1	155-159	5
β4-2	176-180	5
loops		
n°	posición	longitud
L1	1-14	14
L2	23-32	10
L3	37-39	3
L4	45-48	4
L5	54-56	3
L6	62-65	4
L7	70-81	12
L8	95-112	18
L9	123-126	4
L10	137-154	18
L11	160-175	16
L12	181-193	11

La numeración de las estructuras α y β está basada en la comparación entre las G1 de todos los arenavirus (ver **Capítulo 13**)

Tabla 11.c. Estructuras secundarias predichas para la glicoproteína G2 de CD1

α hélices		
n°	posición	longitud
α1	19-33	15
α2	50-100	51
α3	142-163	22
α4	170-191	22
β hojas plegadas		
n°	posición	longitud
β1	3-6	4
β2	37-41	5
β3	108-114	7
β4	125-129	5
loops		
n°	posición	longitud
L1	7-18	12
L2	34-36	3
L3	42-49	8
L4	101-107	7
L5	116-124	9
L6	130-141	12
L7	164-169	6
L8	192-234	43

La numeración de las estructuras α y β está basada en la comparación entre las G2 de todos los arenavirus (ver **Capítulo 13**)

Tabla 11.d. Sitios potenciales identificados en las glicoproteínas G1 y G2

Sitio	G1		G2		Código Prosite
	Cant.	Posición	Cant.	Posición	
N-glicosilación	4	95, 105, 166, 178	4	357, 365, 382, 387	PDOC00001
{S/T} fosforilación (proteínquinasas y caseínquinasas)	5	107, 111, 136, 168, 179	9	206, 249, 257, 262, 317, 322, 412, 421, 446	PDOC00004, PDOC00005 y PDOC00006
N-miristoilación	3	79, 103, 183	3	330, 465, 466	PDOC00008

PROTEINA DE LA NUCLEOCAPSIDE (N)

Estructura de la proteína N

El marco de lectura abierto de polaridad viral complementaria existente en la mitad 3' del RNA S, delimitado por los codones AUG₈₂₋₈₄-UAA₁₇₇₄₋₁₇₇₆ (numeración VC), codifica para un polipéptido de 564 aminoácidos (Figura 11.4) y un peso molecular estimado de 63123 Da.

La proteína N está constituida por 28,7 % de aminoácidos hidrofóbicos, 37,6 % de aminoácidos hidrofílicos y 33,7 % de

aminoácidos ambivalentes, un análisis más detallado se presenta en la Tabla 11.e.

La carga neta estimada es de +9,5 a pH neutro, en clara dependencia de la relativa riqueza en Arg y Lys, aminoácidos que se hallan dispersos en toda la proteína N, con varios grupos de dos, tres y hasta cuatro residuos básicos (Figura 11.4). Esta característica está en relación con la función estructural de la proteína N de asociarse con el RNA viral para formar las nucleocápsides.

MAHSKEVPSF	RWTQSL	GL	SQFTQTVKSD	VLKDAKLIAD	SIDFNQEAQV	50
QRAL	T	G	EEDLNKLRDL	NKEVDRLMSM	RSVQRNTVFK	100
ELASDLEKLF	TKI	AETGS	QGVYMGNLSQ	SQLA	SEIL	150
GGNGVVRVWD	VKDPSKLNQ	FGSVPALTIA	CMTVQGGETM	NSVIQALTSL		200
GLLYTVKYPN	LSDLDRLTQE	HDCLQIVTKD	ESSINISGYN	FLSAAVKAG		250
ASILDGGNML	ETIRVTPDNF	SSLIKSTIQV	K	EGMFIDE	KPGNRNPYEN	300
LLYKLCLSGD	GWPYIG	<u>SRSO</u>	<u>IIGRS</u>	WDNTS	IDLT	350
LRLANLTEIQ	EAVIREAVGK	LDPTNTLWLD	IEGPATDPVE	MALFQPAGKQ		400
YIHC	PHD	EKGFKNGSRH	SHGILMKDIE	DAMPGVLSYV	IPLLPPDMVV	450
TTQGSDDI	LFDLHG	DL	KLVDVRLTSE	QARQFDQQVW	EKFGHLCKHH	500
NGVVVS		DKDAPFKLAS	SEPHCALLDC	IMFQSVLDGK	LYEEELTPLL	550
PPSLLFLPKA	AYAL					564

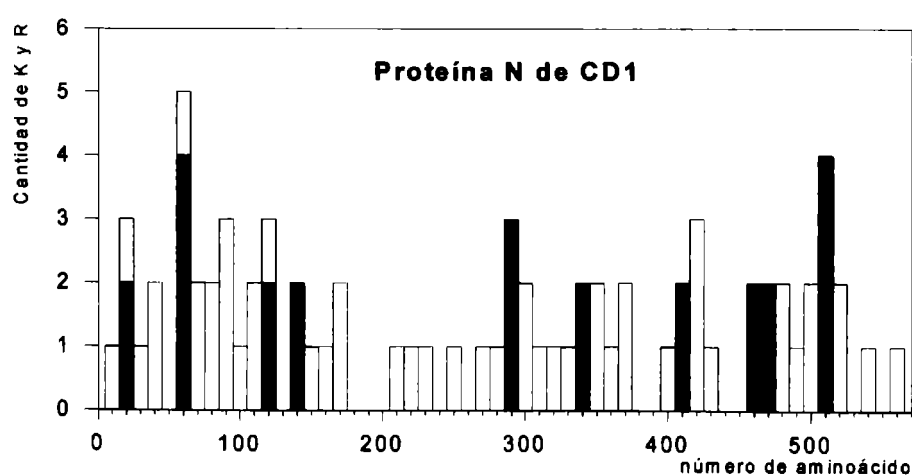


Figura 11.4. Proteína de la nucleocápside (N) de la cepa CD1 del virus Junín. La secuencia aminoacídica fue deducida a partir de la secuencia nucleotídica mostrada en el capítulo 10 (Figura 10.1). Con recuadros grisados se destacan los conjuntos de aminoácidos básicos Arg (R) y Lys (K). El doble subrayado indica la señal de dimerización predicha. El gráfico de barras muestra la cantidad de R y K cada 10 aminoácidos. Las barras blancas indican el total de los mismos. Las barras grisadas superpuestas corresponden a la fracción representada por uno o más conjuntos de 2, 3 o 4 aminoácidos básicos.

Tabla 11. e. Composición aminoacídica de la proteína N

Aminoácido	Cantidad	Frecuencia	Aminoácido	Cantidad	Frecuencia
Ala (A)	31	5,4 (%)	Leu (L)	66	11,7 (%)
Arg (R)	35	6,2 (%)	Lys (K)	41	7,3 (%)
Asn (N)	24	4,3 (%)	Met (M)	14	2,4 (%)
Asp (D)	39	6,9 (%)	Phe (F)	18	3,2 (%)
Cys (C)	7	1,2 (%)	Pro (P)	25	4,4 (%)
Gln (Q)	31	5,5 (%)	Ser (S)	42	7,4 (%)
Glu (E)	31	5,5 (%)	Thr (T)	29	5,1 (%)
Gly (G)	41	7,3 (%)	Trp (W)	6	1,0 (%)
His (H)	11	1,9 (%)	Tyr (Y)	11	1,9 (%)
Ile (I)	27	4,8 (%)	Val (V)	35	6,2 (%)

Agrupamiento según Karlin *et al.* (1990): en rojo los hidrofóbicos, en negro los hidrofílicos y en fucsia los ambivalentes (no son ni fuertemente hidrofílicos ni fuertemente hidrofóbicos)

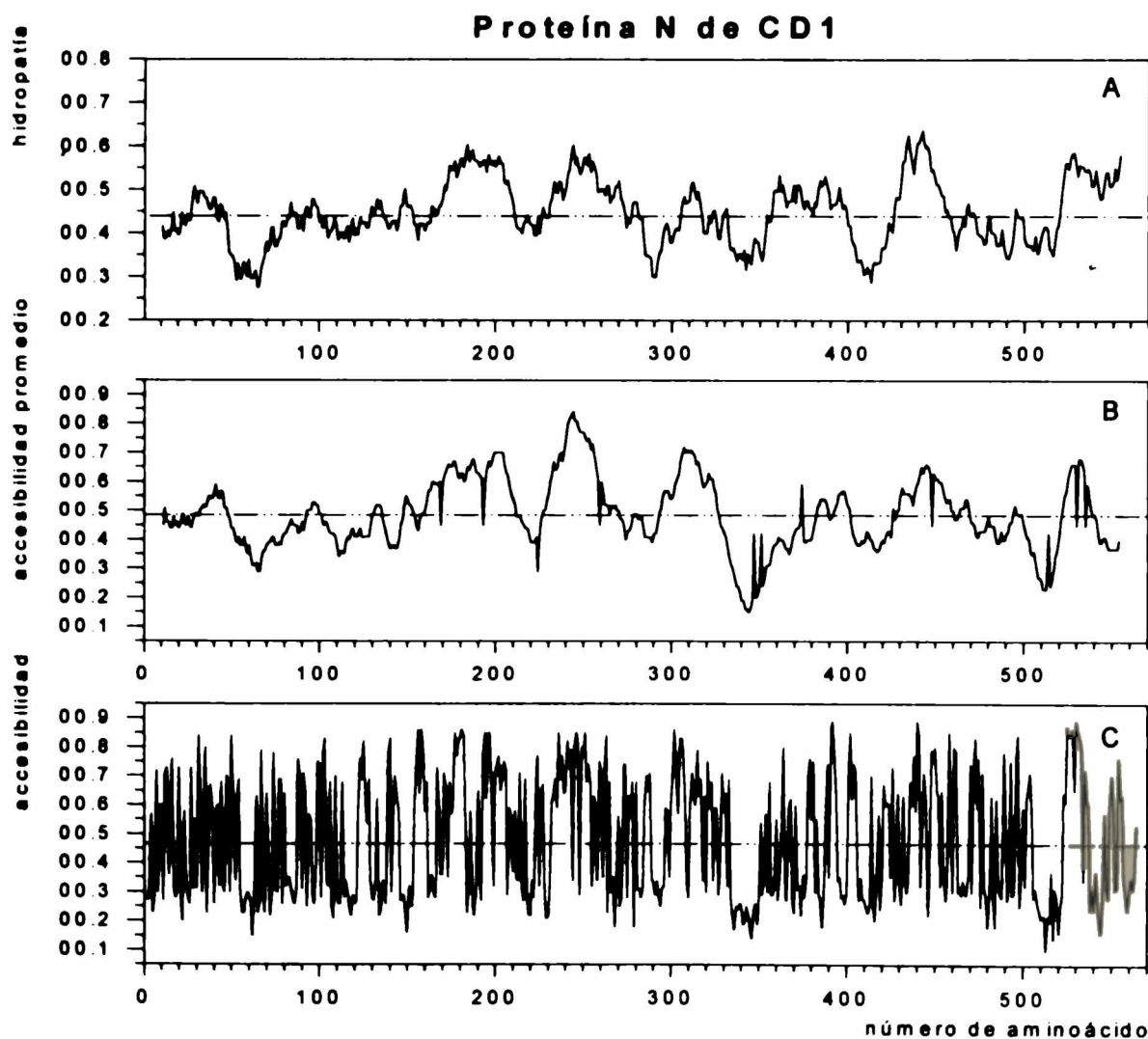


Figura 11.5. Análisis de hidropatía y accesibilidad de la proteína N de CD1. Las líneas quebradas corresponden a los valores promedio para la proteína completa. A: perfil de hidropatía calculado con el programa AASCALE (PC/Gene™, IntelliGenetics Inc.) en base a la tabla de hidrofobicidades de Kyte & Doolittle (1982), se utilizó una ventana de 21 residuos y los datos finales fueron normalizados a una escala de 0 a 1. B y C: perfiles de accesibilidad al solvente. Los datos de accesibilidad fueron calculados con el programa PHDacc (Rost & Sander, 1994c) y procesados posteriormente con el programa DATMANIP (Ghiringhelli, P.D., 2002, esta Tesis) para confeccionar ambos gráficos (ver Métodos y Miscelánea). Los valores fueron normalizados a una escala de 0 a 1, en donde a mayor valor menor accesibilidad. En B se representa la accesibilidad promedio calculada utilizando una ventana de 21 residuos. En C se representa la accesibilidad particular de cada residuo.

El análisis del perfil de hidropatía (**Figura 11.5**) muestra varias regiones de alta hidrofobicidad (residuos 28-48, 164-210, 231-273, 357-399, 426-456 y 520-564) y varias regiones de baja hidrofobicidad (residuos 49-83, 281-304, 332-353, 400-425, 472-494 y 496-519). Los perfiles de accesibilidad promedio y accesibilidad individual (**Figura 11.5**) permiten obtener resultados coincidentes con los datos de hidrofobicidad.

El estudio de la distribución de aminoácidos que pueden afectar directamente el plegamiento de la proteína, prolina (induce torsiones) y glicina (incrementa la flexibilidad), permite observar detalles estructurales interesantes (**Figura 11.6**). El análisis de la distribución de prolina señala la existencia de tres regiones de menor rigidez (residuos 9-163, 210-266 y 447-514), donde la ausencia de este residuo particular dificulta la formación de estructuras de tipo β -turn. La observación de la cantidad y distribución de glicinas sugiere que la flexibilidad de la cadena polipeptídica es alta.

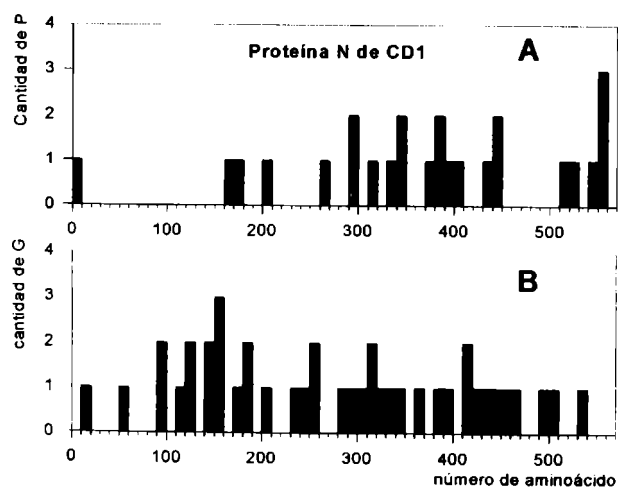


Figura 11.6. Distribución de prolina y glicinas en la proteína N. Se graficó la cantidad del residuo particular cada 10 aminoácidos. **A:** distribución de prolina. **B:** distribución de glicinas.

El estudio computacional de la potencial estructura secundaria de la proteína N indica la presencia de 17 regiones de α hélice, 17 regiones β hoja plegada y 35 regiones definidas como *loops*, donde podrían existir segmentos con estructura *random coil*, β -turn y/o γ -turn (PHDsec, Rost and Sander, 1993 a, b, c, 1994a, b, d) (**Tabla 11.f**).

La búsqueda en la base de datos PROSITE (Fuchs, 1991; Bairoch, 1993) predice la existencia de varios sitios potenciales de glicosilación, fosforilación, etc. (**Tabla 11.g**). Hasta el momento, no existen datos experimentales que demuestren la miristilación, la sulfatación o la amidación en N, por lo cual se desconoce si estos sitios son biológicamente significativos. Respecto a los sitios de N-glicosilación, se conoce que no son utilizados. Por otra parte, la señal de dimerización podría estar asociada con las interacciones proteína N-proteína N en la nucleocápside y/o con la interacción proteína N-proteína X (viral o celular) para el ejercicio de funciones regulatorias. Por último, en referencia a los sitios de fosforilación, no hay resultados experimentales concluyentes acerca de su utilización *in vivo*, pero existen evidencias que sugieren la presencia intracelular de la proteína N en dos estados, fosforilado (Bruns *et al.*, 1986) y no fosforilado. Esto último, podría estar en relación con posibles funciones regulatorias de la proteína.

Tabla 11.f. Estructuras secundarias predichas para la proteína N de CD1

α hélices			loops		
n°	posición	longitud	n°	posición	longitud
$\alpha 1$	11-19	9	L1	1-10	10
$\alpha 2$	27-39	13	L2	20-26	7
$\alpha 3$	45-55	11	L3	40-44	5
$\alpha 4$	64-79	16	L4	56-63	8
$\alpha 5$	96-113	18	L5	80-86	7
$\alpha 6$	129-143	15	L6	92-95	4
$\alpha 7$	176-185	10	L7	114-122	9
$\alpha 8$	193-200	8	L8	126-128	3
$\alpha 9$	241-254	14	L9	144-155	12
$\alpha 10$	271-282	12	L10	162-175	14
$\alpha 11$	356-370	12	L11	186-192	7
$\alpha 12$	426-431	6	L12	201-203	3
$\alpha 13$	437-444	8	L13	208-225	18
$\alpha 14$	457-465	9	L14	230-240	11
$\alpha 15$	480-500	21	L15	255-260	6
$\alpha 16$	524-538	14	L16	267-270	4
$\alpha 17$	556-558	3	L17	283-288	6
β hojas plegadas			L18	291-301	11
n°	posición	longitud	L19	309-313	5
$\beta 1$	87-91	5	L20	317-320	4
$\beta 2$	123-125	3	L21	324-330	7
$\beta 3$	156-161	6	L22	334-355	22
$\beta 4$	204-207	4	L23	371-377	7
$\beta 5$	226-229	4	L24	383-389	7
$\beta 6$	261-266	6	L25	396-401	6
$\beta 7$	288-290	3	L26	408-425	18
$\beta 8$	302-308	7	L27	432-436	5
$\beta 9$	314-316	3	L28	445-449	5
$\beta 10$	321-323	3	L29	454-456	3
$\beta 11$	331-333	3	L30	466-471	6
$\beta 12$	378-382	5	L31	477-479	3
$\beta 13$	390-395	6	L32	501-503	3
$\beta 14$	402-407	6	L33	509-523	15
$\beta 15$	450-453	4	L34	539-555	17
$\beta 16$	472-476	5	L35	559-564	6
$\beta 17$	504-508	5			

Tabla 11.g. Sitios potenciales identificados en la proteína N

Sitio	Cantidad	Posición	Código Prosite
N-glicosilación	8	127, 210, 235, 240, 269, 328, 355, 416	PDOC00001
{S/T} fosforilación (proteínquinasas y caseínquinasas)	13	9, 15, 26, 57, 79, 104, 205, 212, 228, 252, 262, 329, 535	PDOC00005 y PDOC00006
N-miristoilación	6	122, 149, 172, 201, 417, 502	PDOC00008
Amidación	1	465	PDOC00009
Señal de dimerización (Helix-loop-helix)	1	317	PDOC00038
Y-fosforilación Tirosínquinasasa	1	295	PDOC00007

PARTE IV

ANALISIS COMPARATIVO DE SECUENCIAS Y DISCUSION

Capítulo 12

Ácidos nucleicos de arnavirus - RNAs S

Sección I - Aspectos informativos y estructurales

Consideraciones generales

Usualmente, las caracterizaciones de secuencias se resumen en responder tres preguntas

1) ¿Mi secuencia es similar, homóloga, o muy diferente de las presentes en los bancos de datos?

Una vez identificados uno o más marcos de lectura,

2) ¿Qué puedo aprender de la secuencia proteica, con respecto a sus características, estructura y función?

Y, finalmente,

3) ¿Qué más puedo aprender de la secuencia nucleotídica no caracterizada?

Normalmente, se pone mucho énfasis en responder mejor las primeras dos preguntas.

Respecto a la última pregunta, en general, en la bibliografía virológica, las secuencias nucleotídicas son descritas mediante caracterizaciones estándar:

- ⇒ Longitud de las secuencias
- ⇒ Composición nucleotídica porcentual
- ⇒ Presencia de marcos de lectura
- ⇒ Frecuencia de uso de codones
- ⇒ Presencia de secuencias regulatorias
- ⇒ Sitios de unión de ribosomas (en RNA)
- ⇒ Estructuras secundarias (por ej. en RNA)
- ⇒ Comparaciones con bancos de datos
- ⇒ Construcción de dendrogramas o árboles filogenéticos

En este momento cabría hacerse una pregunta adicional:

- ¿Es suficiente todo lo anterior?

Teniendo en cuenta que los ácidos

nucleicos, al igual que las proteínas, son macromoléculas informativas por excelencia, conocer una secuencia nucleotídica novedosa implica disponer de información particular (en la secuencia *per se*) y general (en la familia de secuencias a la cual pertenece).

En síntesis, un esquema adicional de trabajo es aquél que utiliza un análisis basado en la teoría de la información, que estudia las propiedades de los mensajes (States and Boguski, 1991).

Desde este punto de vista, la información encerrada en la secuencia nucleotídica debe ser extraída y analizada, en particular y en general, en función de distintos parámetros que pueden considerarse informativos:

⇒ Complejidad local de composición:

- En principio, es una medida global de la organización de la información (Konopka and Owens, 1990). Analizada en forma comparativa, en una familia de moléculas, es una medida indirecta de los requerimientos informativos y las restricciones evolutivas existentes.
- En segundo lugar, la identificación de regiones de alta complejidad y su posible correlación con regiones de secuencia constantes, permitirían identificar zonas con escasa variabilidad, potencialmente útiles para metodologías de detección.
- En tercer lugar, el análisis particular de regiones de baja complejidad en moléculas nucleotídicas de virus con genoma de RNA, permitiría identificar regiones donde, potencialmente, podría patinar la RNA polimerasa-RNA dependiente (*slippage*).

- ⇒ Regiones constantes y variables:
 - La determinación de las mismas, en forma independiente de la complejidad o no de la información, es una medida indirecta de cuán crucial es el mensaje particular para el organismo en cuestión.
- ⇒ La información acerca de la distribución de nucleótidos individuales, o agrupados de diferentes maneras: purinas vs. pirimidinas, A+U (o T) vs. C+G, oligonucleótidos de corta longitud, etc. son, además, detalles informativos adicionales (Karlin and Cardon, 1994; Robinson *et al.*, 1995).
- ⇒ Las repeticiones, directas e invertidas, posiblemente están en relación con detalles estructurales asociados con funciones particulares y con regiones potencialmente recombinogénicas (Karlin and Cardon, 1994; Blinov *et al.*, 1994).
- ⇒ Por último, la existencia de periodicidad en la información puede estar en relación, también con cuestiones estructurales o funcionales.

Las consideraciones anteriores no pretenden agotar la discusión acerca de estos temas, sino simplemente destacar algunas de las bases teóricas de análisis.

Comparación de los RNAs S de arenavirus

Tamaño y composición

Los RNAs S de los 11 arenavirus analizados poseen una longitud promedio de 3413 nucleótidos (3375 nt. en LCA y LCE - 3534 nt. en OLI) (Tabla 12.a), con un contenido promedio de A+U de 55,4 % (53,9 % en LCA - 57,2 % en SAB) y de G+C de 44,6 % (42,8 % en SAB - 46,1 % en LCA).

Homología de secuencias

La homología global entre los RNAs S completos de arenavirus distintos (Tabla 12.b) oscila entre 46,3 % en PIC/LCA y 68,5 % en CD1/TAC. Como es lógico esperar, estos porcentajes se incrementan significativamente cuando se comparan distintas cepas del mismo arenavirus: 98,2 % en CD1/MC2, 77,9 % en LAG/LAJ y 84,5 % en LCA/LCE.

Tabla 12.a. Longitud y composición nucleotídica de los RNAs S de arenavirus

Arenavirus	Long. RNA S	A	C	G	U
CD1	3413	905	805	717	986
MC2	3400	903	809	716	972
OLI	3534	958	825	743	1008
PIC	3418	916	768	762	972
SAB	3366	907	713	727	1019
TAC	3421	876	751	769	1025
LAG	3417	946	799	707	965
LAJ	3402	909	769	762	972
LCA	3375	868	810	747	950
LCE	3375	881	789	724	981
MOP	3419	875	819	736	989
Composición media		904	787	737	985
Composición porcentual promedio		26,5 %	23,0%	21,6 %	28,9 %

CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia

Tabla 12.b. Homología de secuencias nucleotídicas

	NUEVO MUNDO						VIEJO MUNDO				
	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
CD1	100 %	98,2 %	54,8 %	50,7 %	62,5 %	68,5 %	48,3 %	47,9 %	47,7 %	48,1 %	47,9 %
MC2		100 %	54,1 %	50,3 %	62,3 %	67,8 %	48,3 %	47,6 %	47,3 %	47,6 %	47,8 %
OLI			100 %	55,8 %	52,5 %	53,1 %	48,0 %	48,1 %	47,7 %	48,6 %	49,1 %
PIC				100 %	50,8 %	50,8 %	47,9 %	47,3 %	46,3 %	46,8 %	48,5 %
SAB					100 %	60,2 %	47,4 %	47,5 %	47,6 %	46,7 %	47,5 %
TAC						100 %	48,1 %	48,3 %	46,8 %	47,5 %	47,5 %
LAG							100 %	77,9 %	57,8 %	57,8 %	65,7 %
LAJ								100 %	58,6 %	57,7 %	66,5 %
LCA									100 %	84,5 %	59,1 %
LCE										100 %	59,1 %
MOP											100 %

CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia
 En negrita y grisado se destacan las comparaciones entre distintas cepas de un mismo arnavirus

Por otra parte, la homología nucleotídica de los RNAs S de miembros del nuevo mundo oscila entre 50,3 % en MC2/PIC y 68,5 % en CD1/TAC, mientras que la correspondiente a miembros del viejo mundo oscila entre 57,7 % en LAJ/LCE y 66,5% en LAJ/MOP. Asimismo, cuando se comparan los integrantes del nuevo

mundo con los del viejo mundo, los porcentajes de homología son inferiores: 46,3 % en PIC/LCA y 49,1 % en OLI/MOP.

Cuando se analiza la distribución de la homología en función de un alineamiento múltiple (ver Apéndice 1) o de un gráfico resumen (Figura 12.1), se observa que las regiones de

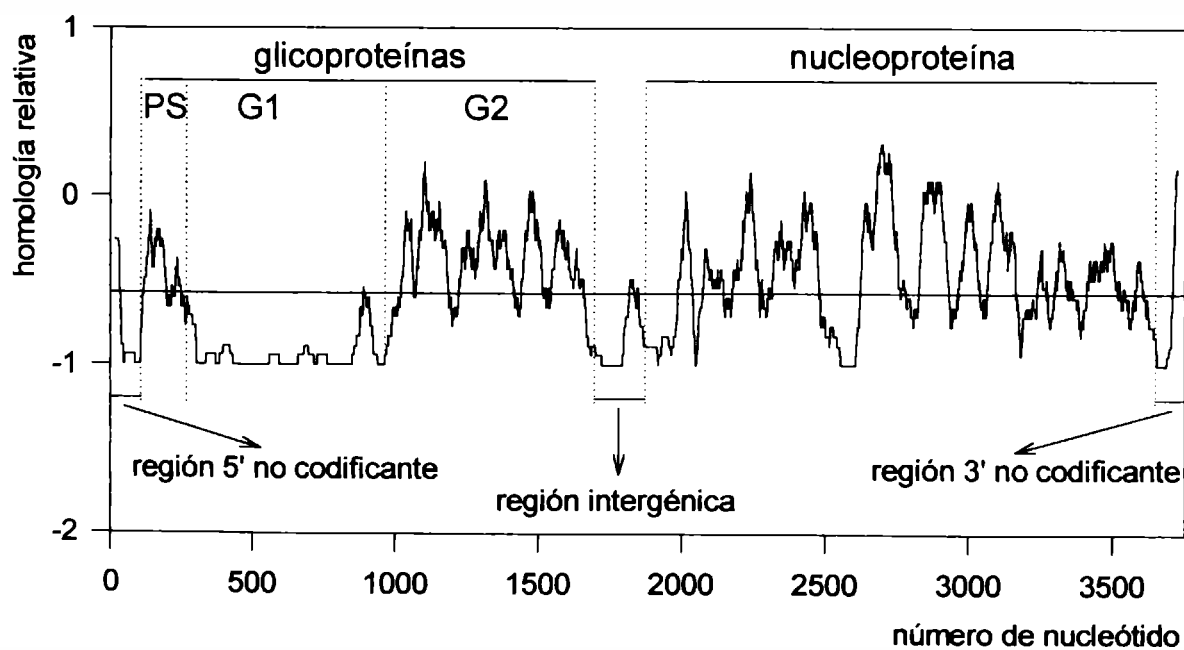


Figura 12.1. Regiones variables y regiones conservadas en los RNAs S de los arnavirus Junín MC2, Junín Candid #1, Oliveros, Pichindé, Sabiá, Tacaribe, Lassa Josiah, Lassa Nigeria, LCM Armstrong, LCM WE y Mopeia. El gráfico fue calculado con la rutina HOMGRAPH (Ghiringhelli, P.D., 2002, esta tesis), utilizando una ventana de 35 nucleótidos y una tabla de equivalencias de +1,0 para las identidades y -1,0 para la ausencia de identidad. La línea llena horizontal indica la media de los valores de homología relativa. Los picos por encima de la media corresponden a regiones conservadas y los picos por debajo corresponden a las regiones variables. En el marco de lectura que codifica para la proteína de la nucleocápside (N) se observa una alternancia de regiones con mayor o menor grado de conservación y regiones variables. En el marco de lectura que codifica para el precursor de las glicoproteínas (GPC), en cambio, es donde se encuentran los mayores contrastes; se observa una región conservada correspondiente al péptido señal (PS), seguida de una gran región no conservada (glicoproteína G1) y una región con alternancia de zonas conservadas y no conservadas (glicoproteína G2). En las regiones no codificantes se observan pequeños conjuntos de nucleótidos conservados en los extremos no codificantes.

mayor conservación corresponden en gran medida a las regiones codificantes para la proteína de la nucleocápside, la glicoproteína G2 y el péptido señal del precursor de las glicoproteínas. Mientras que las regiones de mayor variabilidad corresponden a la región codificante para la glicoproteína G1 y las regiones no codificantes.

Es interesante destacar que los pequeños conjuntos de nucleótidos conservados en las regiones intergénicas se correlacionan con la presencia de estructuras equivalentes, como se verá más adelante.

Complejidad local de la información

Desde este punto de vista, los patrones de LCC (*local composition complexity*) se encuentran muy conservados entre los RNAs S completos de los distintos arnavirus.

La homología global de LCC (Tabla 12.c) oscila entre 62,2 % en OLI/SAB y 67,1 % en LAG/LCE. Como es lógico esperar, estos porcentajes se incrementan cuando se comparan distintas cepas del mismo arnavirus: 71,9 % en CD1/MC2, 68,5 % en LAG/LAJ y 73,2 % en LCA/LCE.

Por otra parte, la homología de LCC de los RNAs S de miembros del nuevo mundo oscila entre 62,2 % en OLI/SAB y 66,3 % en MC2/SAB, mientras que la correspondiente a miembros del viejo mundo oscila entre 66,2 % en

LAJ/LCA y 67,1% en LAG/LCE. Asimismo, cuando se comparan los integrantes del nuevo mundo con los del viejo mundo, los porcentajes de homología oscilan entre 64,3 % en SAB/LCA y 66,3 % en MC2/LAG.

Cuando se analiza el patrón global de LCC en un gráfico resumen general de la familia (Figura 12.2.A), generado en función de un alineamiento múltiple (ver Apéndice 1 y Capítulo 18), se observa que las regiones de mayor complejidad (LCC > de 0,900 hasta un máximo de 0,994) corresponden a las zonas codificantes. Mientras que las regiones de menor complejidad ($\leq 0,900$) corresponden a zonas de muy baja homología y, en general, con abundantes *gaps*. Estas son las regiones no codificantes y un segmento de aproximadamente 60 nucleótidos cercanos a la región codificante para el extremo amino terminal de la glicoproteína G1 (Figura 12.2.B). Por otra parte, en gráficos resumen correspondientes a los arnavirus del Nuevo Mundo y del Viejo Mundo (Figuras 12.3 y 12.4), generados en función de un alineamiento múltiple, se observa que en los del Nuevo Mundo hay mayor dispersión que en los del Viejo Mundo. Esto puede deberse a la existencia dentro del grupo del Nuevo Mundo de virus pertenecientes a diferentes linajes (A, B o C).

Tabla 12.c. Homología de LCC de los RNAs S de arnavirus

	NUEVO MUNDO						VIEJO MUNDO				
	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
CD1	100 %	71,9 %	64.5 %	65.9 %	65.9 %	64.9 %	65.5 %	65.8 %	65.1 %	65.2 %	64.8 %
MC2		100 %	63.8 %	65.5 %	66.3 %	65.0 %	66.3 %	65.7 %	65.1 %	65.2 %	65.2 %
OLI			100 %	64.6 %	62.2 %	63.8 %	65.1 %	64.9 %	64.6 %	64.5 %	64.9 %
PIC				100 %	64.2 %	65.7 %	65.1 %	65.8 %	65.4 %	65.6 %	66.2 %
SAB					100 %	63.6 %	66.0 %	65.5 %	64.3 %	65.1 %	64.7 %
TAC						100 %	64.5 %	64.8 %	65.2 %	65.1 %	64.5 %
LAG							100 %	68,5 %	66.3 %	67.1 %	66.5 %
LAJ								100 %	66.2 %	66.4 %	66.7 %
LCA									100 %	73,2 %	66.8 %
LCE										100 %	66.4 %
MOP											100 %

CD1: Junin-Candid #1; MC2: Junin-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia
En negrita y grisado se destacan las comparaciones entre distintas cepas de un mismo arnavirus

En la **Figura 12.5** se representan los patrones de LCC individuales de las cepas MC2 y Candid #1 del virus Junín. Como se puede observar, los perfiles son casi idénticos, con algunas ligeras diferencias, principalmente en el

marco de lectura de la glicoproteína G2, y dos zonas de complejidad muy diferente en el marco de lectura de la glicoproteína G1 y en la zona complementaria al marco de lectura de la proteína N.

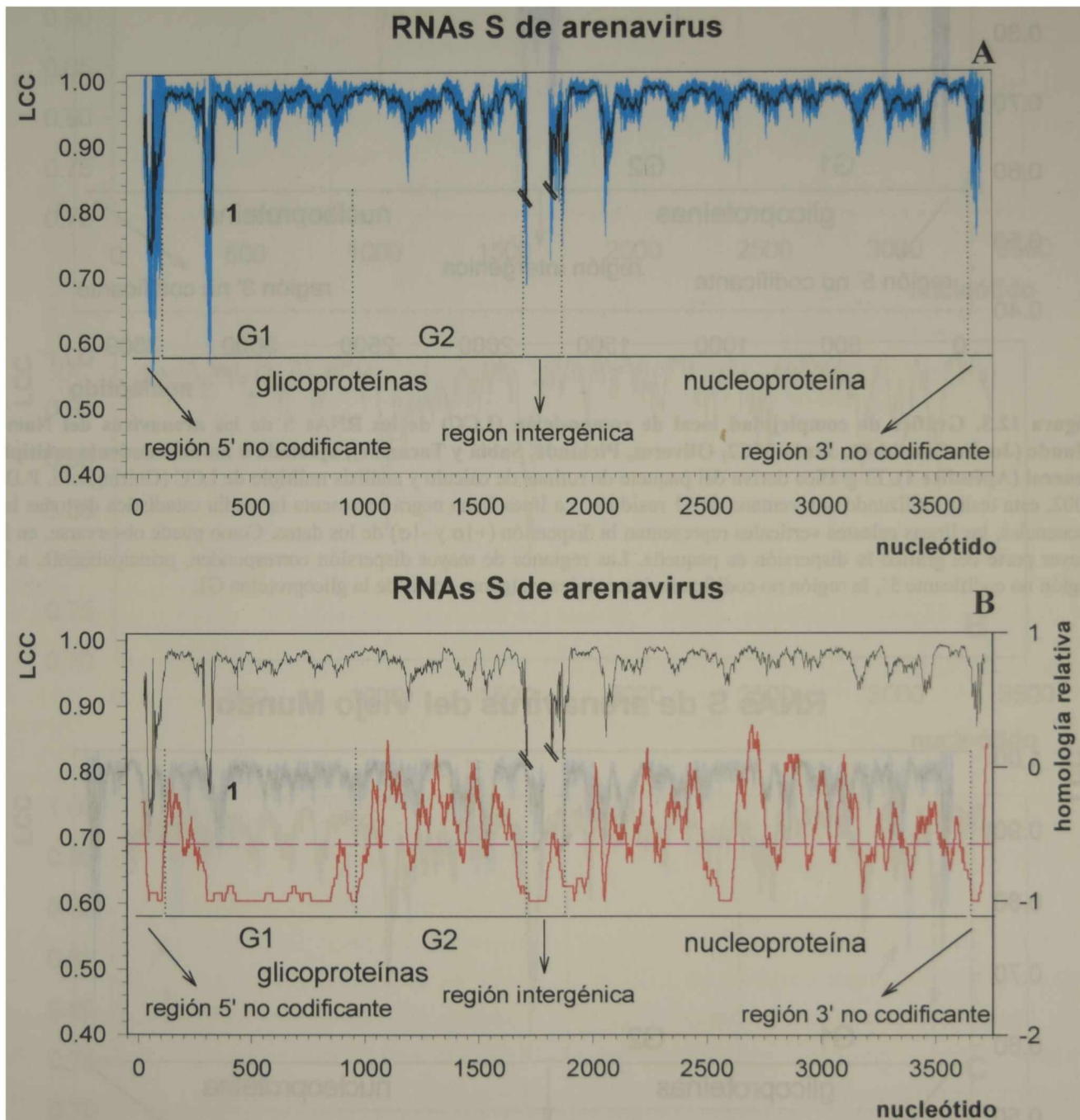


Figura 12.2. Análisis de complejidad local de composición. A: Gráfico de complejidad local de composición (LCC) de los RNAs S de 11 arnavirus (Junín-Candid #1, Junín-MC2, Oliveros, Pichindé, Sabiá, Tacaribe, Lassa-Nigeria, Lassa-Josiah, LCM-Armstrong, LCM-WE y Mopeia) (ver **Apéndice 4: gráficos de LCC individuales**), ajustado a un alineamiento múltiple general (**Apéndice 1**). El gráfico deriva del paquete de rutinas de cálculo y análisis múltiple de LCC (Ghiringhelli, P.D., 2002, esta tesis), utilizando una ventana de 51 residuos. La línea llena negra representa la media estadística de todas las secuencias, las líneas celestes verticales representan la dispersión ($+1\sigma$ y -1σ) de los datos. Como puede observarse, en la mayor parte del gráfico la dispersión es pequeña. Las regiones de mayor dispersión corresponden, principalmente, a las regiones no codificantes 5' y 3', la región no codificante intergénica y algunas zonas de menor complejidad en el resto de las secuencias. En particular, es notoria la dispersión de los datos en una región de baja complejidad en el marco de lectura correspondiente a la glicoproteína G1 (1). Por otra parte, es de destacar que los arnavirus del nuevo mundo y el virus Mopeia, a diferencia de los otros arnavirus, presentan, al menos, dos estructuras tipo *hairpin-loop* en la región intergénica. Por esta razón, el gráfico ha sido segmentado en la región intergénica. B: Gráfico A (en negro) y **Figura 12.1** (en rojo) sobrepuestas para mostrar la correlación existente entre la conservación de patrones de LCC y la homología nucleotídica (ver leyenda **Figura 12.1**). La región marcada como 1 es una zona caracterizada por la existencia de inserciones y deleciones cuando se comparan los distintos arnavirus entre sí. Es de destacar que, aunque son dos cepas del mismo virus, lo mismo ocurre entre Junín-MC2 y Junín-Candid #1.

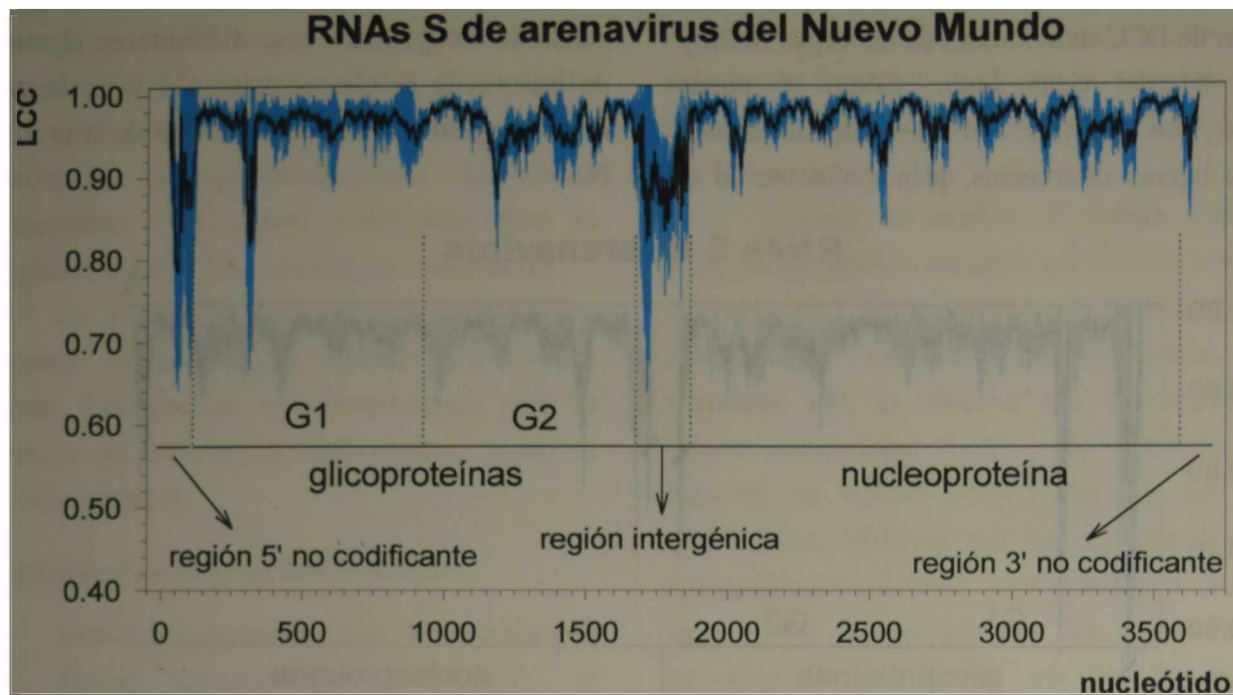


Figura 12.3. Gráfico de complejidad local de composición (LCC) de los RNAs S de los arnavirus del Nuevo Mundo (Junín-Candid #1, Junín-MC2, Oliveros, Pichindé, Sabiá y Tacaribe), ajustado a un alineamiento múltiple general (Apéndice 1). El gráfico deriva del paquete de rutinas de cálculo y análisis múltiple de LCC (Ghiringhelli, P.D., 2002, esta tesis), utilizando una ventana de 51 residuos. La línea llena negra representa la media estadística de todas las secuencias, las líneas celestes verticales representan la dispersión ($+1\sigma$ y -1σ) de los datos. Como puede observarse, en la mayor parte del gráfico la dispersión es pequeña. Las regiones de mayor dispersión corresponden, principalmente, a la región no codificante 5', la región no codificante intergénica y algunas zonas de la glicoproteína G1.

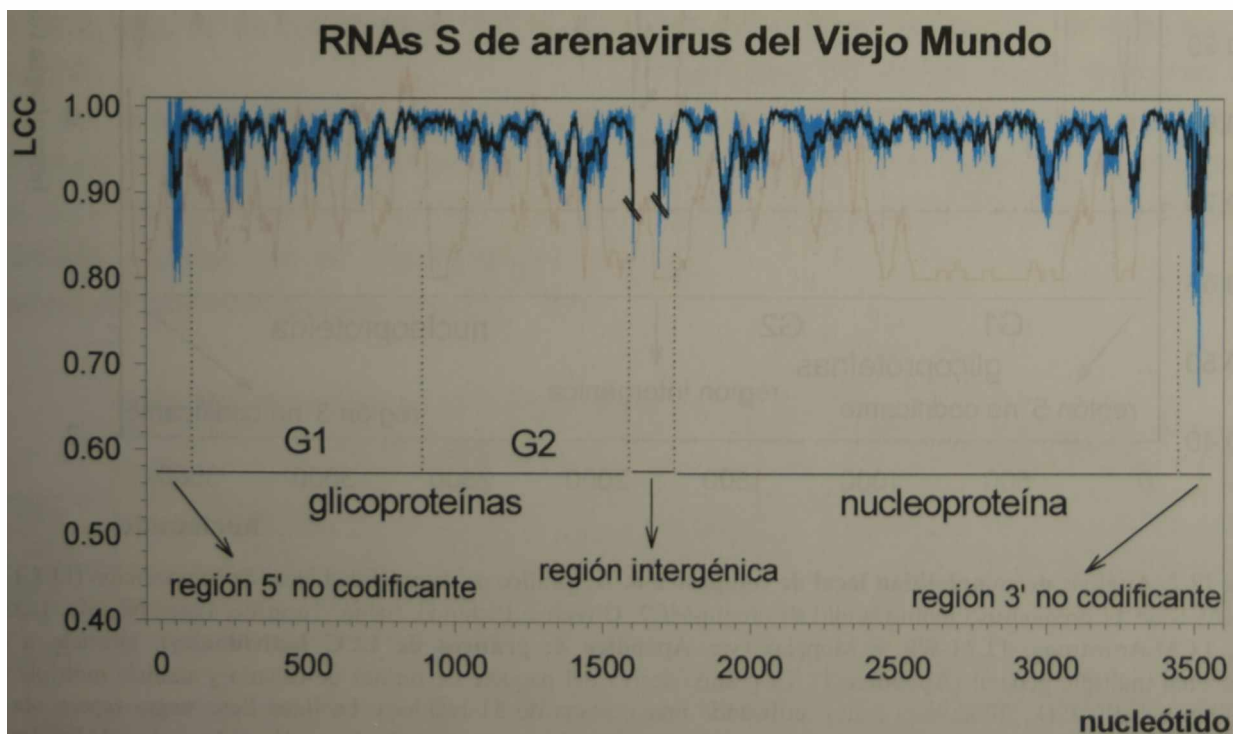


Figura 12.4. Gráfico de complejidad local de composición (LCC) de los RNAs S de los arnavirus del Viejo Mundo (Lassa-Nigeria, Lassa-Josiah, LCM-Armstrong, LCM-WE y Mopeia), ajustado a un alineamiento múltiple general (Apéndice 1). El gráfico deriva del paquete de rutinas de cálculo y análisis múltiple de LCC (Ghiringhelli, P.D., 2002, esta tesis), utilizando una ventana de 51 residuos. La línea llena negra representa la media estadística de todas las secuencias, las líneas celestes verticales representan la dispersión ($+1\sigma$ y -1σ) de los datos. Como puede observarse, en la mayor parte del gráfico la dispersión es pequeña. Las regiones de mayor dispersión corresponden, principalmente, a las regiones no codificantes 5' y 3', la región no codificante intergénica y algunas zonas de menor complejidad en el resto de las secuencias. Es de destacar que el virus Mopeia, a diferencia de los otros virus del Viejo Mundo, presenta dos estructuras tipo *hairpin-loop* en la región intergénica. Por esta razón, el gráfico ha sido segmentado en la región intergénica.

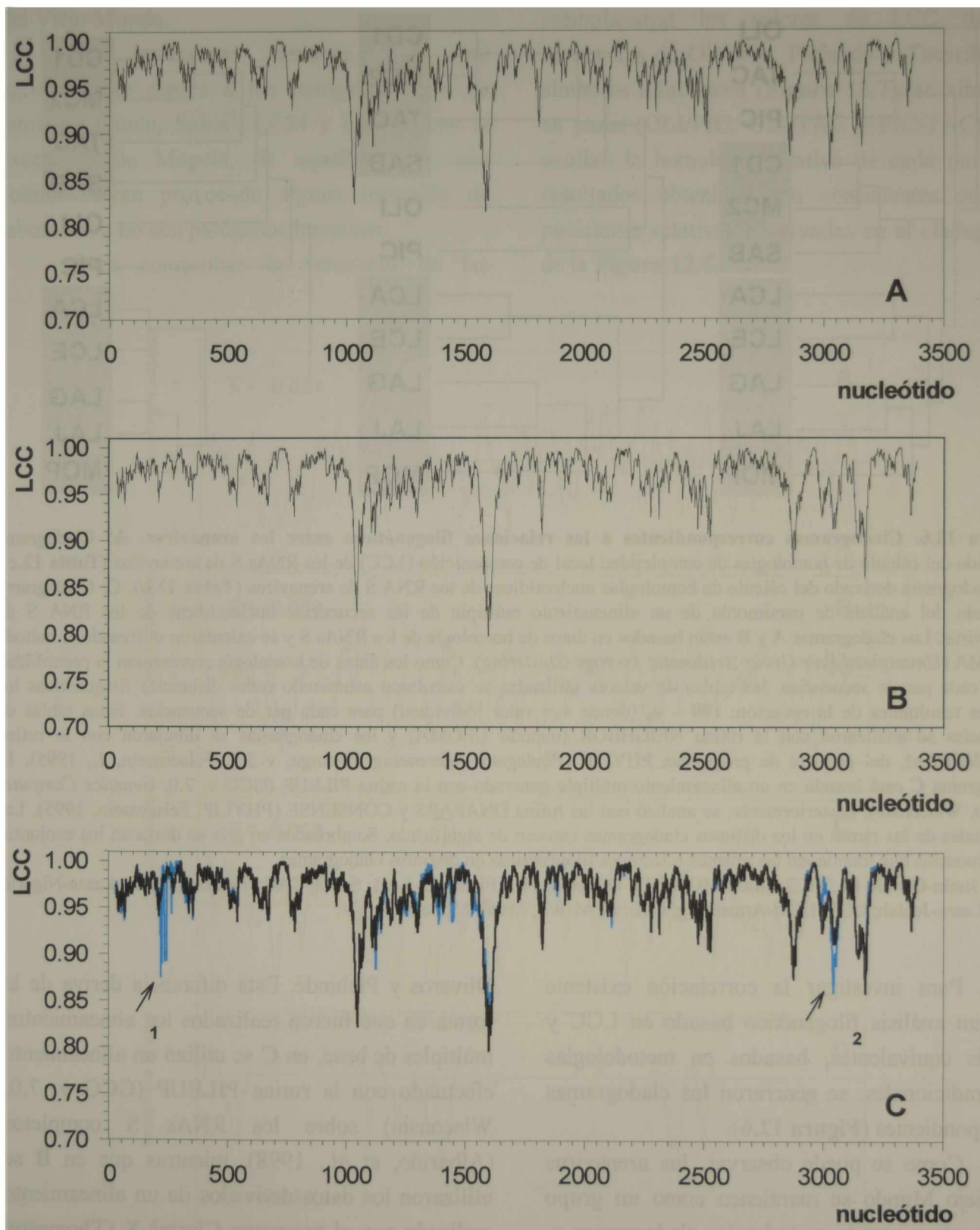


Figura 12.5. Gráficos de complejidad local de composición (LCC) de los RNAs S de las cepas MC2 y Candid #1 del virus Junín. A. RNA S de la cepa MC2. **B.** RNA S de la cepa Candid #1. **C.** Gráficos de ambas cepas superpuestos, en color celeste la cepa MC2 y en color negro la cepa Candid #1. Ambos gráficos derivan de los individuales ajustados al alineamiento correspondiente (Apéndice 1). Como se puede observar, ambos gráficos son casi idénticos. Con flechas se indican las dos regiones de mayor diferencia; **1:** corresponde a una zona donde existen inserciones y deleciones en ambas cepas en el marco de lectura de la glicoproteína G1 y **2:** corresponde a una región donde existen varias transiciones y transversiones en la región complementaria al marco de lectura de la proteína N. Todos los gráficos se generaron con el paquete de rutinas de cálculo y análisis múltiple de LCC (Ghiringhelli, P.D., 2002, esta tesis), utilizando una ventana de 51 residuos.

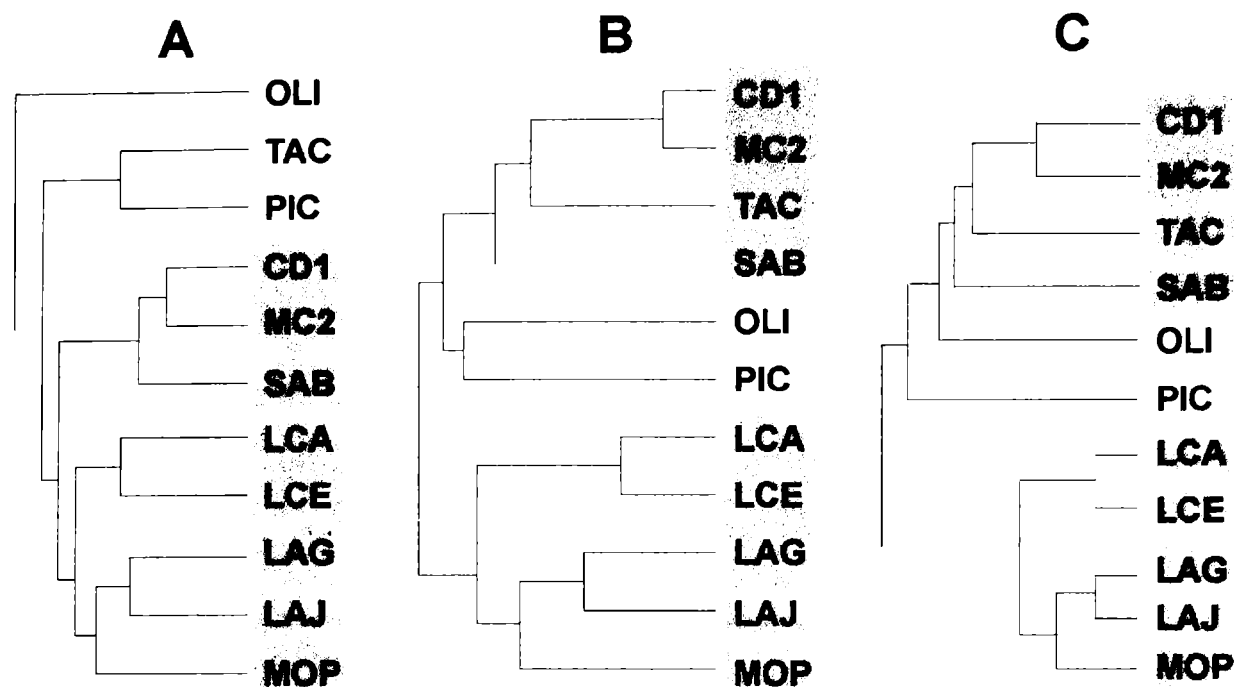


Figura 12.6. Cladogramas correspondientes a las relaciones filogenéticas entre los arnavirus. A: Cladograma derivado del cálculo de homologías de complejidad local de composición (LCC) de los RNAs S de arnavirus (Tabla 12.c). B: Cladograma derivado del cálculo de homologías nucleotídicas de los RNA S de arnavirus (Tabla 12.b). C: Cladograma derivado del análisis de parsimonia de un alineamiento múltiple de las secuencias nucleotídicas de los RNA S de arnavirus. Los cladogramas A y B están basados en datos de homología de los RNAs S y se calcularon utilizando el método UPGMA (*Unweighted Pair Group Arithmetic Average Clustering*). Como los datos de homología representan la proximidad entre cada par de secuencias, las tablas de valores utilizadas se calcularon asumiendo como distancias filogenéticas los valores resultantes de la ecuación: $100 - v_n$ (donde v_n = valor individual) para cada par de secuencias. Estas tablas de distancias se analizaron con la rutina NEIGHBOR (método UPGMA) y los cladogramas se dibujaron con la rutina DRAWGRAM, del paquete de programas PHYLIP (*Phylogeny Inferencing Package*, v 3.5, Felsenstein, J., 1995). El cladograma C está basado en un alineamiento múltiple generado con la rutina PILEUP (GCG v. 7.0, *Genetics Computer Group*, Wisconsin), posteriormente, se analizó con las rutinas DNAPARS y CONSENSE (PHYLIP, Felsenstein, 1995). Las longitudes de las ramas en los distintos cladogramas carecen de significado. Sombreados en gris se destacan los conjuntos de secuencias que mantienen las mismas relaciones filogenéticas en distintos cladogramas.

CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia.

Para investigar la correlación existente entre un análisis filogenético basado en LCC y análisis equivalentes, basados en metodologías más tradicionales, se generaron los cladogramas correspondientes (Figura 12.6).

Como se puede observar, los arnavirus del Viejo Mundo se mantienen como un grupo claramente separado en todos los cladogramas y, además, las relaciones filogenéticas entre ellos también son las mismas.

Por otra parte, los arnavirus del Nuevo Mundo muestran un comportamiento diferente según el tipo de análisis empleado. Desde este punto de vista es importante destacar que la única diferencia entre los cladogramas B y C (Figura 12.6) corresponde a la ubicación relativa de

Oliveros y Pichindé. Esta diferencia deriva de la forma en que fueron realizados los alineamientos múltiples de base, en C se utilizó un alineamiento efectuado con la rutina PILEUP (GCG v. 7.0, Wisconsin) sobre los RNAs S completos (Albariño, *et al.*, 1998), mientras que en B se utilizaron los datos derivados de un alineamiento realizado con el programa Clustal X (Thompson *et al.*, 1994, 1997), en forma independiente para las regiones codificantes y no codificantes (ver Apéndice 1). Sin embargo, en ambos cladogramas, tanto Oliveros como Pichindé aparecen ligados al resto de los arnavirus del Nuevo Mundo. En el cladograma A, la posición de los arnavirus Oliveros, Pichindé y Tacaribe es absolutamente distinta (Figura 12.6) siendo

ubicados todos como posibles antecesores del resto de los arnavirus del Nuevo Mundo y de los del Viejo Mundo.

Es interesante destacar que este agrupamiento separa a los patógenos netos de humanos (Junín, Sabiá¹, LCM y Lassa), con la excepción de Mopeia, de aquéllos que, aún cuando hayan provocado alguna infección de laboratorio, no son patógenos humanos.

Para comprobar la veracidad de las

relaciones establecidas con la metodología anterior usando un método independiente, se reemplazaron los valores de LCC de las secuencias de Oliveros, Pichindé y Tacaribe por símbolos alfabéticos (Figura 12.7), se alinearon en pares (OLI/PIC, OLI/TAC y PIC/TAC) y se analizó la homología relativa de cada par. Los resultados obtenidos son consistentes con las posiciones relativas observadas en el cladograma de la Figura 12.6.

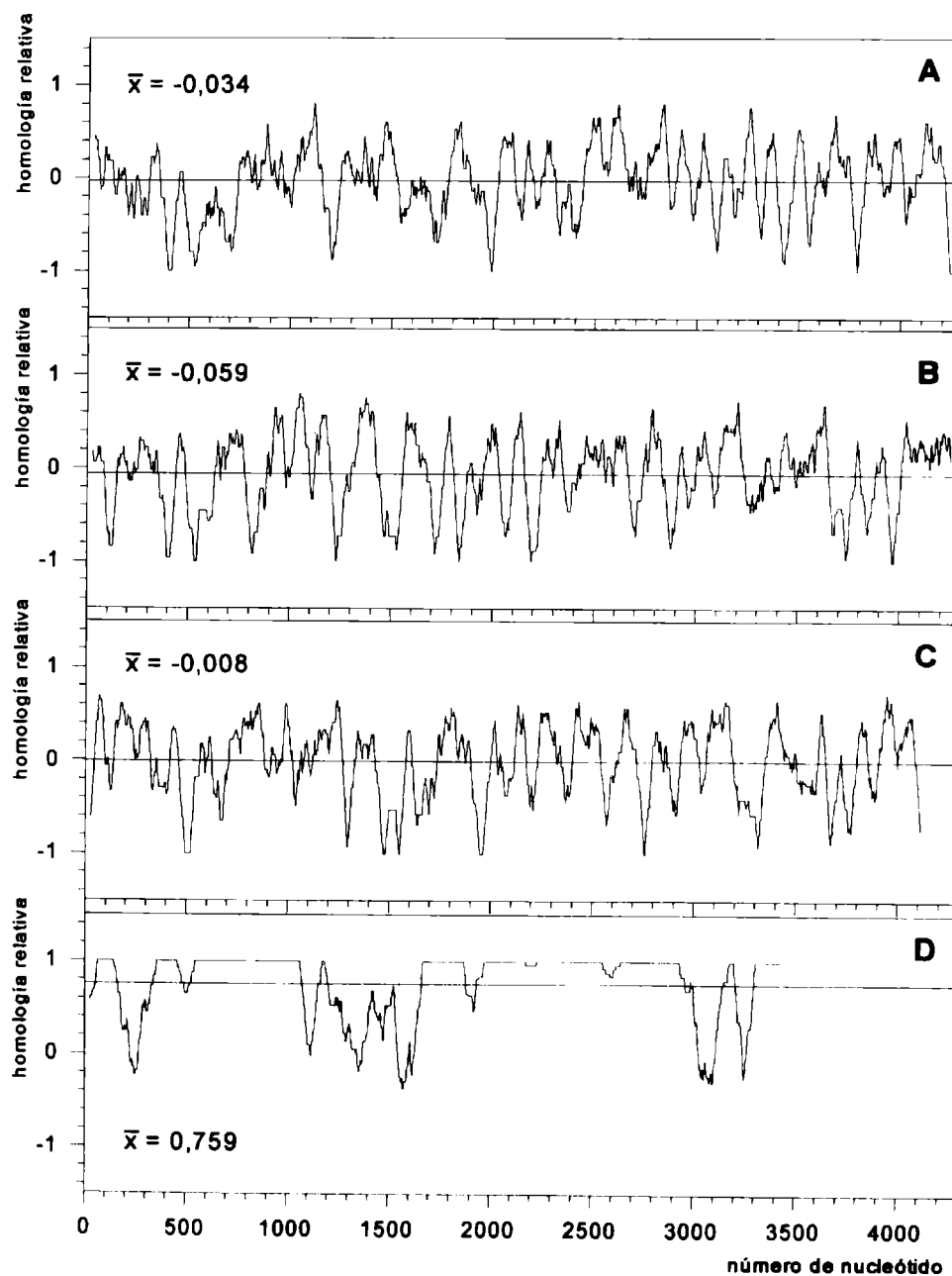


Figura 12.7. Gráficos de homología relativa de LCC de diferentes pares de RNAs S de arnavirus. A. Oliveros y Pichindé. B. Oliveros y Tacaribe. C. Pichindé y Tacaribe. D. Junín-Candid # 1 y Junín-MC2. Para realizar las comparaciones, los valores de LCC de cada secuencia se reemplazaron por símbolos alfabéticos correspondientes a 11 rangos (C: mayor de 0,9900; S: 0,9700-0,9899; P: 0,9500-0,9699; G: 0,9300-0,9499; D: 0,9100-0,9299; Q: 0,8900-0,9099; R: 0,8700-0,8899; M: 0,8500-0,8699; L: 0,8300-0,8499; F: 0,8100-0,8299; Y: menor de 0,8100). Las secuencias alfabéticas obtenidas se alinearon con el programa Clustal X (Thompson *et al.*, 1997), utilizando una matriz de identidades. Los consensos estrictos obtenidos se analizaron con la rutina HOMGRAPH (Ghiringhelli, P.D., 2002, esta tesis), utilizando una ventana de 51 residuos y una tabla de equivalencias de +1,0 para las identidades y -1,0 para la ausencia de identidad. La línea llena horizontal indica la media de los valores de homología relativa y el grado de similitud global de complejidad informativa.

¹ El virus Sabiá fue aislado en 1990 después de la muerte de un agrónomo en Brasil; posteriormente provocó dos infecciones de laboratorio, en Brasil y en USA (Coimbra *et al.*, 1994; González *et al.*, 1996; Peters, 1998).

Frecuencias nucleotídicas

En los arnavirus, las frecuencias de A+U vs C+G y su distribución a lo largo de las secuencias de los RNAs S corresponden, mayoritariamente, al rango de lo esperable para secuencias de orden aleatorio (Figura 12.8.A.).

Sin embargo, existen varias regiones cuyas frecuencias resultan características; de ellas, solo tres se destacan por su extensión. La región 1 (~ 250 nt.) y la región 2 (~ 80 nt.) corresponden a la región central del marco de lectura que codifica para la glicoproteína G2 (Figura 12.8.A.) y muestran una predominancia de A+U. La región 3 (~ 200 nt.), en cambio,

muestra una marcada predominancia de C+G y corresponde a la zona intergénica y regiones adyacentes.

Es de destacar que la predominancia señalada en esta última región está asociada con la existencia de las estructuras tipo *hairpin-loop* de alta estabilidad. Por otra parte, también se debe notar que en esta zona es donde existe una mayor abundancia de *gaps* en el análisis múltiple, motivo por el cual las extensiones reales para las secuencias individuales varían entre ~70 nt. (Lassa-Nigeria) hasta ~200 nt. (Junín-CD1 y Junín-MC2) (Figura 12.8.B. y Apéndice 4).

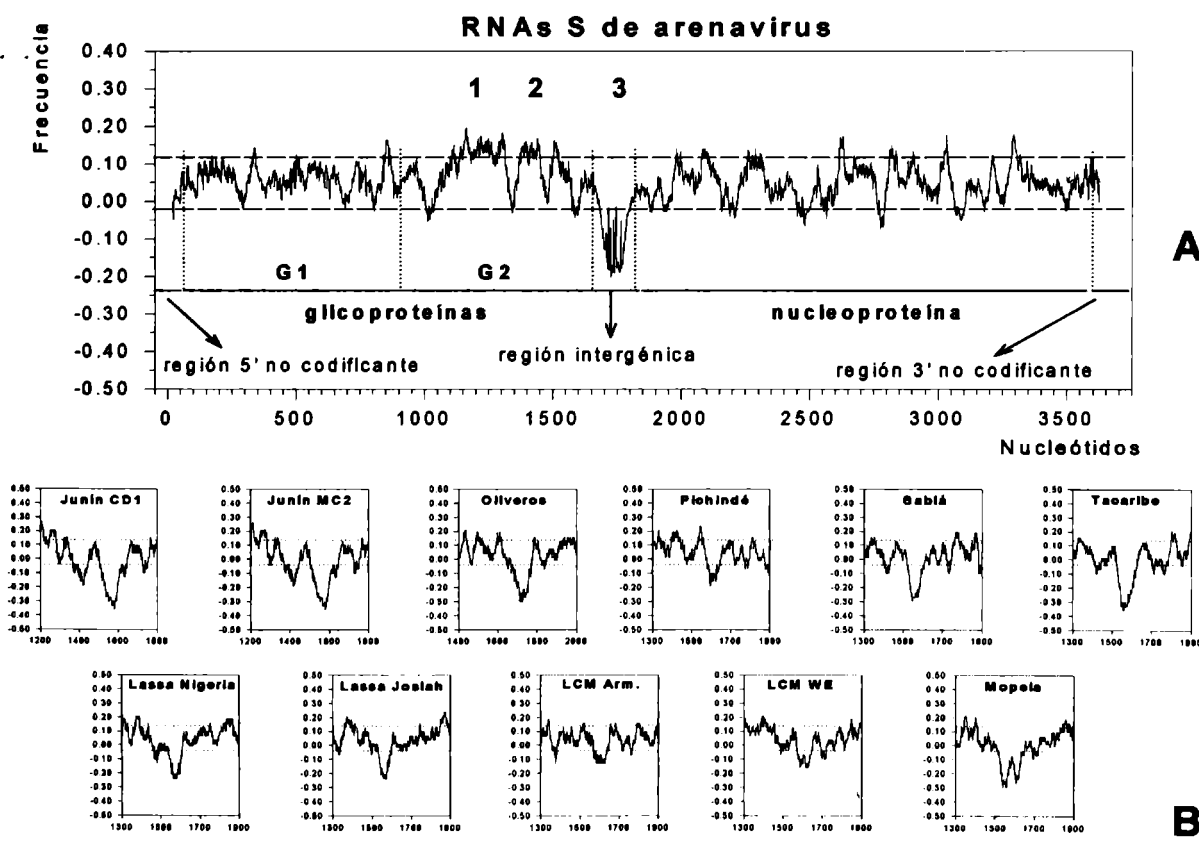


Figura 12.8. Gráficos de frecuencia nucleotídica relativa (A+U vs C+G). Todos los gráficos derivan del paquete de rutinas de cálculo y análisis múltiple de frecuencias nucleotídicas (Ghiringhelli, P.D., 2002, esta tesis). **A.** Gráfico resumen correspondiente al análisis conjunto de los RNAs S de los arnavirus Junín-Candid #1, Junín-MC2, Oliveros, Pichindé, Sabiá, Tacaribe, Lassa-Nigeria, Lassa-Josiah, LCM-Armstrong, LCM-WE y Mopeia, utilizando una ventana de cálculo de 35 residuos. Las líneas de guiones horizontales indican el rango de frecuencias esperables para secuencias nucleotídicas de organización aleatoria (ver Capítulo 18). Con números se señalan las tres regiones características de mayor extensión. La numeración nucleotídica corresponde a la del alineamiento múltiple (ver Apéndice 1). **B.** Detalle correspondiente a la región intergénica, donde existe una abundancia relativa de C+G característica para los arnavirus. Ventana de 35 residuos. Las líneas de guiones horizontales señalan el rango de frecuencias esperables para secuencias nucleotídicas de organización aleatoria. La numeración nucleotídica corresponde a la propia de cada secuencia de RNA.

Abundancia relativa de oligonucleótidos

Análisis global

La abundancia relativa (AR) de oligonucleótidos cortos (di-, tri- y tetranucleótidos) en los RNAs S de arnavirus muestra que la mayor parte los mismos se encuentran dentro del rango estadísticamente esperable (Karlín and Cardon, 1994).

Cuatro de los 16 dinucleótidos posibles, 4 de los 64 trinucleótidos posibles y 18 de los 256 tetranucleótidos posibles presentan algún valor diferencial (sobre- o sub-abundancia) en cualquiera de las polaridades (V o VC) (Tabla 12.d). En el caso de los dinucleótidos, las

sobre-abundancias (CA y UG) o las sub-abundancias (CG y UA) son consistentes para todos los arnavirus. Es de destacar que el dinucleótido CG es extremadamente sub-abundante, lo cual puede estar asociado, tanto con aspectos estructurales (formación de estructuras tipo *hairpin loop*) como con aspectos regulatorios (Karlín & Cardon, 1994).

Respecto de los tri- y tetranucleótidos, el patrón de sobre- y sub-abundancias es variable entre los diferentes arnavirus, algunos tri- o tetranucleótidos son sub-abundantes (o sobre-abundantes) para ciertos arnavirus mientras que son normales para otros. En

Tabla 12.d. Abundancia relativa de oligonucleótidos cortos en los RNAs S de arnavirus

1. Abundancias relativas de dinucleótidos

	NUEVO MUNDO						VIEJO MUNDO				
	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
CA	1,38	1,36	1,48	1,46	1,36	1,32	1,40	1,43	1,36	1,41	1,38
CG	0,25	0,27	0,24	0,30	0,30	0,35	0,24	0,28	0,31	0,24	0,29
UA	0,56	0,56	0,47	0,52	0,54	0,54	0,56	0,60	0,54	0,53	0,49
UG	1,38	1,36	1,48	1,46	1,36	1,32	1,40	1,43	1,36	1,41	1,38

2. Abundancias relativas de trinucleótidos

	NUEVO MUNDO						VIEJO MUNDO				
	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
ACG	0,71	0,67	0,89	0,88	0,84	0,67	0,72	0,82	0,60	0,86	0,70
CCG	1,26	1,31	1,07	1,04	1,21	1,25	1,09	1,18	0,98	1,02	1,37
CGG	1,26	1,31	1,07	1,04	1,21	1,25	1,09	1,18	0,98	1,02	1,37
CGU	0,71	0,67	0,89	0,88	0,84	0,67	0,72	0,82	0,60	0,86	0,70

3. Abundancias relativas de tetranucleótidos

	NUEVO MUNDO						VIEJO MUNDO				
	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
ACCG	0,58	0,65	1,08	1,35	1,35	0,73	1,44	1,15	0,70	0,69	1,20
ACGA	0,75	0,73	0,89	1,01	0,56	1,50	1,16	1,43	0,57	1,18	2,12
ACGU	1,64	1,62	0,00	1,79	1,73	0,68	0,66	0,54	1,10	0,85	0,00
CCGG	1,30	1,13	1,43	1,63	1,08	1,24	0,95	0,74	0,88	0,47	0,62
CGCA	0,52	0,53	1,24	1,26	0,79	1,01	1,22	1,28	1,31	0,78	1,02
CGCC	1,44	1,41	0,82	0,70	0,77	0,66	0,71	1,46	0,59	0,97	1,62
CGCG	2,60	2,05	1,38	0,67	1,20	2,61	1,08	0,00	0,00	0,00	0,00
CGGU	0,58	0,65	1,08	1,35	1,35	0,73	1,44	1,15	0,70	0,69	1,20
CGUA	1,31	1,06	0,87	0,74	0,88	1,39	0,47	0,45	0,63	1,09	0,57
CGUC	0,69	0,72	0,53	1,49	1,00	0,91	1,34	0,65	1,69	1,07	0,74
GACG	0,69	0,72	0,53	1,49	1,00	0,91	1,34	0,65	1,69	1,07	0,74
GCGA	1,27	0,91	1,13	0,75	1,42	0,80	0,70	0,63	0,78	0,39	0,34
GCGC	0,69	1,25	0,41	0,98	0,77	0,75	0,98	1,56	1,85	1,70	1,82
GGCG	1,44	1,41	0,82	0,70	0,77	0,66	0,71	1,46	0,59	0,97	1,62
UACG	1,31	1,06	0,88	0,74	0,88	1,39	0,47	0,45	0,63	1,09	0,57
UCGC	1,27	0,91	1,13	0,75	1,42	0,80	0,70	0,63	0,78	0,39	0,34
UCGU	0,75	0,73	0,89	1,02	0,56	1,50	1,16	1,43	0,57	1,18	2,12
UGCG	0,52	0,53	1,24	1,26	0,78	1,01	1,22	1,28	1,31	0,78	1,02

CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia
 Se muestran sólo las abundancias relativas significativas. En negrita y con fondo gris claro se destacan las abundancias con valores <0,78 o >1,23 (diferentes de lo estadísticamente esperable). En negrita y con fondo gris oscuro se destaca la ausencia absoluta de algunos tetranucleótidos.

particular, existen dos tetranucleótidos especialmente destacables. El tetranucleótido ACGU es sobre-abundante en Junín-Candid #1, Junín-MC2, Pichindé y Sabiá; normal en LCM-Armstrong y LCM-WE; sub-abundante en Tacaribe, Lassa-Nigeria y Lassa-Josiah; y ausente en Oliveros y Mopeia. Por otra parte, el tetranucleótido CGCG es extremadamente sobre-abundante en Junín-Candid #1, Junín-MC2 y Tacaribe; ligeramente sobre-abundante en Oliveros; normal en Sabiá y Lassa-Nigeria; ligeramente sub-abundante en Pichindé y ausente en Lassa-Josiah, LCM-Armstrong, LCM-WE y Mopeia.

Análisis diferencial por regiones

El análisis diferenciado por regiones (no codificantes y las posiciones I+II, II+III y III+I en la sucesión de codones), permite detectar sobre- o sub-abundancias locales que pueden asociarse con una estructura sintáctica de la información nucleotídica (**Figuras 12.9, 12.10, 12.11, 12.12 y 12.13**).

Es interesante destacar que las sobre- o sub-abundancias relativas de dinucleótidos detectadas a nivel global en los RNAs S de arnavirus presentan una predominancia diferencial en distintas regiones, y que éstas son las mismas en todos los arnavirus. El dinucleótido CA es especialmente sobre-abundante en las posiciones III+I de los marcos de lectura, el dinucleótido CG es especialmente sub-abundante en las regiones codificantes (I+II, II+III y III+I de los marcos de lectura), el dinucleótido UA es especialmente sub-abundante en las regiones no codificantes y en las posiciones I+II y II+III de los marcos de lectura y el dinucleótido UG es especialmente sobre-abundante en las posiciones II+III y III+I de los marcos de lectura. Esta homogeneidad en la distribución de sobre- y sub-abundancias refleja, probablemente, una similitud sintáctica en los RNAs S de los arnavirus. El resto de los 16 dinucleótidos posibles, presentan patrones de sobre-abundancia, sub-abundancia o abundancia normal más aleatorios,

probablemente asociados con la información particular de cada genoma.

Análisis diferencial por ventanas

Este tipo de análisis permite obtener perfiles de la distribución de la abundancia relativa de los distintos dinucleótidos en los genomas analizados.

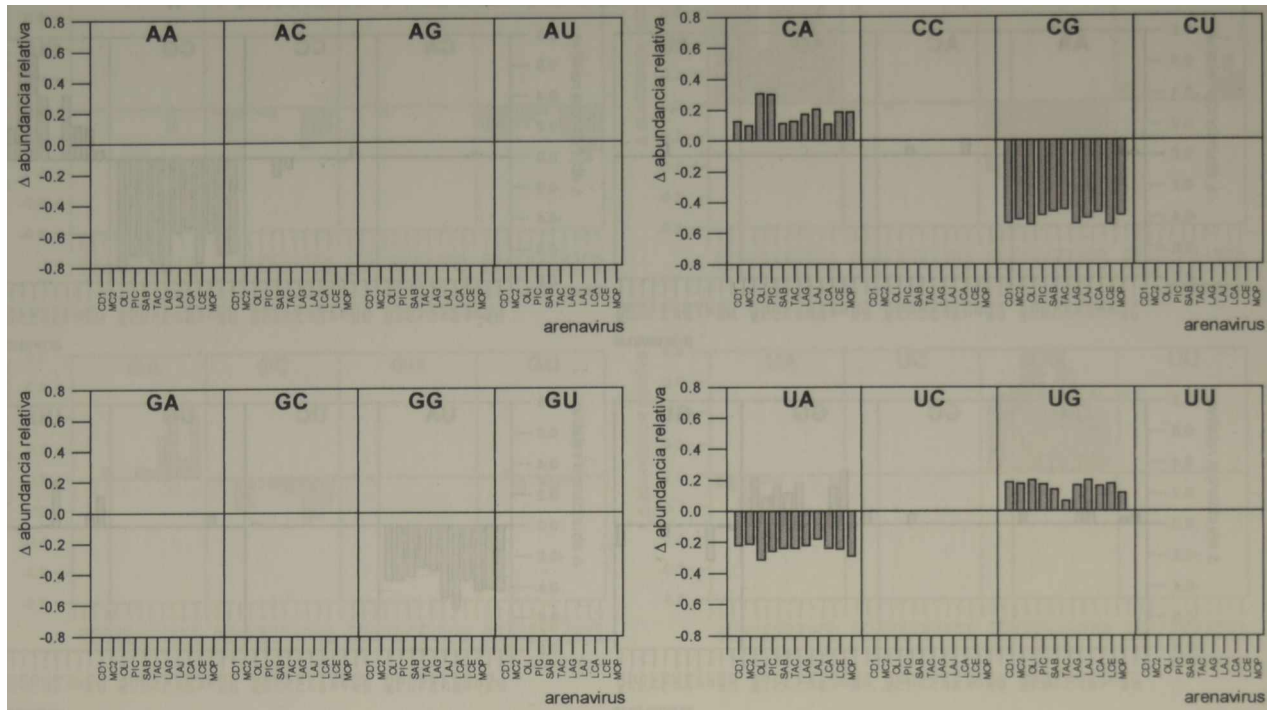
En una comparación general, se puede observar que, para la mayor parte de los dinucleótidos posibles, no existen perfiles de abundancia relativa comunes a todos los arnavirus (**Figuras 12.14.A, 12.14.B, 12.15.A y 12.15.B**). Tampoco se detectan perfiles comunes cuando se comparan dos cepas del mismo virus (Junín-Candid #1/Junín-MC2, Lassa-Nigeria /Lassa-Josiah o LCM-WE/LCM-Armstrong).

En una comparación más detallada, se puede observar que los 4 dinucleótidos significativos (CA, CG, UA y UG), detectados a nivel global en los RNAs S de arnavirus, presentan perfiles de distribución de sobre- o sub-abundancias relativas ligeramente similares en todos los arnavirus.

En particular, el dinucleótido CG presenta las menores sub-abundancias en las regiones no codificantes (extremos y región intergénica). Esto último guarda una íntima correlación con la existencia de estructuras secundarias doble cadena de gran estabilidad, asociadas con la conformación circular de las nucleocápsides (observada en microscopía electrónica) y con la regulación de los fenómenos de transcripción/replicación (Franze Fernández *et al.*, 1987; Ghiringhelli *et al.*, 1991; Rivera Pomar *et al.*, 1993).

Por otra parte, también es interesante destacar que, aún cuando no se observa ninguna periodicidad en los perfiles de los distintos dinucleótidos, sí se observa una asimetría bastante marcada de los perfiles de sobre- o sub-abundancia en la mayor parte de los dinucleótidos cuando se compara la mitad 5' vs. la mitad 3' de los genomas.

A Abundancia relativa de dinucleótidos en los RNAs S completos de arnavirus Polaridad viral



B Abundancia relativa de dinucleótidos en las regiones no codificantes de los RNAs S de arnavirus - Polaridad viral

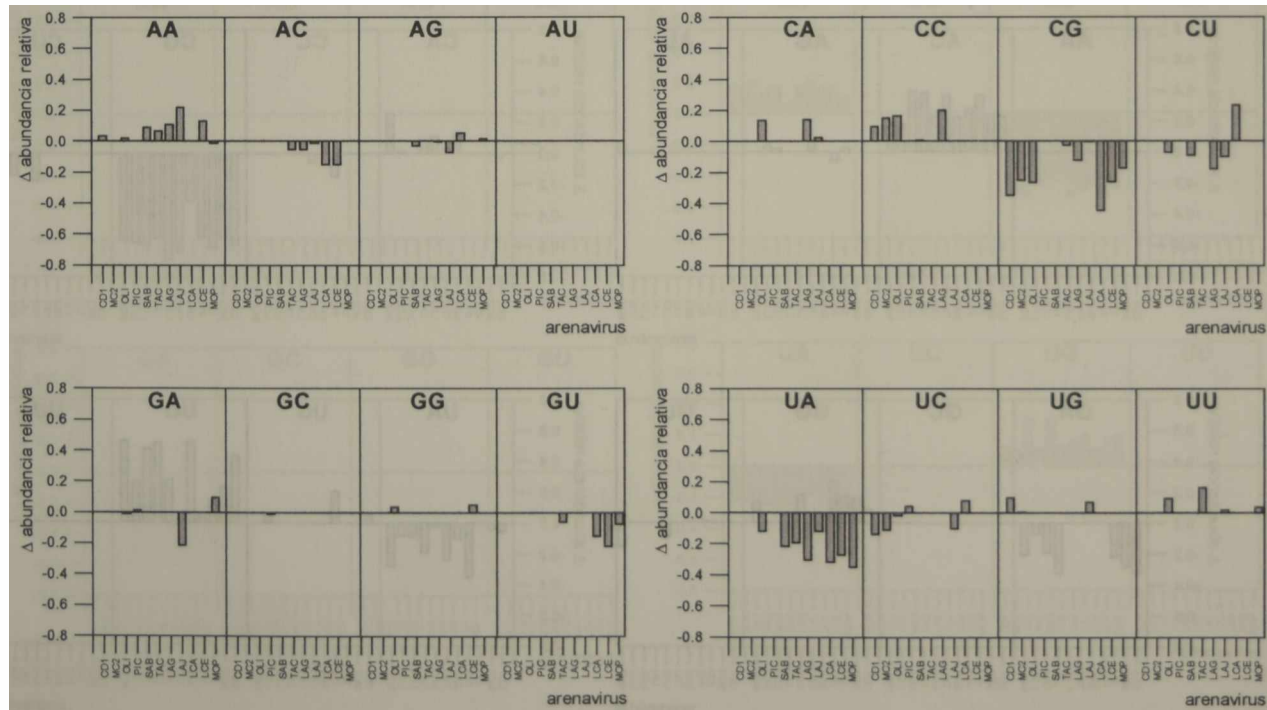


Figura 12.9. Abundancia relativa de dinucleótidos en los RNAs S de arnavirus. **A.** RNAs S completos analizados en polaridad viral. **B.** regiones no codificantes analizadas en la misma polaridad. El análisis se realizó con las rutinas diseñadas y escritas por Ghiringhelli, P.D. (2002), esta tesis. Posteriormente, para facilitar la discriminación rápida entre sub- y sobre-representaciones y la visualización gráfica, los datos originales se modificaron, llevando a 0 todos los valores comprendidos entre 0,78 y 1,23, y asignando valor positivo a la diferencia: $\text{valor}_n (>1,23) - 1,23$ y valor negativo a la diferencia: $0,78 - \text{valor}_n (<0,78)$. CDI1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia.

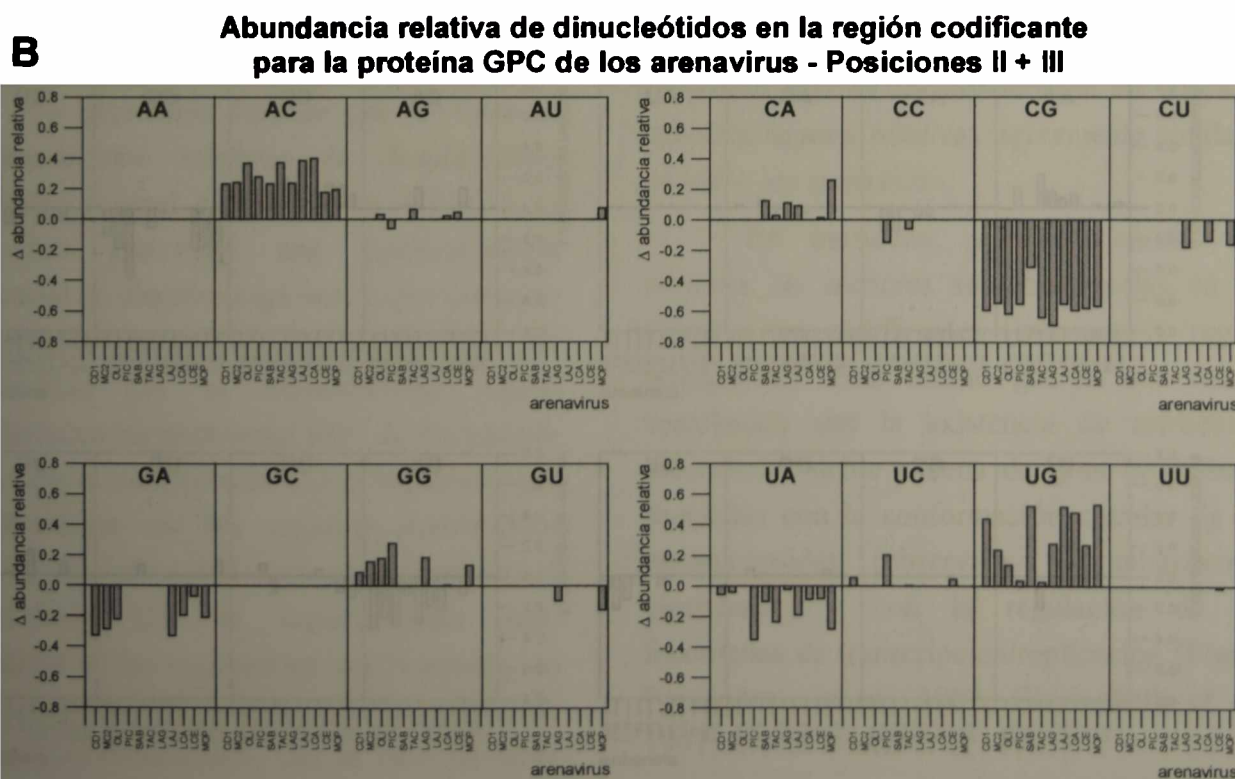
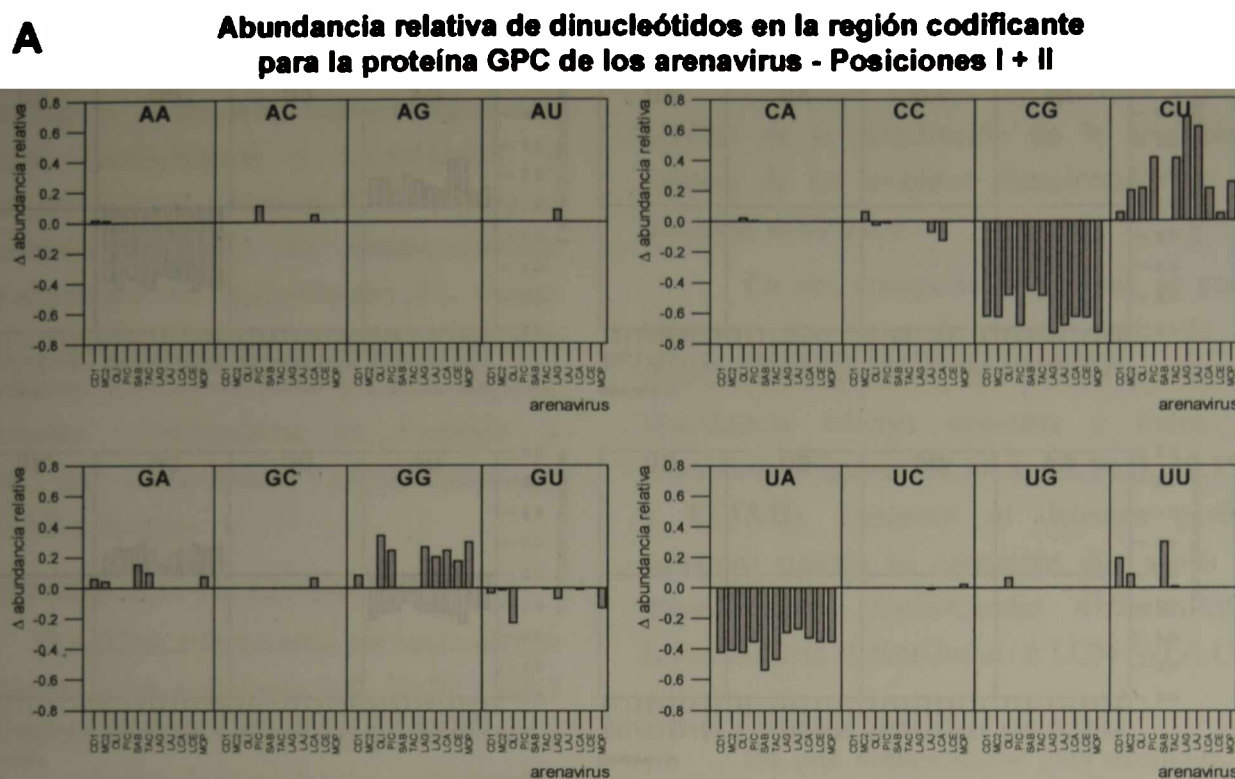


Figura 12.10. Abundancia relativa de dinucleótidos en los RNAs S de arnavirus. A. RNAs S analizados en el marco de lectura codificante para GPC, en las posiciones I + II. B. RNAs S analizados en el marco de lectura codificante para GPC, en las posiciones II + III. El análisis se realizó con las rutinas diseñadas y escritas por Ghiringhelli, P.D. (2002), esta tesis. Posteriormente, para facilitar la discriminación rápida entre sub- y sobre-representaciones y la visualización gráfica, los datos originales se modificaron, llevando a 0 todos los valores comprendidos entre 0,78 y 1,23, y asignando valor positivo a la diferencia: $\text{valor}_n (>1,23) - 1,23$ y valor negativo a la diferencia: $0,78 - \text{valor}_n (<0,78)$. CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia.

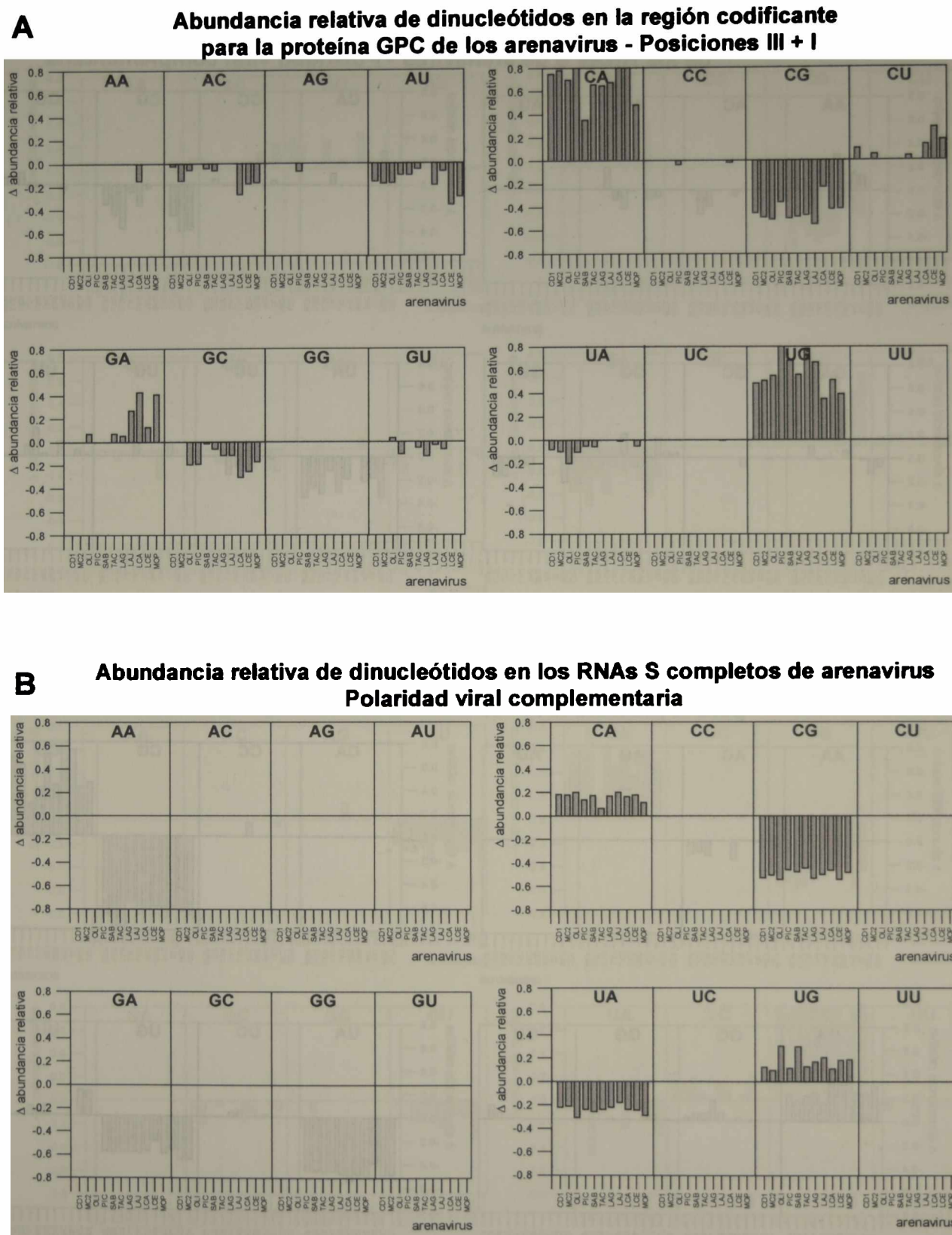


Figura 12.11. Abundancia relativa de dinucleótidos en los RNAs S de arnavirus. A. RNAs S analizados en el marco de lectura codificante para GPC, en las posiciones III + I. B. RNAs S completos analizados en polaridad viral complementaria. El análisis se realizó con las rutinas diseñadas y escritas por Ghiringhelli, P.D. (2002), esta tesis. Posteriormente, para facilitar la discriminación rápida entre sub- y sobre-representaciones y la visualización gráfica, los datos originales se modificaron, llevando a 0 todos los valores comprendidos entre 0,78 y 1,23, y asignando valor positivo a la diferencia: $\text{valor}_n (>1,23) - 1,23$ y valor negativo a la diferencia: $0,78 - \text{valor}_n (<0,78)$. CDI: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia.

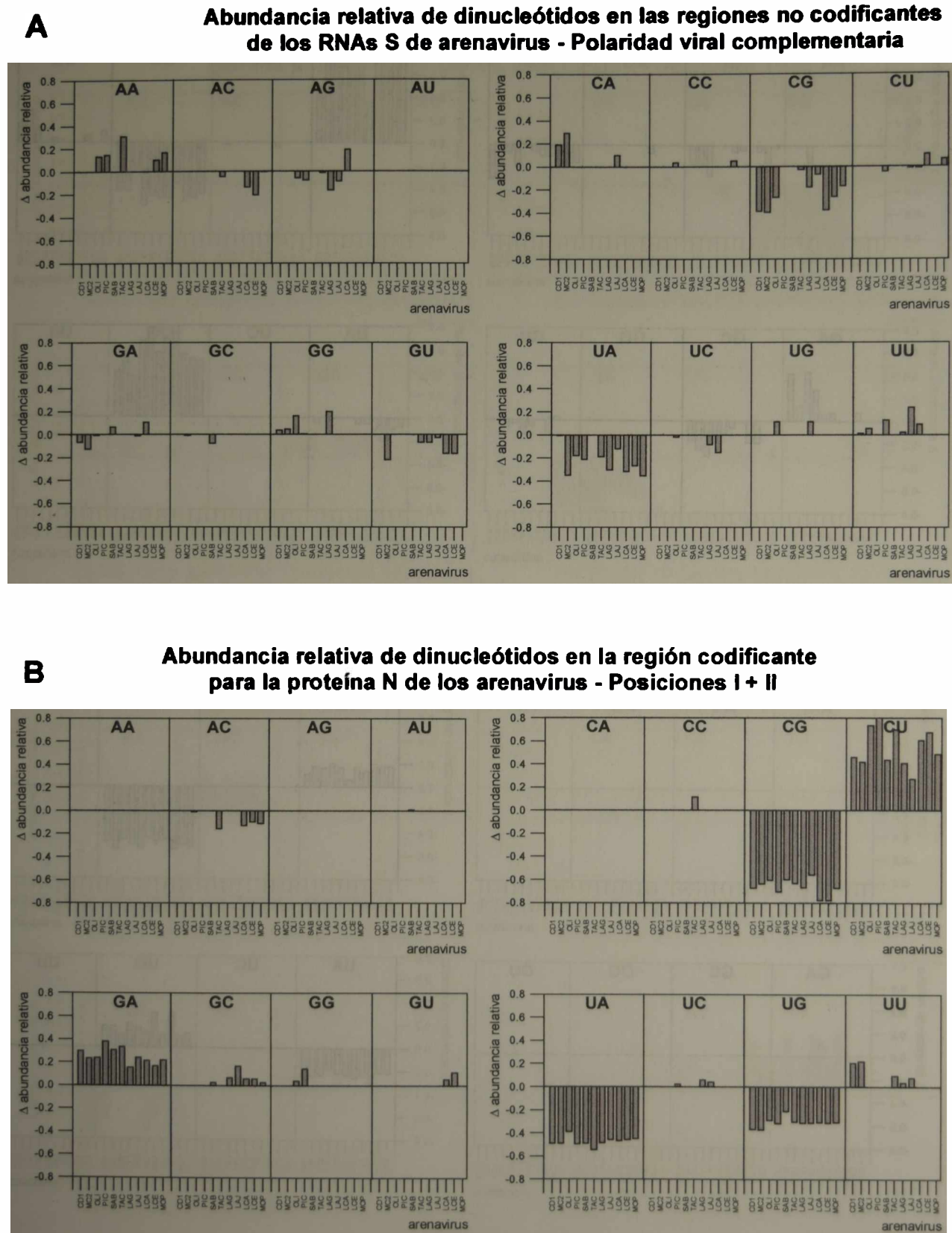


Figura 12.12. Abundancia relativa de dinucleótidos en los RNAs S de arnavirus. A. Regiones no codificantes analizadas en polaridad viral complementaria. **B.** RNAs S analizados en el marco de lectura codificante para N, en las posiciones I + II. El análisis se realizó con las rutinas diseñadas y escritas por Ghiringhelli, P.D. (2002), esta tesis. Posteriormente, para facilitar la discriminación rápida entre sub- y sobre-representaciones y la visualización gráfica, los datos originales se modificaron, llevando a 0 todos los valores comprendidos entre 0,78 y 1,23, y asignando valor positivo a la diferencia: $\text{valor}_n (>1,23) - 1,23$ y valor negativo a la diferencia: $0,78 - \text{valor}_n (<0,78)$. CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia.

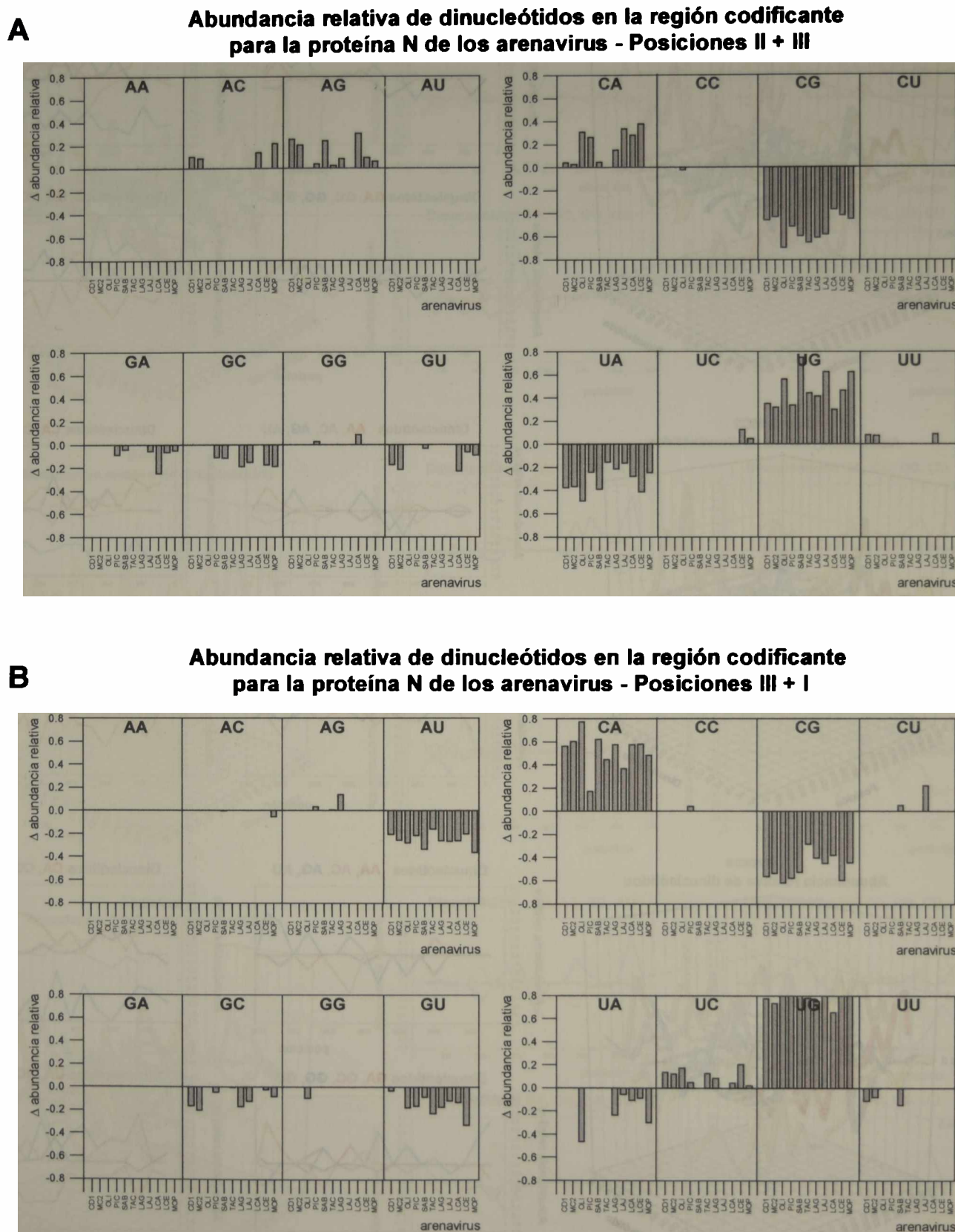


Figura 12.13. Abundancia relativa de dinucleótidos en los RNAs S de arnavirus. A. RNAs S analizados en el marco de lectura codificante para N, en las posiciones II + III. **B.** RNAs S analizados en el marco de lectura codificante para N, en las posiciones III + I. El análisis se realizó con las rutinas diseñadas y escritas por Ghiringhelli, P.D. (2002), esta tesis. Posteriormente, para facilitar la discriminación rápida entre sub- y sobre-representaciones y la visualización gráfica, los datos originales se modificaron, llevando a 0 todos los valores comprendidos entre 0,78 y 1,23, y asignando valor positivo a la diferencia: $\text{valor}_n (>1,23) - 1,23$ y valor negativo a la diferencia: $0,78 - \text{valor}_n (<0,78)$. CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia.

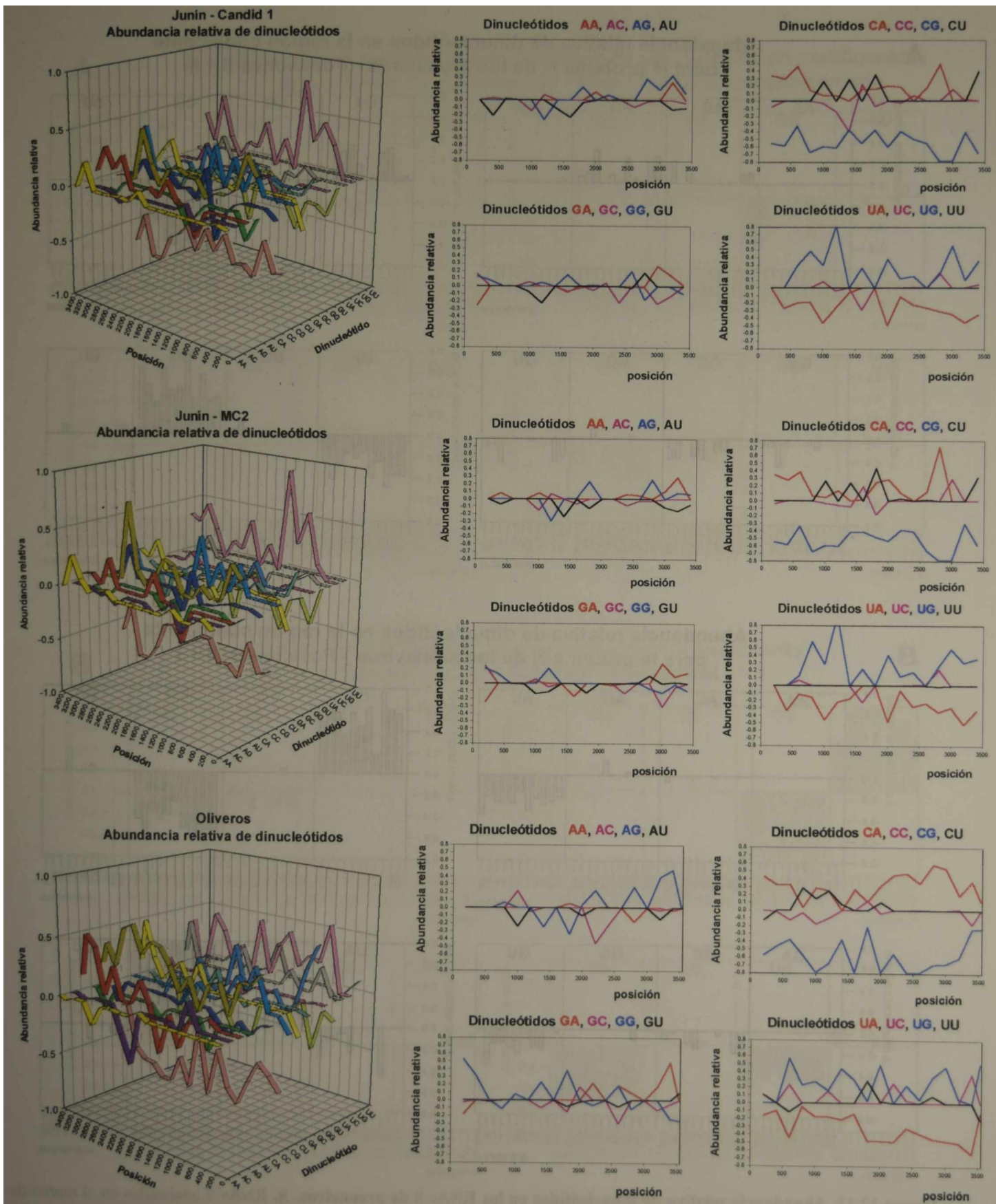


Figura 12.14.A. Perfiles de abundancia relativa de dinucleótidos en los RNAs S virales de los arnavirus Junin-Candid #1, Junin- MC2 y Oliveros. Los datos se calcularon empleando las rutinas diseñadas y escritas por Ghiringhelli, P.D. (2002), esta tesis y utilizando ventanas adyacentes de 200 nucleótidos. Posteriormente, para facilitar la discriminación rápida entre sub- y sobre-representaciones y la visualización gráfica, los datos originales se modificaron, llevando a 0 todos los valores comprendidos entre 0,78 y 1,23, y asignando valor positivo a la diferencia: $\text{valor}_n (>1,23) - 1,23$ y valor negativo a la diferencia: $0,78 - \text{valor}_n (<0,78)$. En los gráficos tridimensionales se muestran las abundancias relativas de los 16 dinucleótidos posibles. En los gráficos individuales en conjuntos de cuatro dinucleótidos.

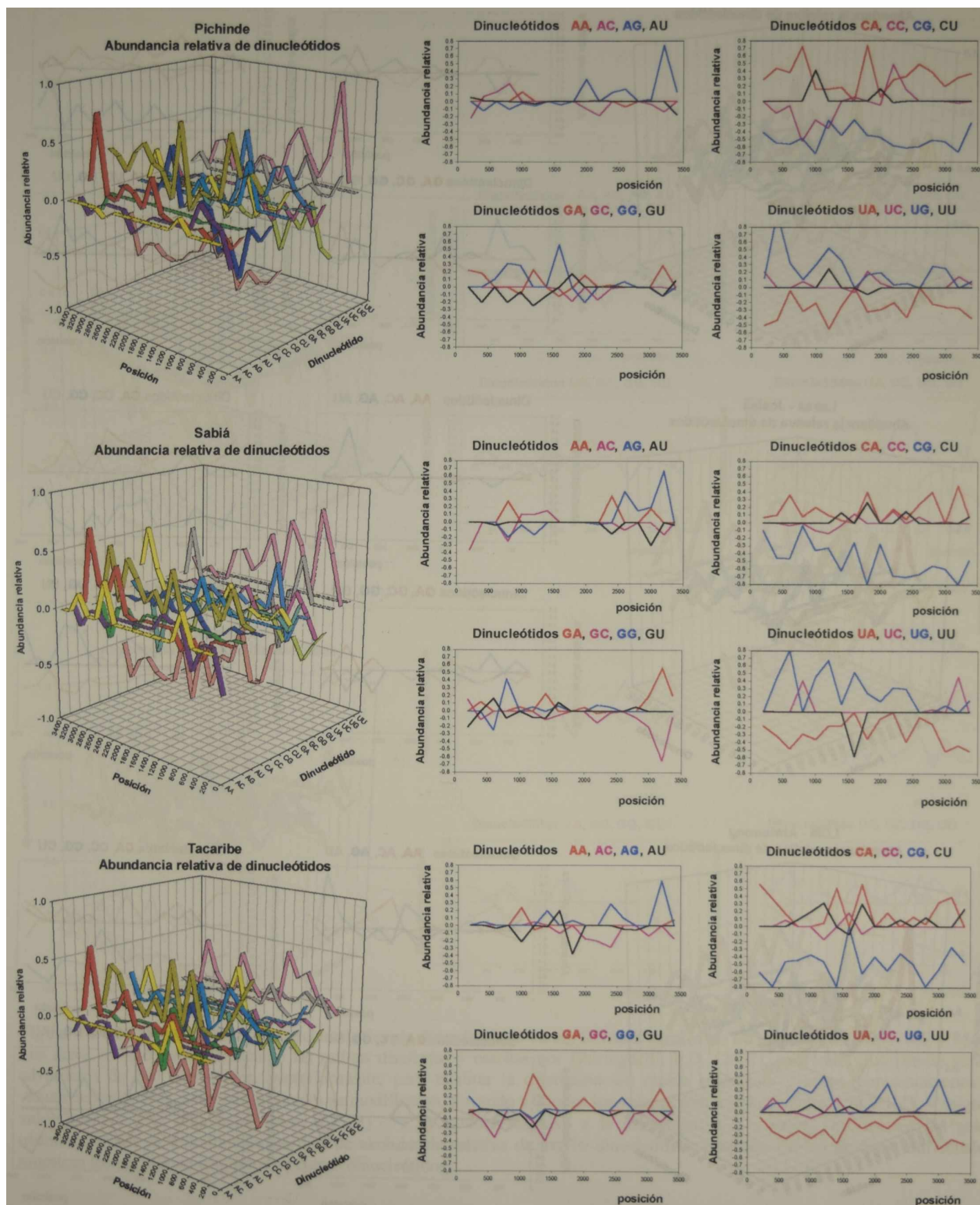


Figura 12.14.B. Perfiles de abundancia relativa de dinucleótidos en los RNAs S virales de los arnavirus Pichindé, Sabiá y Tacaribe. Los datos se calcularon empleando las rutinas diseñadas y escritas por Ghiringhelli, P.D. (2002), esta tesis y utilizando ventanas adyacentes de 200 nucleótidos. Posteriormente, para facilitar la discriminación rápida entre sub- y sobre-representaciones y la visualización gráfica, los datos originales se modificaron, llevando a 0 todos los valores comprendidos entre 0,78 y 1,23, y asignando valor positivo a la diferencia: $\text{valor}_n (>1,23) - 1,23$ y valor negativo a la diferencia: $0,78 - \text{valor}_n (<0,78)$. En los gráficos tridimensionales se muestran los perfiles de abundancia relativa de los 16 dinucleótidos posibles. En los gráficos individuales se muestran los perfiles en conjuntos de cuatro dinucleótidos.

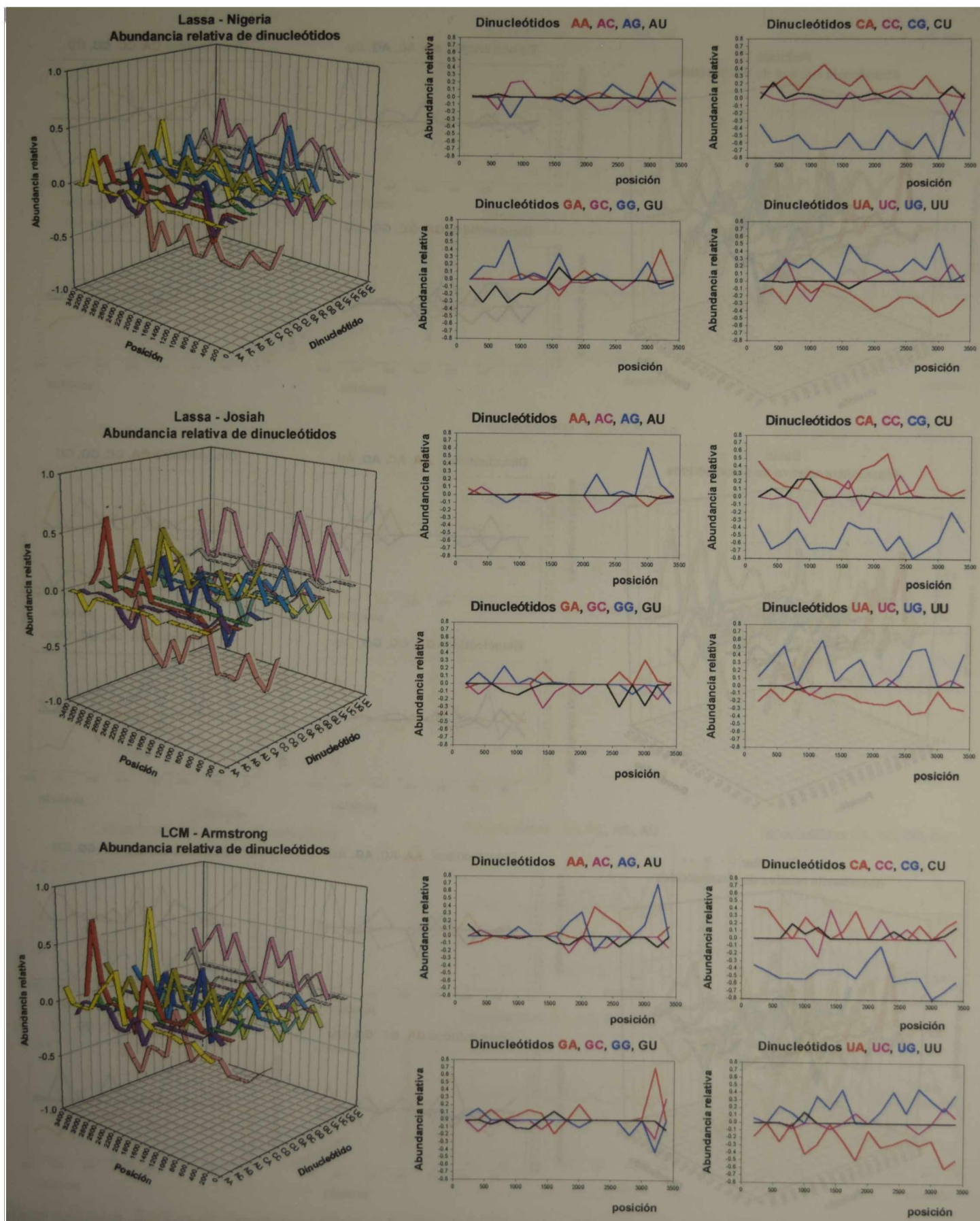


Figura 12.15.A. Perfiles de abundancia relativa de dinucleótidos en los RNAs S virales de los arnavirus Lassa-Nigeria, Lassa-Josiah y LCM-Armstrong. Los datos se calcularon empleando las rutinas diseñadas y escritas por Ghiringhelli, P.D. (2002), esta tesis y utilizando ventanas adyacentes de 200 nucleótidos. Posteriormente, para facilitar la discriminación rápida entre sub- y sobre-representaciones y la visualización gráfica, los datos originales se modificaron, llevando a 0 todos los valores comprendidos entre 0,78 y 1,23, y asignando valor positivo a la diferencia: $\text{valor}_n (>1,23) - 1,23$ y valor negativo a la diferencia: $0,78 - \text{valor}_n (<0,78)$. En los gráficos tridimensionales se muestran los perfiles de abundancia relativa de los 16 dinucleótidos posibles. En los gráficos individuales se muestran los perfiles en conjuntos de cuatro dinucleótidos.

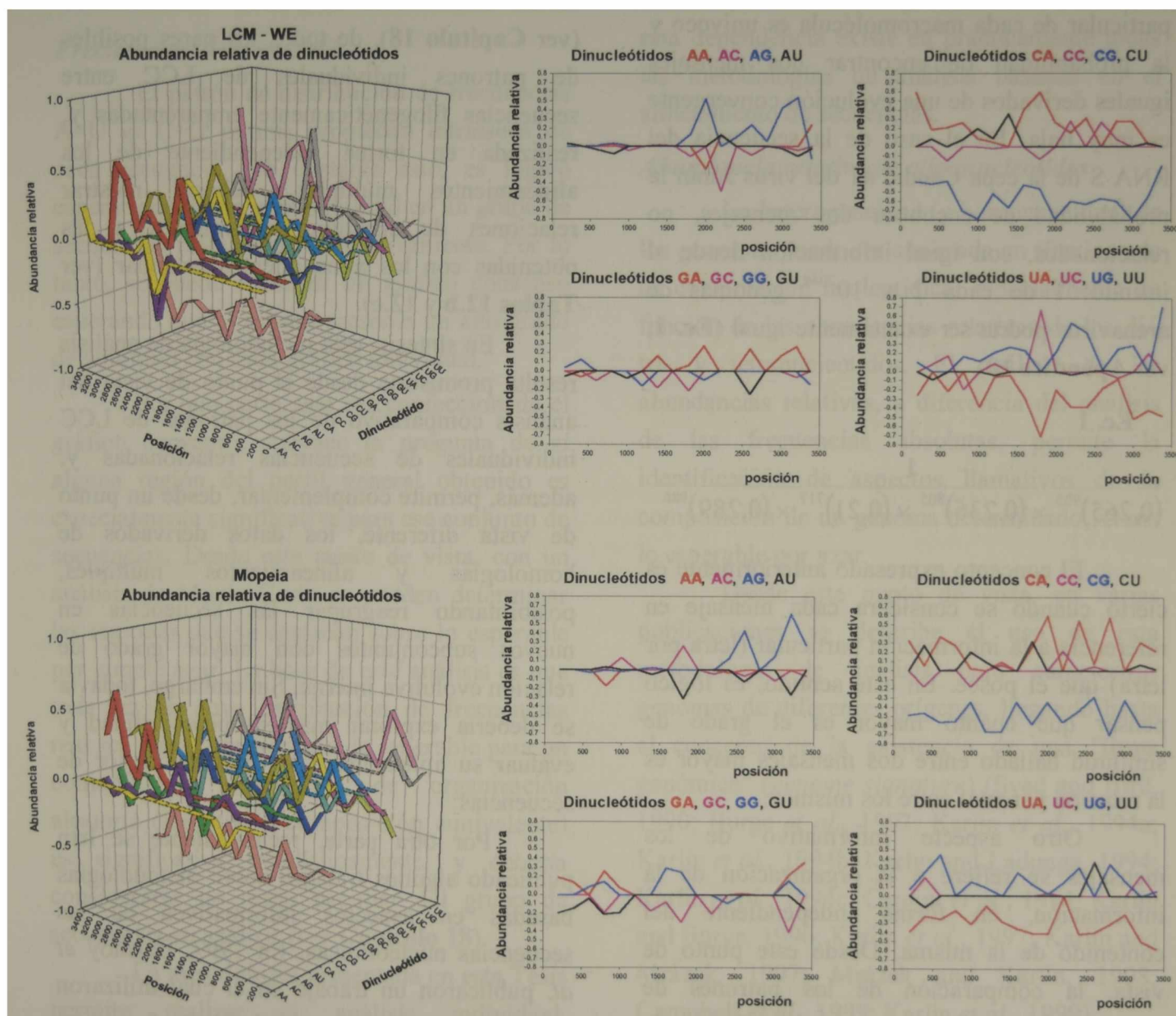


Figura 12.15.B. Perfiles de abundancia relativa de dinucleótidos en los RNAs S virales de los arnavirus LCM-WE y Mopeia. Los datos se calcularon empleando las rutinas diseñadas y escritas por Ghiringhelli, P.D. (2002), esta tesis y utilizando ventanas adyacentes de 200 nucleótidos. Posteriormente, para facilitar la discriminación rápida entre sub- y sobre-representaciones y la visualización gráfica, los datos originales se modificaron, llevando a 0 todos los valores comprendidos entre 0,78 y 1,23, y asignando valor positivo a la diferencia: $\text{valor}_n (>1,23) - 1,23$ y valor negativo a la diferencia: $0,78 - \text{valor}_n (<0,78)$. En los gráficos tridimensionales se muestran los perfiles de abundancia relativa de los 16 dinucleótidos posibles. En los gráficos individuales se muestran los perfiles en conjuntos de cuatro dinucleótidos.

Discusión sobre los aspectos informativos de los RNAs S de arnavirus

Complejidad local de composición (LCC)

Los ácidos nucleicos, son moléculas informativas por excelencia. El mensaje particular de cada macromolécula es unívoco y la probabilidad de encontrar dos mensajes iguales derivados de una evolución convergente es muy baja. En el caso de la secuencia del RNA S de la cepa Candid #1 del virus Junín la probabilidad de encontrar dos mensajes, no relacionados, con igual información tiende al infinito: 1 de cada 2×10^{2044} genomas de arnavirus podría ser exactamente igual (Ec. 1; ver Apéndice 10).

Ec. 1.

$$\frac{1}{(0,265)^{905} \times (0,236)^{805} \times (0,21)^{717} \times (0,289)^{986}}$$

El concepto expresado anteriormente es cierto cuando se considera cada mensaje en referencia a la información particular (letra por letra) que él posee. En este sentido, es lógico pensar que cuanto mayor es el grado de similitud hallado entre dos mensajes mayor es la relación parental entre los mismos.

Otro aspecto informativo de los mensajes se refiere a la organización de la información, en forma independiente del contenido de la misma. Desde este punto de vista, la comparación de los patrones de complejidad local de composición (LCC) de macromoléculas relacionadas evolutivamente muestra un cierto grado de similitud organizativa. Dicha similitud se puede correlacionar con el grado de relación evolutiva de las macromoléculas en cuestión.

Respecto de este último punto, y de una manera cualitativa, un gráfico consenso generado a partir de un conjunto de secuencias relacionadas evolutivamente, presentaría un alto grado de similitud con cada uno de los gráficos individuales. Mientras que un gráfico consenso generado a partir de un conjunto de secuencias

con escasa o nula relación evolutiva, presentaría grandes diferencias con cada uno de los gráficos individuales.

Si las consideraciones anteriores son correctas, una comparación cuantitativa, basada en las intensidades de Fourier características (ver Capítulo 18), de todos los pares posibles de patrones individuales de LCC entre secuencias filogenéticamente emparentadas y, realizada en forma independiente de los alineamientos múltiples, debería mostrar relaciones de similitud equivalentes a las obtenidas con los alineamientos estándar (ver Tablas 12.b y 12.c).

En síntesis, esta metodología de análisis resulta promisoría desde el punto de vista del análisis comparativo de los patrones de LCC individuales de secuencias relacionadas y, además, permite complementar, desde un punto de vista diferente, los datos derivados de homologías y alineamientos múltiples, posibilitando reagrupar las secuencias en nuevos subconjuntos con mayor grado de relación evolutiva interna. Sin embargo, todavía se debería explorar más su potencialidad y evaluar su aplicación a un número mayor de secuencias.

Por otra parte, recientemente se han publicado algunos trabajos sobre metodologías basadas en aspectos informativos de las secuencias nucleotídicas. En 1997, Bolshoy *et al.* publicaron un trabajo en el cual utilizaron métodos basados en la complejidad informativa para detectar la presencia de patrones nucleosomales en regiones de baja complejidad. Los distintos patrones obtenidos fueron caracterizados empleando la transformada de Fourier. Además, en 1999, Oliver *et al.* publicaron un trabajo en el cual utilizaron una medida de la complejidad de composición de secuencias, a la cual llamaron: *Sequence Composition Complexity (SCC)*, para detectar dominios composicionales, estadísticamente significativos, en moléculas de DNA.

Sin embargo, se debe destacar que

ninguno de estos trabajos está orientado hacia el análisis comparativo de una familia de secuencias. La principal ventaja de la metodología presentada en esta Tesis es, justamente, su orientación al análisis comparativo.

Frecuencias A+U vs C+G

El patrón de distribución de frecuencias A+U vs C+G es una propiedad intrínseca de cada secuencia. No obstante esto, es lógico esperar que exista gran similitud en un grupo de secuencias relacionadas evolutivamente. Por lo tanto, se puede construir un gráfico consenso esperando que el mismo presente un alto grado de similitud con cada gráfico individual.

Por otra parte, una vez confeccionado el gráfico consenso, subyace la pregunta de si alguna región del perfil general obtenido es especialmente significativa para ese conjunto de secuencias. Desde este punto de vista, con un análisis de Monte Carlo se pueden determinar las regiones que se destacan sobre lo esperable por simple azar. La base de este análisis es que toda región cuya distribución de frecuencias real sea mayor (menor) que la esperable para un conjunto de secuencias de organización aleatoria (longitud y composición equivalente) es significativa estadísticamente, y debería considerarse un rasgo peculiar del grupo de secuencias en análisis (ver **Capítulo 18**).

La metodología presentada en esta Tesis permite realizar el análisis individual, confeccionar el gráfico consenso en base a un alineamiento múltiple de las secuencias y determinar las regiones características estadísticamente significativas. De esta forma, en los RNAs S de los arnavirus, se han encontrado 3 regiones características: la región 1 (~ 250 nt.) y la región 2 (~ 80 nt.) corresponden a la región central del marco de lectura que codifica para la glicoproteína G2 y muestran una predominancia de A+U, mientras que la región 3 (~ 200 nt.) muestra una marcada predominancia de C+G y corresponde a la zona

intergénica y regiones adyacentes. (ver **Figura 12.8.A** y **Apéndice 5**).

En este punto, es necesario destacar que el patrón consenso generado es altamente dependiente del alineamiento múltiple. No obstante, también es importante considerar que esta dependencia existe en prácticamente todas las metodologías de análisis basadas en el alineamiento de secuencias.

Abundancia relativa de oligonucleótidos

La heterogeneidad (u homogeneidad) de los genomas puede ser analizada en términos de su composición nucleotídica, especialmente en función de las abundancias relativas de los di-, tri- y tetranucleótidos. El análisis de las abundancias relativas, a diferencia del análisis de las frecuencias absolutas, permite la identificación de aspectos llamativos de la composición de un genoma determinado *versus* lo esperable por azar.

Desde este punto de vista, en varias publicaciones se describe el uso de esta metodología de análisis para caracterizar genomas de diferentes orígenes, llegando hasta el concepto de la existencia de una "firma genómica" (*genome signature*) (Sved and Bird, 1990; Burge *et al.*, 1992; Karlin *et al.*, 1994a; Karlin *et al.*, 1994b; Karlin and Ladunga, 1994; Karlin *et al.*, 1994c; Cardon *et al.*, 1994; Karlin and Burge, 1996; Karlin *et al.*, 1997; Karlin and Mrázek, 1997; Mrázek and Karlin, 1998; Campbell *et al.*, 1999; Karlin *et al.*, 1999).

En el caso de los arnavirus, como se puede observar en la **Tabla 12.d**, los dinucleótidos CG y UA son sub-abundantes, mientras que los dinucleótidos UG y CA son sobre-abundantes. Por otra parte, las sub- o sobre-abundancias de los tri- y tetranucleótidos están directamente relacionadas con la sub-abundancia del dinucleótido CG, ya que éste es un componente de todos aquellos que se han detectado con valores extremos.

Cuando se realiza un análisis por regiones, resulta llamativo observar que el

dinucleótido CG es especialmente sub-abundante en cualquiera de las tres posiciones en las regiones codificantes, mientras que los dinucleótidos UG y CA son especialmente sobre-abundantes en la posición III+I de los marcos de lectura (ver Figuras 12.9, 12.10, 12.11, 12.12 y 12.13).

Además, también resulta llamativo observar que, en la mayor parte de los arnavirus, existe una cierta correlación entre la sub-abundancia del dinucleótido CG y la sobre-abundancia del dinucleótido UG cuando se realiza un análisis por ventanas (ver Figuras 12.14.A, 12.14.B, 12.15.A y 12.15.B); en la mayor parte de las ventanas los picos de sub-abundancia del dinucleótido CG son acompañados por picos de sobre-abundancia del dinucleótido UG.

Por otra parte, la información concerniente a las sub- y sobre-abundancias de dinucleótidos es coincidente con lo encontrado en otros virus con genoma de RNA, tales como los picornavirus, los flavivirus, los coronavirus, los retrovirus, los calicivirus, los rhabdovirus, los paramyxovirus y los orthomyxovirus (Karlin *et al.*, 1994b).

La constancia en la sub-abundancia del dinucleótido CG, genera varios interrogantes:

1) ¿El contenido y/o patrón de distribución del dinucleótido CG puede generar una estructura definida que afecte, en particular, la interacción con proteínas?. En función de esto, ¿Existen restricciones estructurales y/o regulatorias intrínsecas al dinucleótido CG?

2) ¿Existen mecanismos de mutación preferencial, mediante deaminación u otras fuerzas selectivas que afecten especialmente al dinucleótido CG?

3) ¿El contenido del dinucleótido CG afecta de alguna manera la frecuencia de uso de aminoácidos tales como la arginina?, o ¿Sólo afecta la frecuencia de uso de codones?

Respecto de los cuestionamientos planteados en 1), no se han encontrado evidencias que justifiquen una hipótesis

asociada con cuestiones estructurales y/o regulatorias. La distribución observada en el análisis por ventanas (ver Figuras 12.14.A, 12.14.B, 12.15.A y 12.15.B) es sumamente variable entre los distintos arnavirus.

Respecto del cuestionamiento 2), un análisis de correlación entre la sub-abundancia del dinucleótido CG y las sobre-abundancias de los dinucleótidos UG y CA (Figura 12.16) muestra una cierta coincidencia entre las mismas. Esto sugiere que podría existir una deaminación preferencial de las Cs (Lewin, 1997) que forman parte del dinucleótido CG durante el ciclo replicativo, lo cual llevaría, en forma directa, a un incremento en el dinucleótido UG y, en forma indirecta, a un incremento posterior en el dinucleótido CA. Por otra parte, el mismo tipo de análisis realizado entre la sub-abundancia del dinucleótido UA y las sobre-abundancias de los dinucleótidos UG y CA (Figura 12.17) sugiere que la sub-abundancia del dinucleótido UA está asociada con mecanismos diferentes e independientes.

Respecto de los cuestionamientos planteados en 3), la frecuencia promedio de arginina en las proteínas GPC y N de los arnavirus es 4,96%, lo cual coincide con lo esperable por simple azar. En la Tabla 12.e se puede observar que el único efecto de la sub-abundancia del dinucleótido CG está asociado con la frecuencia de uso de codones.

Tabla 12.e. Frecuencias de uso de los codones de arginina en las proteínas GPC y N de los arnavirus

Virus	Proteína GPC		Proteína N	
	CGN	AGR	CGN	AGR
CDI	3	15	3	32
LAG	1	21	3	26
LAJ	2	15	6	24
LCA	4	17	0	28
LCE	3	18	0	30
MC2	3	17	4	32
MOP	1	21	3	30
OLI	6	14	5	26
PIC	2	21	5	28
SAB	6	20	2	27
TAC	6	21	4	26

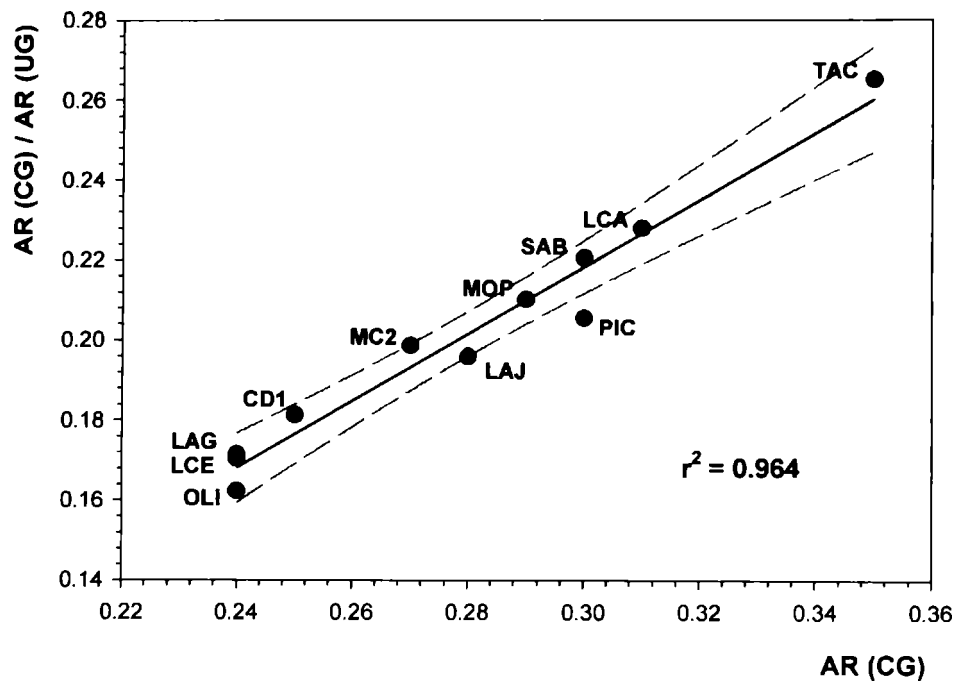


Figura 12.16. Correlación entre las abundancias relativas de los dinucleótidos CG y UG. En el eje Y se representa el cociente de las abundancias relativas (AR) de los dinucleótidos CG y UG, mientras que en el eje X se representa la abundancia relativa del dinucleótido CG (ver **Tabla 12.d**). La línea negra representa la correlación lineal, calculada por el método de mínimos cuadrados. Las líneas rojas representan el intervalo de confianza del 99%. Se indica el coeficiente de correlación r^2 . Un gráfico idéntico se obtiene cuando se calcula la correlación entre los dinucleótidos CG y CA. CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia.

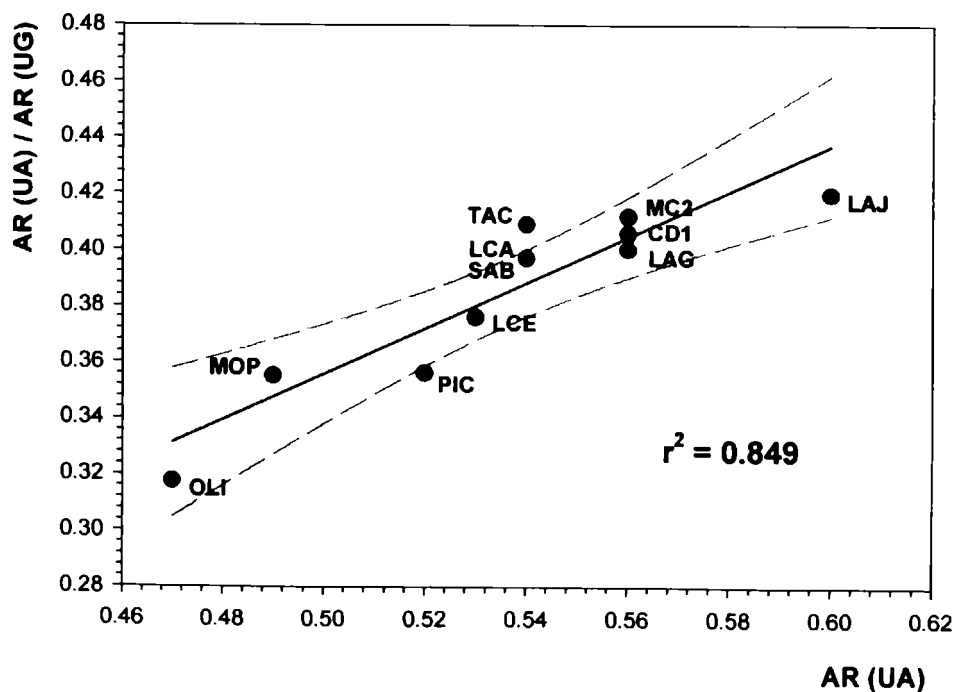


Figura 12.17. Correlación entre las abundancias relativas de los dinucleótidos UA y UG. En el eje Y se representa el cociente de las abundancias relativas (AR) de los dinucleótidos UA y UG, mientras que en el eje X se representa la abundancia relativa del dinucleótido UA (ver **Tabla 12.d**). La línea negra representa la correlación lineal, calculada por el método de mínimos cuadrados. Las líneas rojas representan el intervalo de confianza del 99%. Se indica el coeficiente de correlación r^2 . Un gráfico idéntico se obtiene cuando se calcula la correlación entre los dinucleótidos UA y CA. CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia.

Predicción de la estructura secundaria del RNA

Las moléculas de RNA pueden formar tanto estructuras secundarias como terciarias. Dado que las estructuras terciarias presentan interacciones mucho más débiles que las secundarias (Riesner and Romer, 1973; Crothers *et al.*, 1978), la mayor parte de los trabajos y algoritmos desarrollados se han concentrado en estas últimas.

La determinación de las estructuras secundarias puede ser predicha a partir de las secuencias individuales de RNA, siguiendo reglas bastante simples y un esquema de minimización de energías, o mediante análisis de tipo filogenético, utilizando un alineamiento como base, con el posterior cálculo de plegamiento de las moléculas individuales en estructuras equivalentes (Jaeger *et al.*, 1990).

El método basado en minimización de energías requiere, esencialmente, de tablas de parámetros termodinámicos con los valores de energía libre de los distintos motivos estructurales posibles (ver **Apéndice 7**) y capacidad (y tiempo) de computación para determinar un conjunto de estructuras de mínima energía. La principal ventaja de este método es que permite encontrar los códigos estructurales mínimos propios de una secuencia y analizarlos comparativamente con los de un conjunto de secuencias relacionadas.

De acuerdo a Jaeger *et al.* (1990) y Michael Zuker (1994), el conocimiento actual es adecuado para predecir correctamente alrededor del 70% de la estructura secundaria de una molécula de RNA.

Estructuras secundarias predichas para los RNAs S de arnavirus

Estructuras secundarias completas

En función de lo mencionado anteriormente, se trabajó sobre las secuencias individuales de los RNAs S de 11 arnavirus, realizando los cálculos y predicciones en el servidor de Michel Zuker (<http://bioinfo.math.rpi.edu/~zukerm/>).

Desde un punto de vista general, y a grandes rasgos, dividiendo arbitrariamente en tres conjuntos de motivos estructurales básicos: interacciones locales (rango ± 200 nucleótidos), interacciones a distancia (rango $>$ de 200 nucleótidos y ausencia de interacción (regiones simple cadena), en la **Tabla 12.f** se puede observar que todos los arnavirus exhiben una gran similitud en cuanto a los porcentajes de las moléculas involucrados en cada conjunto de motivos estructurales. En promedio, el 64,1% de las secuencias nucleotídicas de los RNAs S de los arnavirus están involucradas en estructuras de doble cadena.

En las **Figuras 12.18** y **12.19** se muestran las estructuras secundarias predichas para los RNAs S de los virus Junín-MC2 y Junín-Candid #1, respectivamente. Esta visión global, por sí misma interesante, no permite determinar, sin embargo, la existencia de motivos estructurales comunes a todos los arnavirus.

En función de la disposición circular de la nucleocápsides en las observaciones de microscopía electrónica y de lo descrito para otros sistemas virales, tales como influenza (Hsu *et al.*, 1987), se analizaron con mayor

Tabla 12.f. Estructura secundaria de los RNAs S de arnavirus

	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP	PROM
Interacciones en un rango +/- 200 nt	44,1	45,3	40,4	49,2	43,2	49,3	44,1	44,0	42,5	50,8	47,5	45,5
Interacciones en un rango > 200 nt	19,2	18,2	26,0	14,4	21,9	17,5	18,6	19,9	22,8	13,3	15,4	18,6
Ausencia de interacciones	36,7	36,5	35,6	36,4	34,9	33,2	37,3	36,1	34,7	36,9	37,1	35,9

CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia

detalle las posibles estructuras tipo *panhandle* formadas por los extremos 5' y 3' de los RNAs virales.

Por otra parte, los mismos criterios se aplicaron al estudio detallado de las regiones intergénicas, ya que todas las publicaciones sobre secuencias de RNAs S de arnavirus han predicho la existencia local de estructuras tipo *hairpin loop* (Auperin *et al.*, 1984; Clegg and Oram, 1985; Romanowski *et al.*, 1985; Auperin *et al.*, 1986; Franze Fernández *et al.*, 1987; Auperin and McCormick, 1988; Ghiringhelli *et al.*, 1991; Gonzalez *et al.*, 1996; Bowen *et al.*, 1996; Ghiringhelli *et al.*, 1997; Albariño *et al.*, 1997).

Estructuras secundarias en los extremos y en las regiones intergénicas de los RNAs S

La determinación de motivos estructurales en las regiones mencionadas se realizó mediante el análisis comparativo de las estructuras 5-óptimas (*5-optimal*, Zuker, 1989; Jaeger *et al.*, 1990; Zuker *et al.*, 1991) para cada RNA S individual (Tablas 12.g y 12.h) (ver **Materiales y Métodos** y **Capítulo 18**).

Los motivos estructurales de los extremos de los RNAs se determinaron en base al cálculo de plegamiento de moléculas quiméricas. Las mismas se construyeron con 400-600 nucleótidos de cada extremo viral, separados por una secuencia nucleotídica homopolimérica de ~200 residuos.

Los motivos estructurales de las regiones intergénicas se determinaron en base al cálculo de plegamiento de moléculas de RNA S parciales. En general, se utilizaron moléculas de ~2800-3000 nucleótidos de longitud, a partir del nucleótido 241 del RNA viral.

En síntesis, mediante esta metodología se ha determinado la existencia de motivos de estructura secundaria comunes en regiones de los RNAs S importantes desde el punto de vista conformacional y regulatorio. Estos motivos podrían formar parte de un código estructural general para los arnavirus.

En las **Figuras 12.20, 12.21 y 12.22** se pueden observar distintos motivos estructurales característicos de los RNAs S de las dos cepas del virus Junín.

En las **Figuras 12.23, 12.24 y 12.25** se muestran comparativamente los distintos motivos estructurales en todos los arnavirus.

Tabla 12.g. Estructuras secundarias 5-óptimas de los extremos 5' y 3' virales

Virus	Estructuras 5-óptimas analizadas	Estructura seleccionada	
		número	R _S *
CD1	24	3	1,16
LAG	29	1	1,00
LAJ	24	1	1,00
LCA	21	1	1,00
LCE	31	1	1,00
MC2	9	1	1,00
MOP	29	1	1,00
OLI	21	1	1,00
PIC	27	1	1,00
SAB	28	4	1,09
TAC	23	1	1,00

* R_S: relación entre el *score* promedio de la estructura seleccionada y el *score* promedio de la estructura predicha como óptima en base al ΔG₀.

Tabla 12.h. Estructuras secundarias 5-óptimas de las regiones intergénicas virales

Virus	Estructuras 5-óptimas analizadas	Estructura seleccionada	
		número	R _S *
CD1	4	1	1,00
LAG	4	1	1,00
LAJ	4	1	1,00
LCA	4	1	1,00
LCE	4	1	1,00
MC2	4	1	1,00
MOP	4	1	1,00
OLI	4	1	1,00
PIC	4	1	1,00
SAB	4	1	1,00
TAC	4	1	1,00

* R_S: relación entre el *score* promedio de la estructura seleccionada y el *score* promedio de la estructura predicha como óptima en base al ΔG₀.

RNA S viral del virus Junín MC2 -
Estructura secundaria

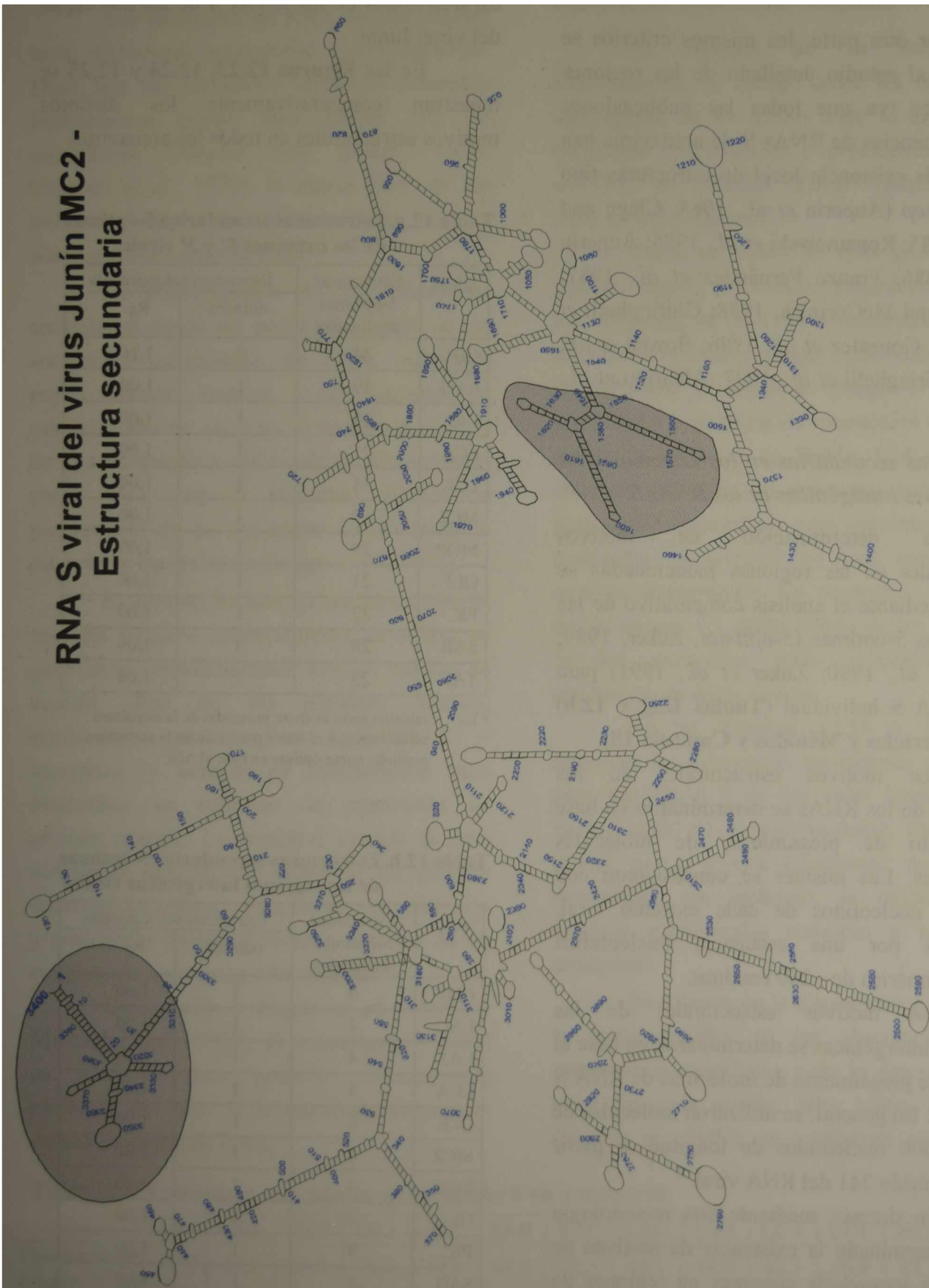


Figura 12.18. Estructura secundaria del RNA S viral del virus Junín-MC2. La estructura fue calculada de acuerdo a lo descrito en *Materiales y Métodos* y en el *Capítulo 18*. Con campos sombreados en gris claro se destacan: la estructura tipo *panhandle* formada por el apareamiento de los nucleótidos correspondientes a los extremos 5' y 3', y las estructuras tipo *hairpin loop* intergénicas. La estructura mostrada para los nucleótidos 1 a 600 y 2801 a 3400 corresponde a la estructura consenso derivada del análisis comparativo de 9 estructuras (incluyendo la óptima y las subóptimas). La estructura mostrada para la región 201 a 3200 corresponde a la estructura consenso derivada del análisis comparativo de 4 estructuras (incluyendo la óptima y las subóptimas). En las regiones de superposición, correspondientes a los nucleótidos 201 a 600 y 2801 a 3200, se analizó la conservación de la identidad de interacciones nucleotídicas.

RNA S viral del virus Junín Candid 1 - Estructura secundaria

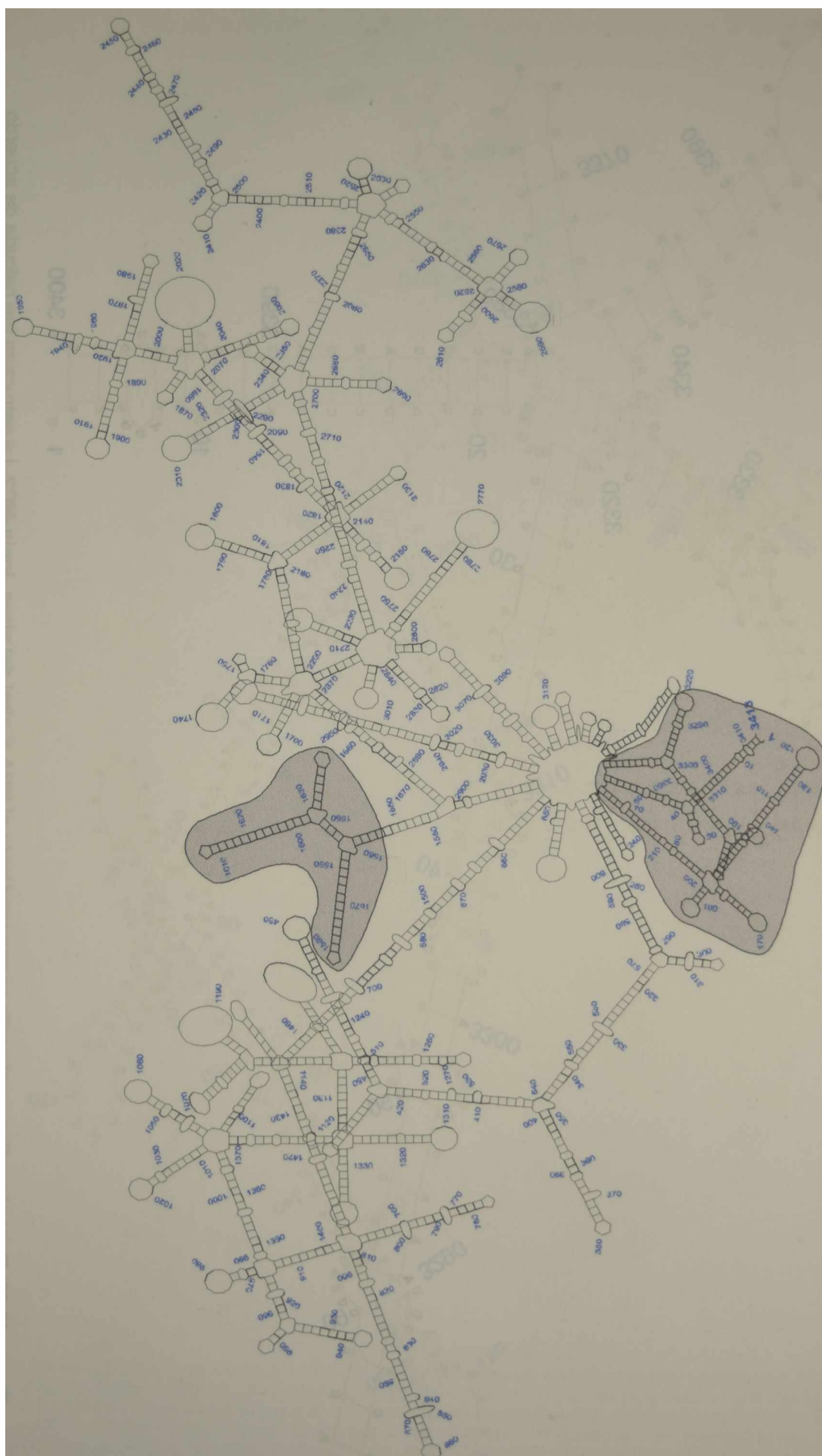


Figura 12.19. Estructura secundaria del RNA S viral del virus Junín-Candid #1. La estructura fue calculada de acuerdo a lo descrito en **Materiales y Métodos** y en el **Capítulo 18**. Con campos sombreados en gris claro se destacan: la estructura tipo *panhandle* formada por el apareamiento de los nucleótidos correspondientes a los extremos 5' y 3', y las estructuras tipo *hairpin loop* intergénicas. La estructura mostrada para los nucleótidos 1 a 600 y 2814 a 3413 corresponde a la estructura consenso derivada del análisis comparativo de 24 estructuras (incluyendo la óptima y las subóptimas). La estructura mostrada para la región 241 a 3213 corresponde a la estructura consenso derivada del análisis comparativo de 4 estructuras (incluyendo la óptima y las subóptimas). En las regiones de superposición, correspondientes a los nucleótidos 241 a 600 y 2814 a 3213, se analizó la conservación de la identidad de interacciones nucleotídicas.

RNA S viral del virus Junín MC2 - Estructura tipo panhandle

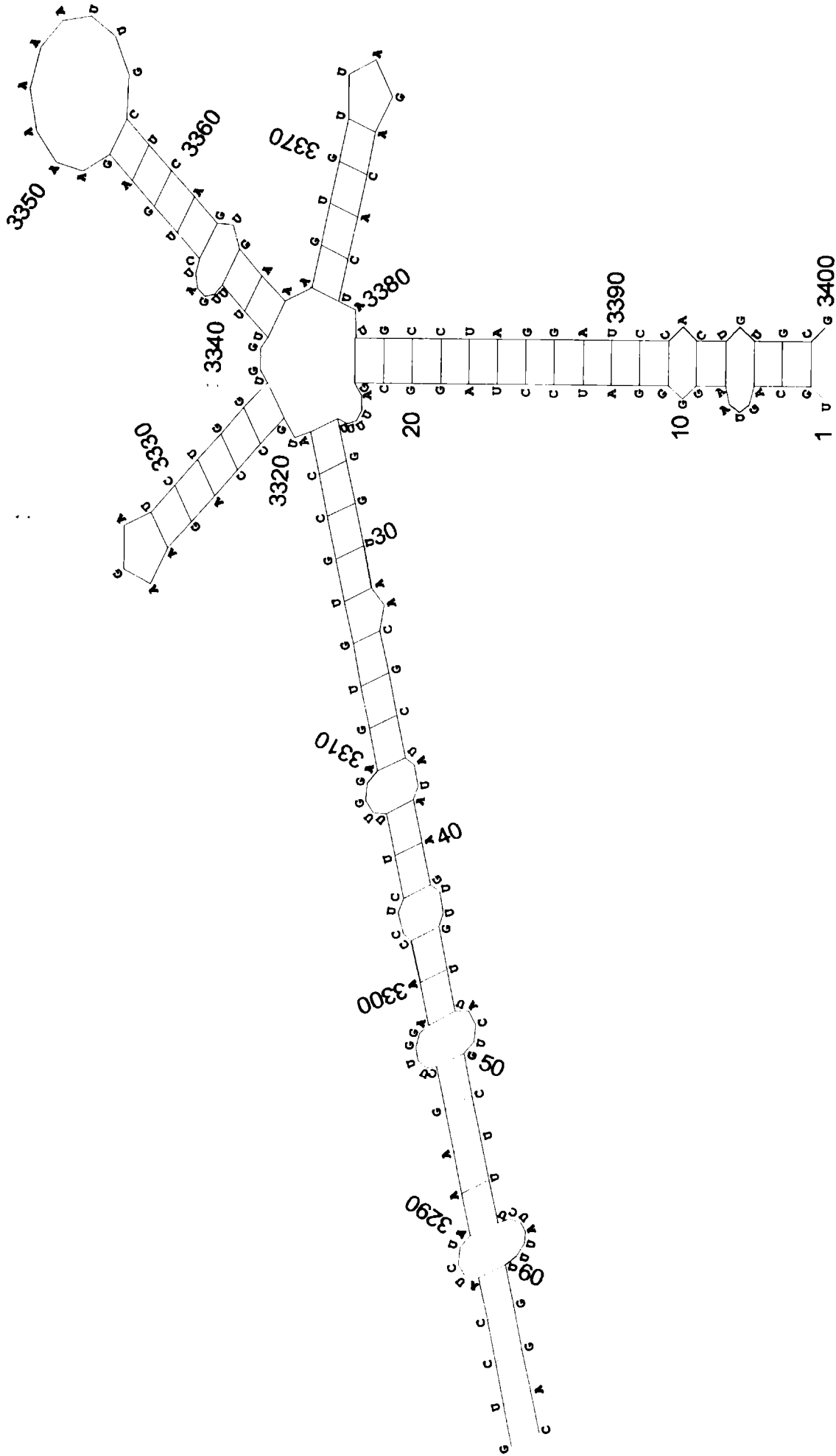


Figura 12.20. Estructura tipo *panhandle* correspondiente al RNA S viral del virus Junín-MC2. La estructura fue calculada de acuerdo a lo descrito en *Materiales y Métodos* y en el Capítulo 18. La estructura mostrada es una ampliación de la región correspondiente de la estructura completa.

RNA S viral del virus Junín Candid 1 - Estructura tipo panhandle

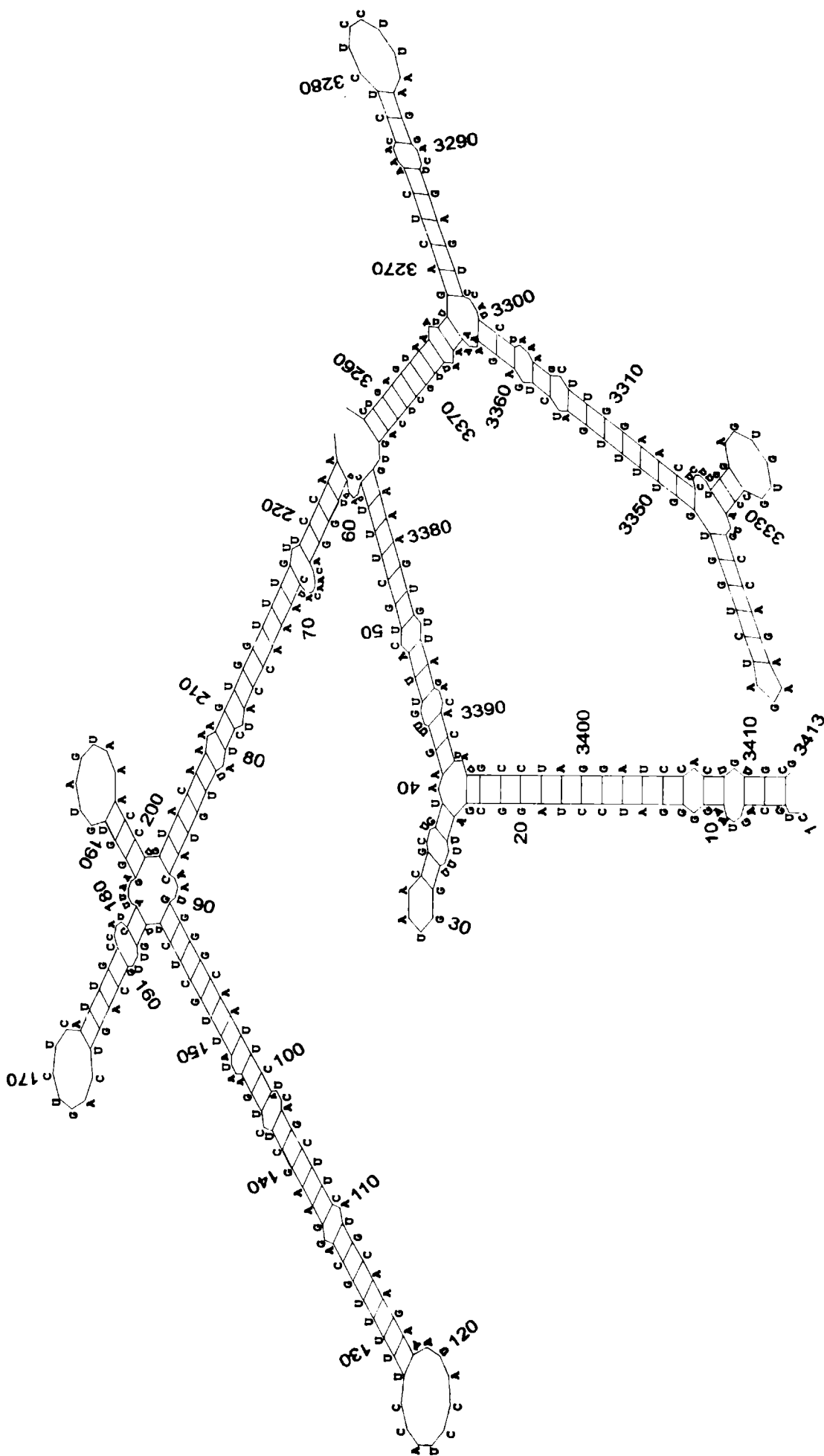


Figura 12.21. Estructura tipo panhandle correspondiente al RNA S viral del virus Junín-Candid #1. La estructura fue calculada de acuerdo a lo descrito en Materiales y Métodos y en el Capítulo 18. La estructura mostrada es una ampliación de la región correspondiente de la estructura completa.

**RNAs S virales de los virus Junin-MC2 y Junin-Candid #1
Estructuras tipo *hairpin loop* intergénicas**

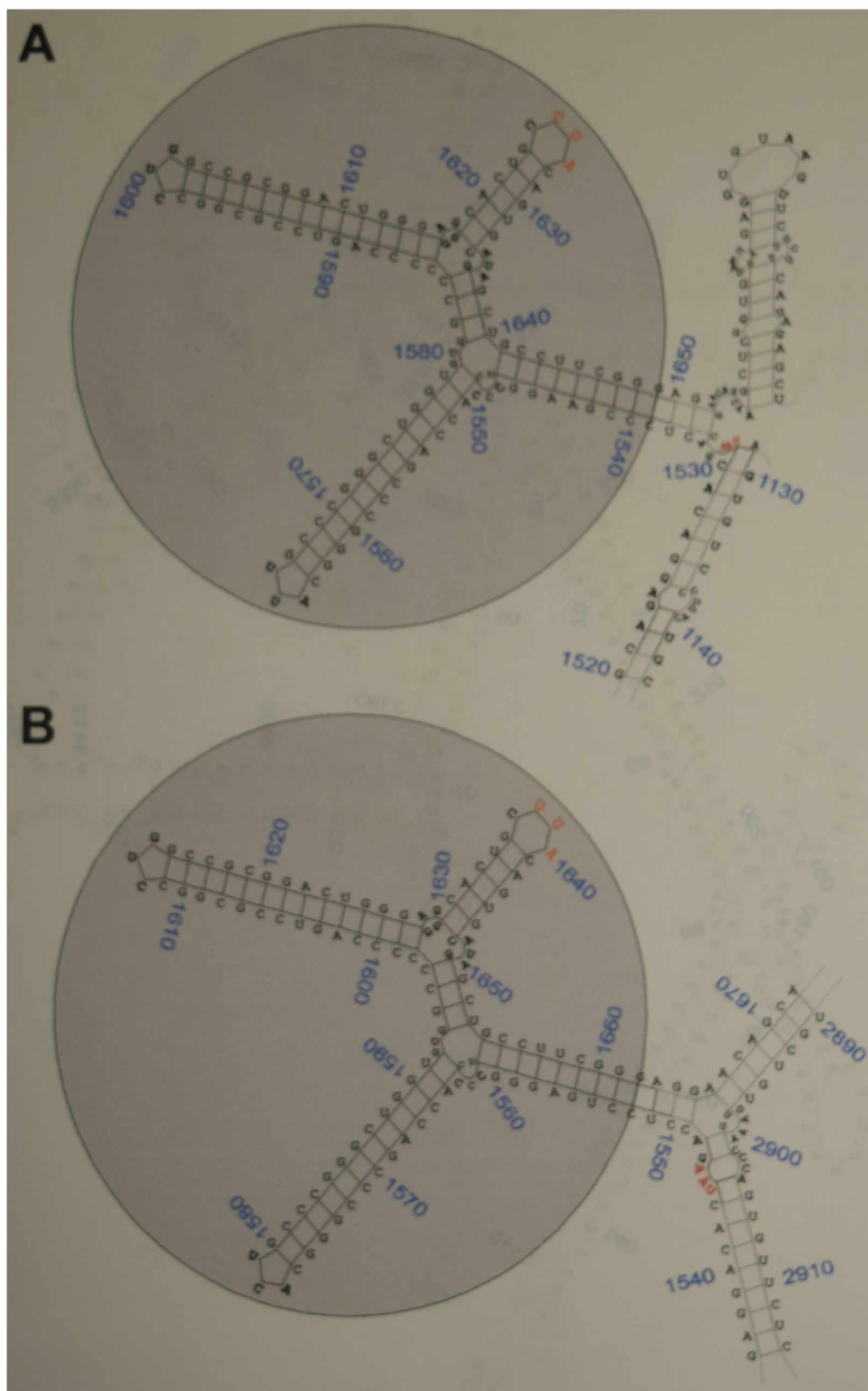


Figura 12.22. Estructuras tipo *hairpin loop* correspondientes a los RNAs S virales de los virus Junin-MC2 y Junin-Candid #1. A. Estructura de la región intergénica del virus Junin-MC2. B. Estructura de la región intergénica del virus Junin-Candid #1. Las estructuras fueron calculadas de acuerdo a lo descrito en *Materiales y Métodos* y en el *Capítulo 18*. Las estructuras mostradas son una ampliación de las regiones correspondientes de las estructuras completas respectivas. En rojo se indican los codones de terminación de los marcos de lectura codificantes para GPC y en naranja los tripletes complementarios a los codones de terminación de los marcos de lectura codificantes para N.

RNAS S virales de arenavirus - Estructuras tipo panhandle

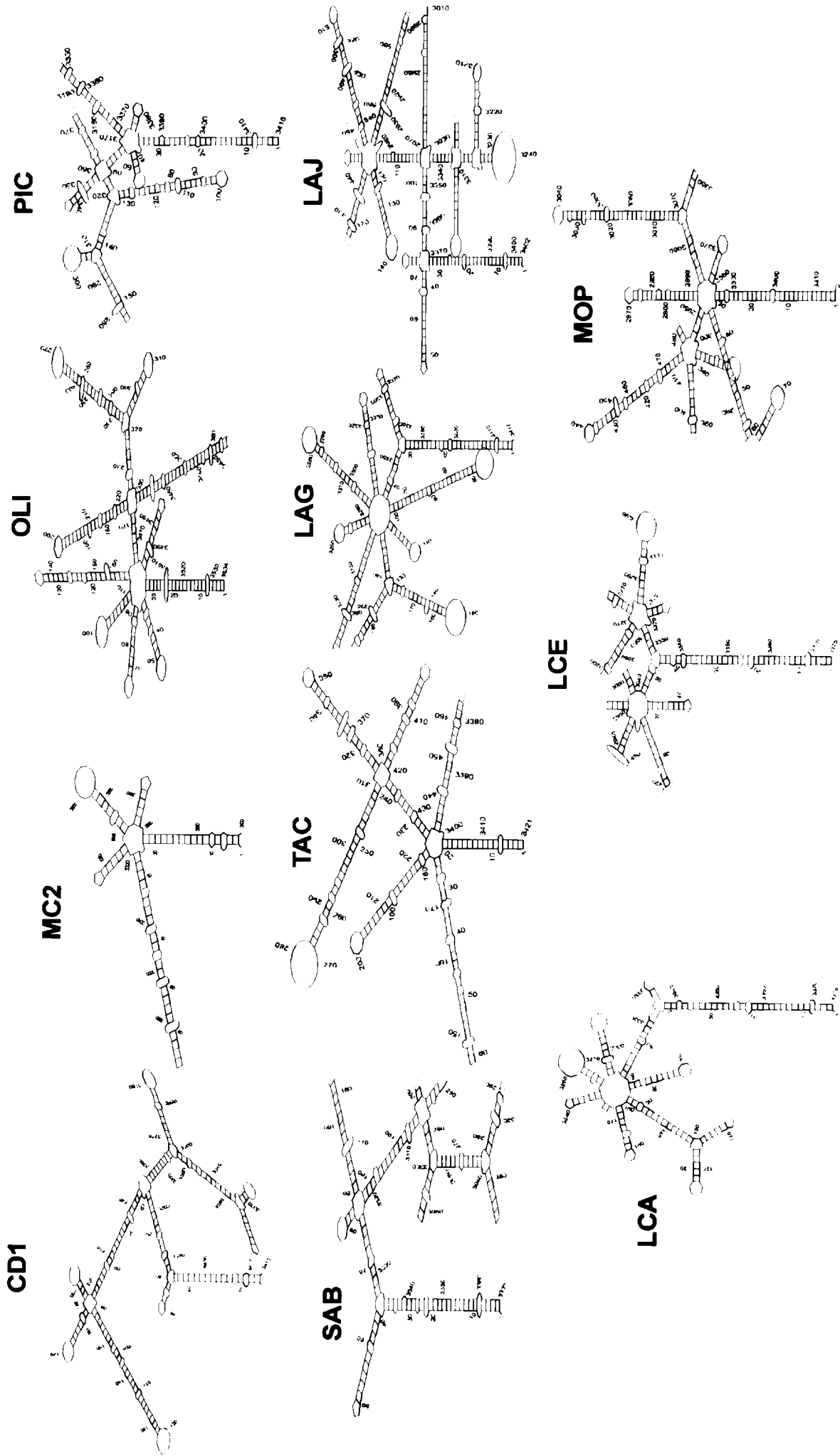


Figura 12.23. Estructuras tipo panhandle correspondientes a los RNAs S virales de 11 arenavirus. Las estructuras fueron calculadas de acuerdo a lo descrito en *Materiales y Métodos* y en el Capítulo 18. Las estructuras mostradas corresponden a segmentos de la regiones respectivas de las estructuras completas. CD1: Junin-Candid #1; MC2: Junin-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Nigeria; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia.

RNAs S virales de arnavirus del Nuevo Mundo Estructuras tipo *hairpin loop* intergénicas

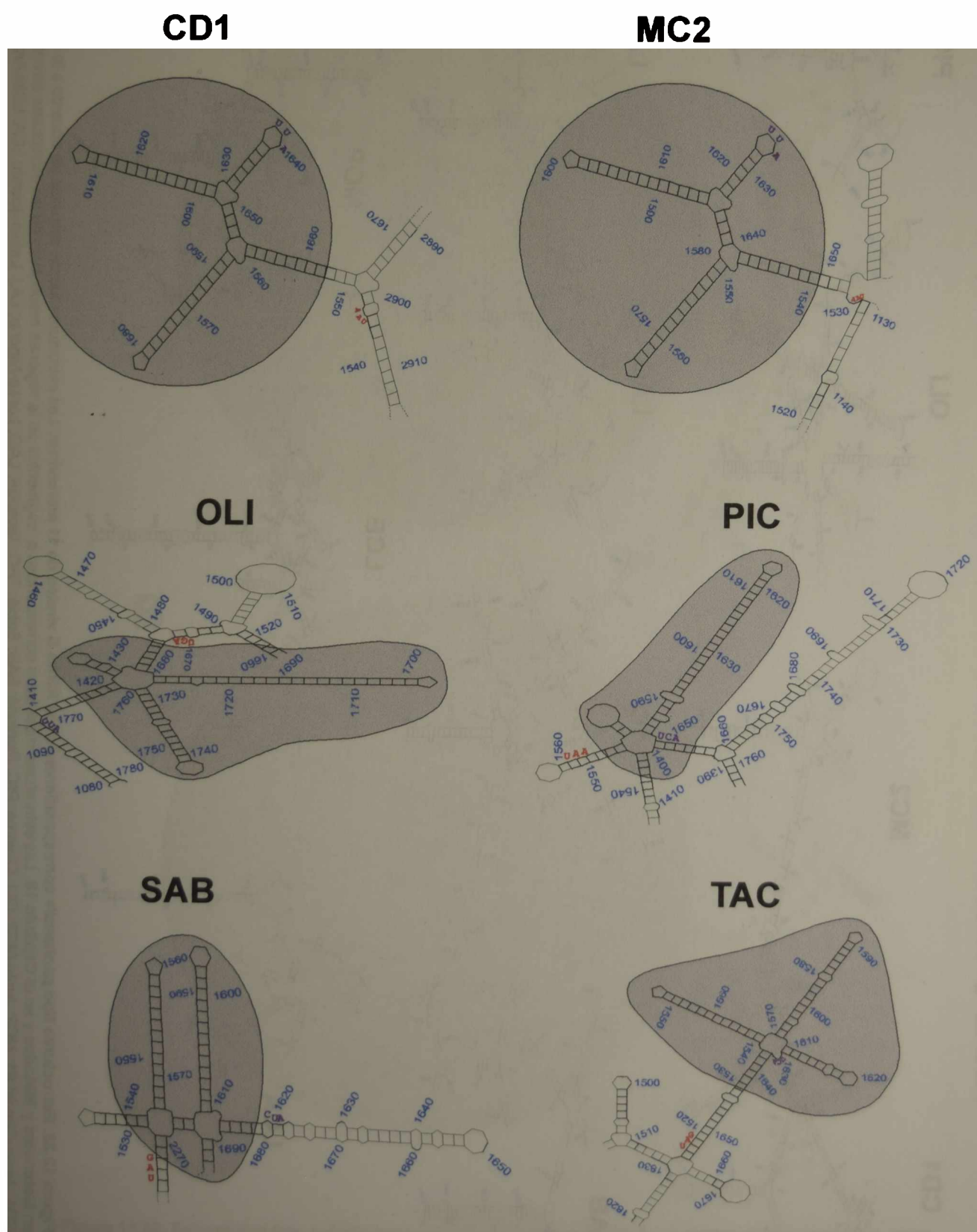


Figura 12.24. Estructuras tipo *hairpin loop* correspondientes a los RNAs S virales de arnavirus del Nuevo Mundo. Las estructuras fueron calculadas de acuerdo a lo descrito en **Materiales y Métodos** y en el **Capítulo 18**. Las estructuras mostradas son una ampliación de las regiones correspondientes de las estructuras completas respectivas. En rojo se indican los codones de terminación de los marcos de lectura codificantes para GPC y en violeta los tripletes complementarios a los codones de terminación de los marcos de lectura codificantes para N.
CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe

RNAs S virales de arnavirus del Viejo Mundo Estructuras tipo *hairpin loop* intergénicas

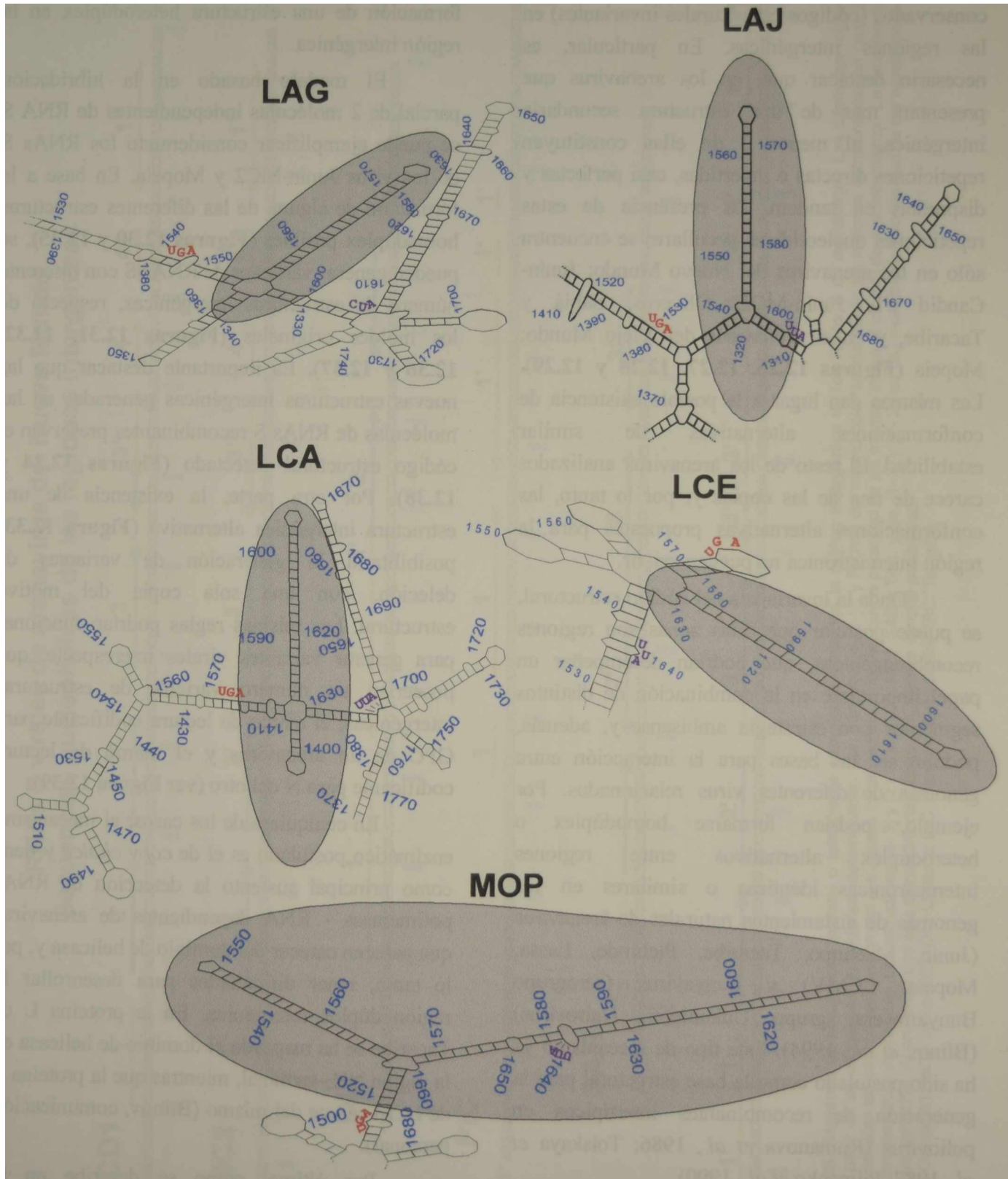


Figura 12.25. Estructuras tipo *hairpin loop* correspondientes a los RNAs S virales de arnavirus del Viejo Mundo. Las estructuras fueron calculadas de acuerdo a lo descrito en **Materiales y Métodos** y en el **Capítulo 18**. Las estructuras mostradas son una ampliación de las regiones correspondientes de las estructuras completas respectivas. En rojo se indican los codones de terminación de los marcos de lectura codificantes para GPC y en violeta los triplete complementarios a los codones de terminación de los marcos de lectura codificantes para N.
LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia

Estructura secundaria de RNA y mecanismos de recombinación

Un análisis detallado de las secuencias nucleotídicas de los RNAs S de 11 arnavirus revela la existencia de elementos altamente conservados (códigos estructurales invariantes) en las regiones intergénicas. En particular, es necesario destacar que, en los arnavirus que presentan más de una estructura secundaria intergénica, al menos 2 de ellas constituyen repeticiones directas e invertidas, casi perfectas y dispuestas en tandem. La presencia de estas repeticiones nucleotídicas peculiares se encuentra sólo en los arnavirus del Nuevo Mundo: Junín-Candid #1, Junín-MC2, Oliveros, Sabiá y Tacaribe, y en el arnavirus del Viejo Mundo: Mopeia (Figuras 12.26, 12.27, 12.28 y 12.29). Las mismas dan lugar a la posible existencia de conformaciones alternativas de similar estabilidad. El resto de los arnavirus analizados carece de una de las copias y, por lo tanto, las conformaciones alternativas propuestas para la región intercistronica no pueden existir.

Dada la invariancia del código estructural, se puede postular que estas zonas son regiones recombinogénicas, que podrían desempeñar un papel importante en la combinación de distintos segmentos con estrategia ambisense y, además, podrían ser las bases para la interacción entre genomas de diferentes virus relacionados. Por ejemplo, podrían formarse homodúplex o heterodúplex alternativos entre regiones intercistronicas idénticas o similares en los genomas de aislamientos naturales de arnavirus (Junín, Machupo, Tacaribe, Pichindé, Lassa, Mopeia, LCMV) y bunyavirus (serogrupo Bunyamwera, grupo Uukuniemi, Nairovirus) (Blinov *et al.*, 1994). Este tipo de mecanismo ya ha sido postulado como la base estructural para la generación de recombinantes intertípicos en poliovirus (Romanova *et al.*, 1986; Tolskaya *et al.*, 1987; Pilipenko *et al.*, 1990).

Analizando modelos específicos para arnavirus se pueden considerar, al menos, dos situaciones que posibilitarían la generación de RNAs S recombinantes intraespecie y, al menos una, que posibilitaría la generación de recombinantes interespecie. En el primer caso, el

modelo se basa en la hibridación parcial de 2 moléculas independientes de RNA S y/o en la coexistencia de estructuras intergénicas alternativas. En el segundo caso, el modelo se basa en la coinfección con dos arnavirus relacionados y la formación de una estructura heterodúplex en la región intergénica.

El modelo basado en la hibridación parcial de 2 moléculas independientes de RNA S se puede ejemplificar considerando los RNAs S de los virus Junín-MC2 y Mopeia. En base a la formación de alguna de las diferentes estructuras homodúplex posibles (Figuras 12.30 y 12.35), se pueden generar variantes de RNAs S con diferente número de estructuras intergénicas, respecto de los moldes originales (Figuras 12.31, 12.32, 12.36 y 12.37). Es importante destacar que las nuevas estructuras intergénicas generadas en las moléculas de RNAs S recombinantes preservan el código estructural detectado (Figuras 12.34 y 12.38). Por otra parte, la existencia de una estructura intergénica alternativa (Figura 12.33) posibilitaría la generación de variantes de delección, con una sola copia del motivo estructural. Las mismas reglas podrían funcionar para generar variantes virales interespecie, que poseerían un número variable de estructuras intergénicas, el marco de lectura codificante para GPC de un arnavirus y el marco de lectura codificante para N del otro (ver Figura 12.39).

En cualquiera de los casos, el mecanismo enzimático postulado es el de *copy choice* y tiene como principal sustento la detección de RNAs polimerasas - RNA dependientes de arnavirus que parecen carecer del dominio de helicasa y, por lo tanto, tener dificultades para desenrollar la región dúplex intergénica. En la proteína L de Tacaribe se ha mapeado el dominio de helicasa en la región NH₂-terminal, mientras que la proteína L de LCM carece del mismo (Blinov, comunicación personal).

Por último, como se describe en la discusión, la existencia de repeticiones directas e invertidas no se limita sólo a las regiones intergénicas. En función de esto, la probabilidad de emergencia viral, mediante alguno de los mecanismos de recombinación postulados, aumenta significativamente.

RNAs S virales de arnavirus del Nuevo Mundo
Secuencias nucleotídicas de las regiones intergénicas

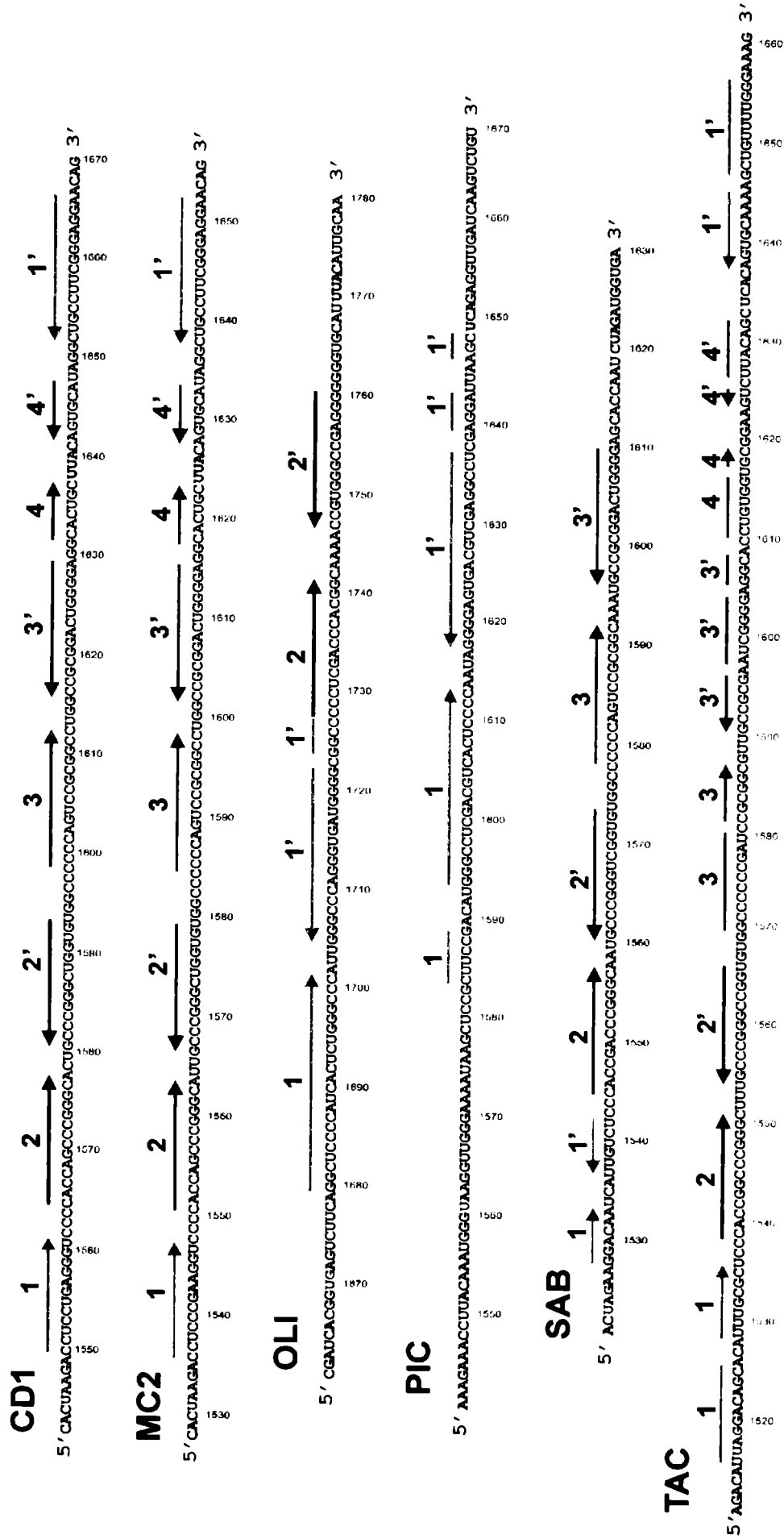


Figura 12.26. Regiones de secuencias parcial o totalmente complementarias en la región intergénica de los RNAs S de arnavirus. Las diferentes regiones fueron ubicadas a partir de las estructuras secundarias determinadas. Con números (ej. 1-1') y flechas del mismo color se identifican las regiones complementarias en cada secuencia. Las flechas señalan las secuencias que forman estructuras doble cadena, mientras que la polaridad de las mismas indica el sentido de apareamiento. La cantidad de regiones con estas características es variable. Las estructuras [2-3] y [2'-3'] en CD1, MC2, SAB y TAC, y las [1-2] y [1'-2'] en OLI muestran la peculiaridad de ser repeticiones directas e invertidas casi perfectas. Al presentar una sola estructura intergénica, PIC carece de esta posibilidad. CD1: Junin-Candid #1; MC2: Junin-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe.

RNAs S virales de arnavirus Comparación de estructuras tipo hairpin loop intergénicas

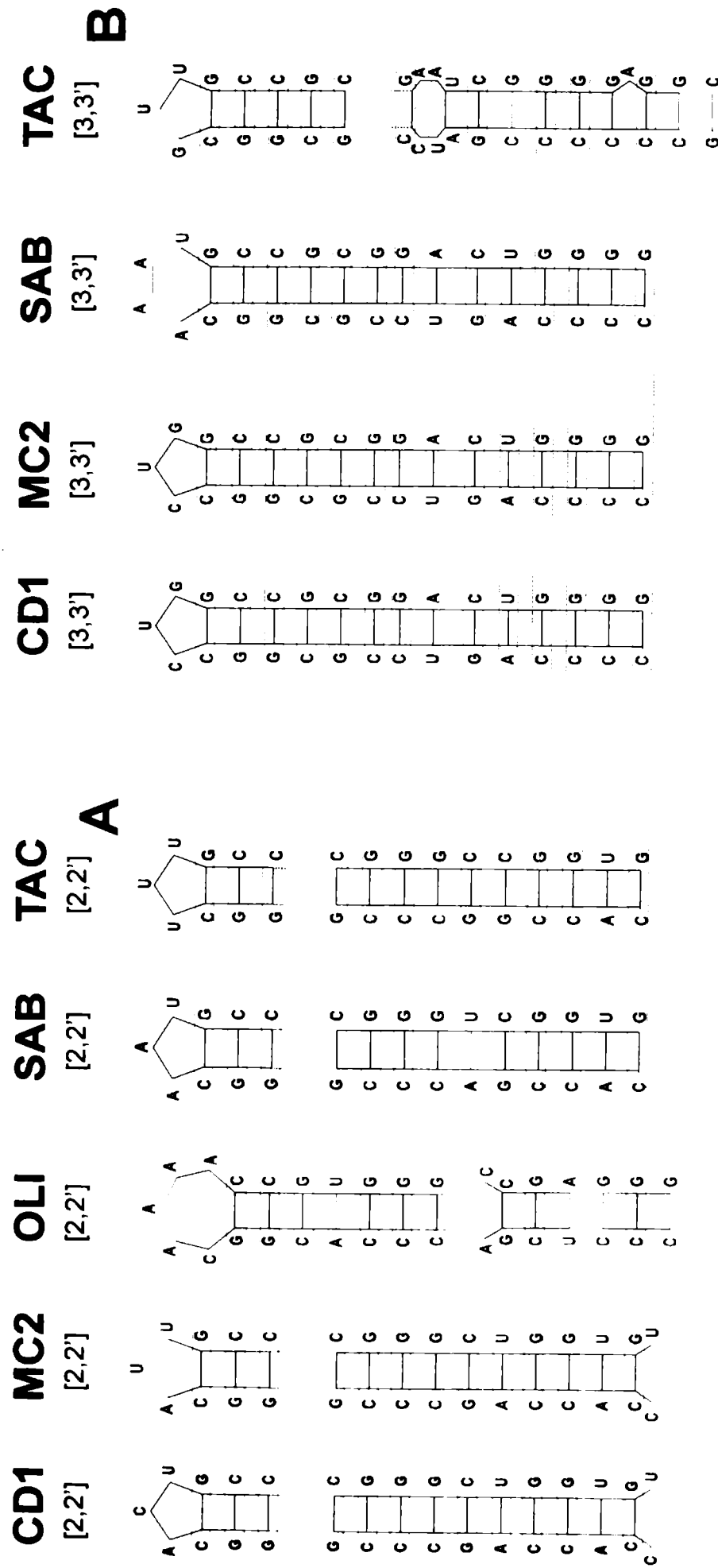


Figura 12.27. Comparación de las estructuras intergénicas de los arnavirus con doble estructura tipo hairpin loop. Se comparan solamente las estructuras correspondientes a las repeticiones directas e invertidas. Para facilitar la comparación, en algunas de las estructuras se han introducido gaps. Para destacar la conservación del motivo estructural, las identidades de apareamientos nucleotídicos en las mismas posiciones dentro del grupo se indican con recuadros naranjas sombreados de amarillo pálido y con recuadros verdes cuando están conservadas en ambos grupos. **A.** Comparación de las estructuras formadas por [2-2'] de CD1, MC2, OLI, SAB y TAC. **B.** Comparación de las estructuras formadas por [3-3'] de CD1, MC2, SAB y TAC. La estructura formada por [1-1'] de OLI no se ha incluido debido a las diferencias que presenta con el resto. Es notorio el alto grado de conservación en ambos tipos de estructura. En particular, las estructuras [3-3'] se caracterizan por ser prácticamente idénticas en la región dúplex. CD1: JunIn-Candid #1; MC2: JunIn-MC2; SAB: Sabiá; TAC: Tacaribe.

**RNAs S virales de arnavirus del Viejo Mundo
Secuencias nucleotídicas de las regiones intergénicas**

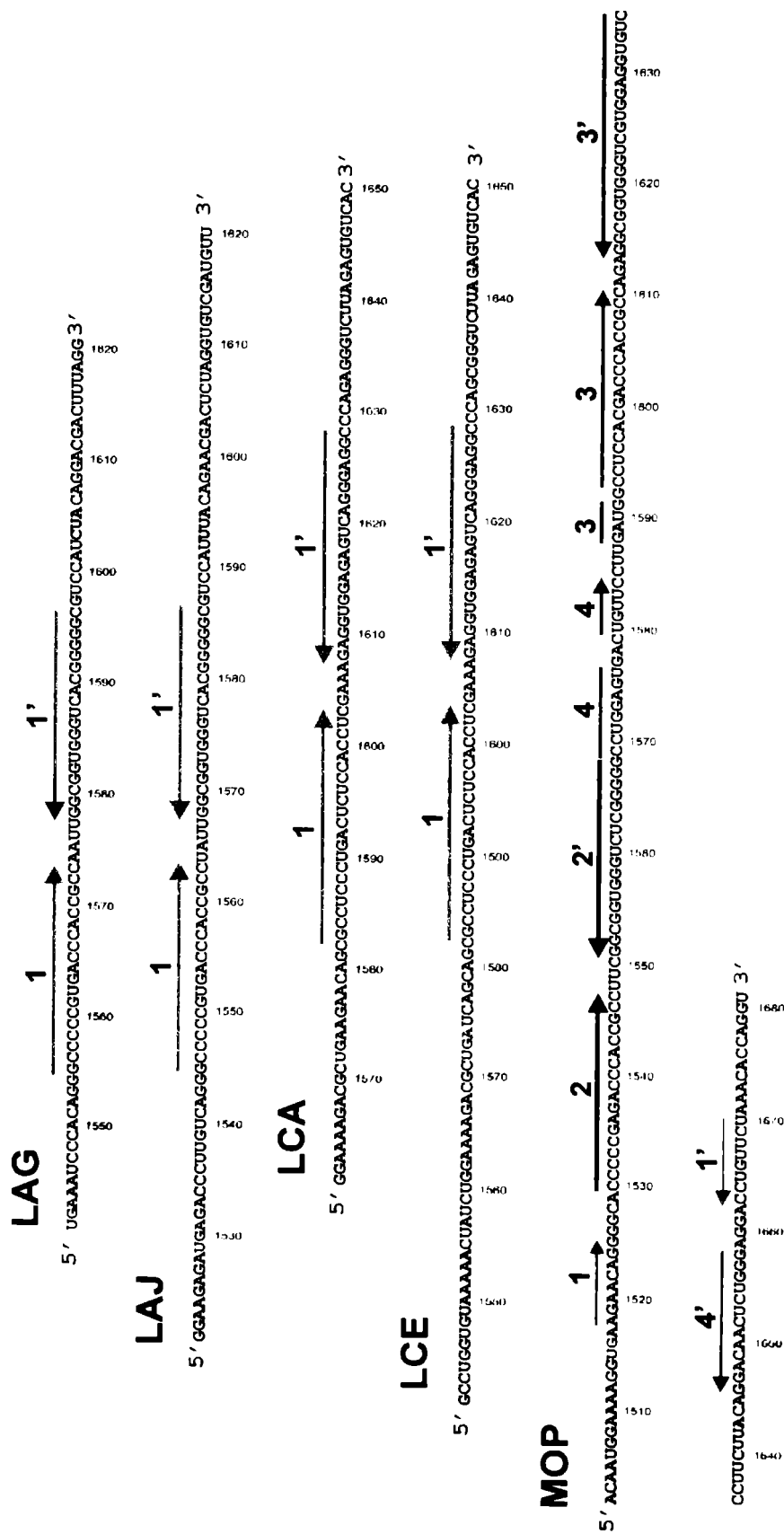


Figura 12.28. Regiones de secuencias parcial o totalmente complementarias en la región intergénica de los RNAs S de arnavirus. Las diferentes regiones fueron ubicadas a partir de las estructuras secundarias determinadas. Con números (ej. 1-1') y flechas del mismo color se identifican las regiones complementarias en cada secuencia. Las flechas señalan las secuencias que forman estructuras doble cadena, mientras que la polaridad de las mismas indica el sentido de apareamiento. Salvo MOP, el resto de los arnavirus del Viejo Mundo presenta una única estructura intergénica. Las estructuras [2-3] y [2'-3'] en MOP muestran la peculiaridad de ser repeticiones directas e invertidas casi perfectas. Al presentar una sola estructura intergénica, el resto carecen de esta posibilidad.
LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia.

RNAs S virales de arnavirus Comparación de estructuras tipo hairpin loop intergénicas

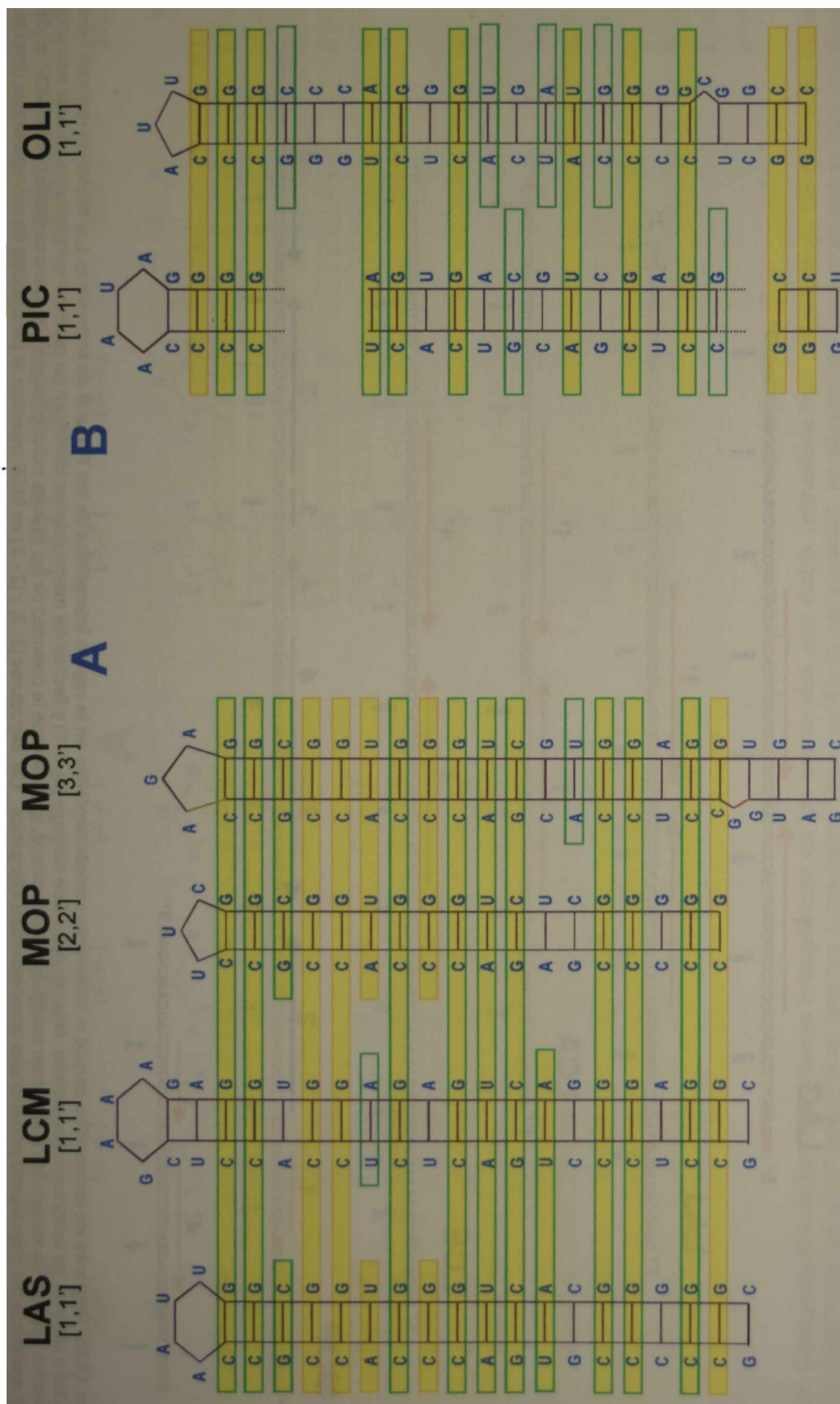


Figura 12.29. Comparación de las estructuras intergénicas de los arnavirus con simple estructura tipo hairpin loop. Se comparan solamente las estructuras correspondientes a las repeticiones directas e invertidas. Para facilitar la comparación, en algunas de las estructuras se han introducido gaps. Para destacar la conservación del motivo estructural, las identidades de apareamientos nucleotídicos en las mismas posiciones dentro del grupo se indican con recuadros naranjas sombreados de amarillo pálido y con recuadros verdes cuando están conservadas en ambos grupos. A. Comparación de las estructuras formadas por [1-1] de LAS y LCM, y [2-2] y [3-3] de MOP. B. Comparación de las estructuras formadas por [1-1] de PIC y OLI. Es notorio el alto grado de conservación dentro del panel A. LAS: Lassa-Nigeria y Lassa-Josiah; LCM: LCM-Armstrong y LCM-WÉ; MOP: Mopeia; OLI: Oliveros; PIC: Pichindé.

RNA S viral del virus Junín MC2
Posibles estructuras doble cadena en la región intergénica

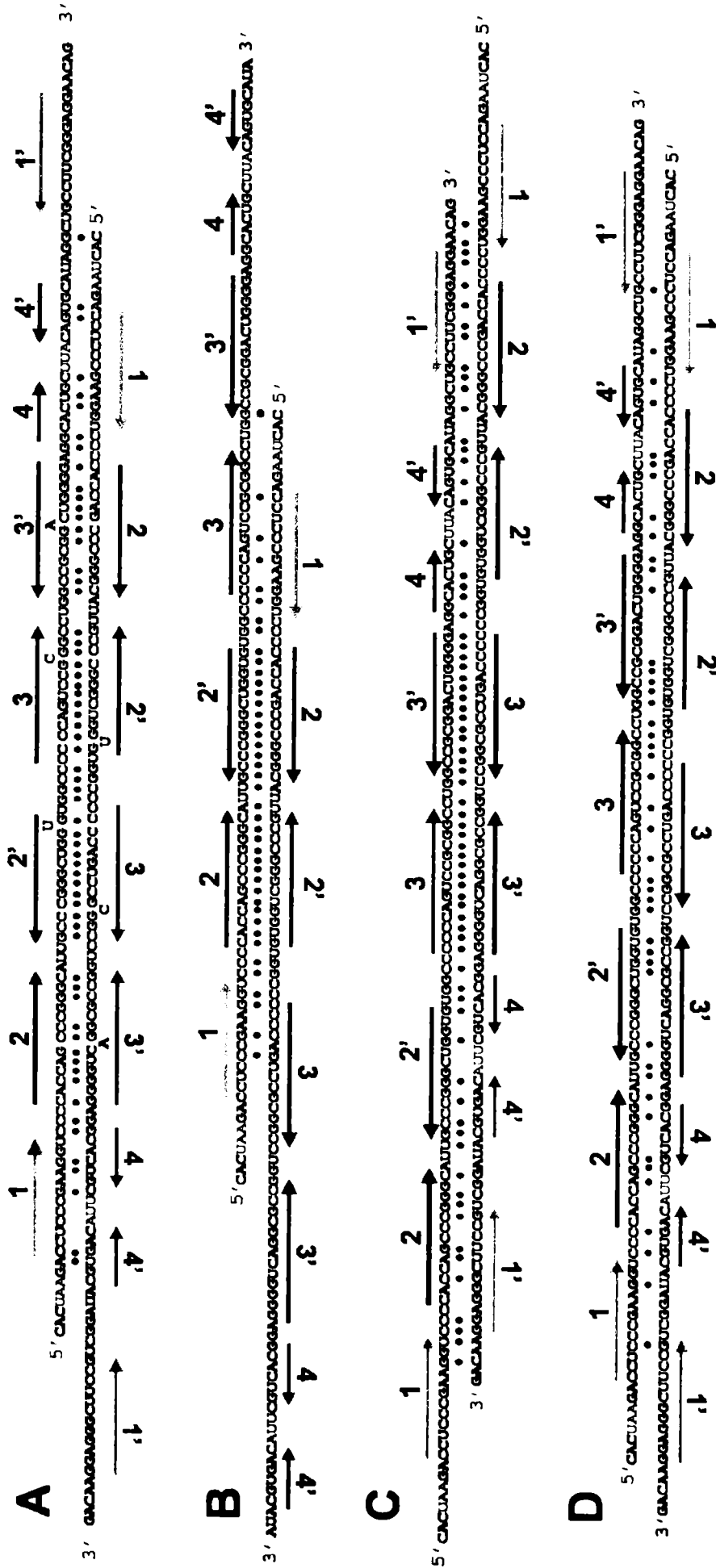


Figura 12.30. Diferentes estructuras doble cadena formadas por apareamiento parcial en la región intergénica de dos moléculas de RNA S viral del virus Junín-MC2. A. Estructura doble cadena potencial estabilizada por apareamiento de las regiones [2-3], [2-3], [3-2] y [3'-2]. B. Estructura doble cadena potencial estabilizada por apareamiento de las regiones [2-2] y [2'-2]. C. Estructura doble cadena potencial estabilizada por apareamiento de las regiones [3-3] y [3'-3]. D. Estructura doble cadena potencial estabilizada por apareamiento imperfecto de las regiones [2-3], [2'-3], [3-2] y [3'-2]. En rojo se muestra el codón de terminación del marco de lectura codificante para N.

RNA S viral del virus Junín MC2 Mecanismo que permite el incremento en el número de estructuras de la región intergénica

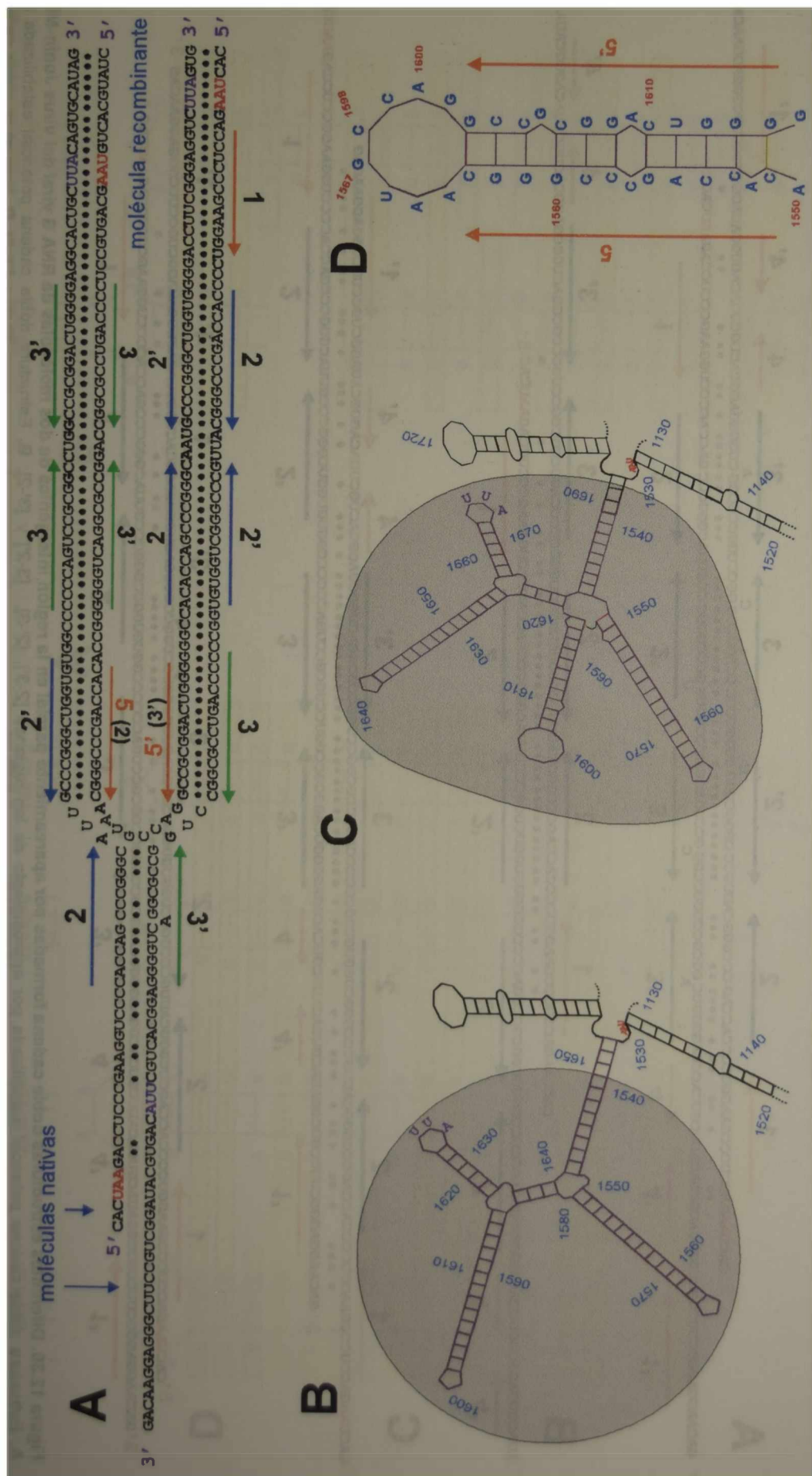


Figura 12.31. Generación de un recombinante del virus Junín-MC2 con triple estructura intergénica. A. Secuencia nucleotídica del híbrido parcial entre dos cadenas de RNA S viral (ver Figura 12.30) y secuencia nucleotídica de la molécula recombinante. B. Detalle de las estructuras tipo *hairpin-loop* originales. C. Detalle de las estructuras tipo *hairpin-loop* recombinantes. D. Detalle estructural del marco de terminación del marco de lectura codificante para GPC y en violeta el triplete complementario al marco de lectura codificante para N.

RNA S viral del virus Junín MC2 Mecanismo que permite la disminución del número de estructuras de la región intergénica

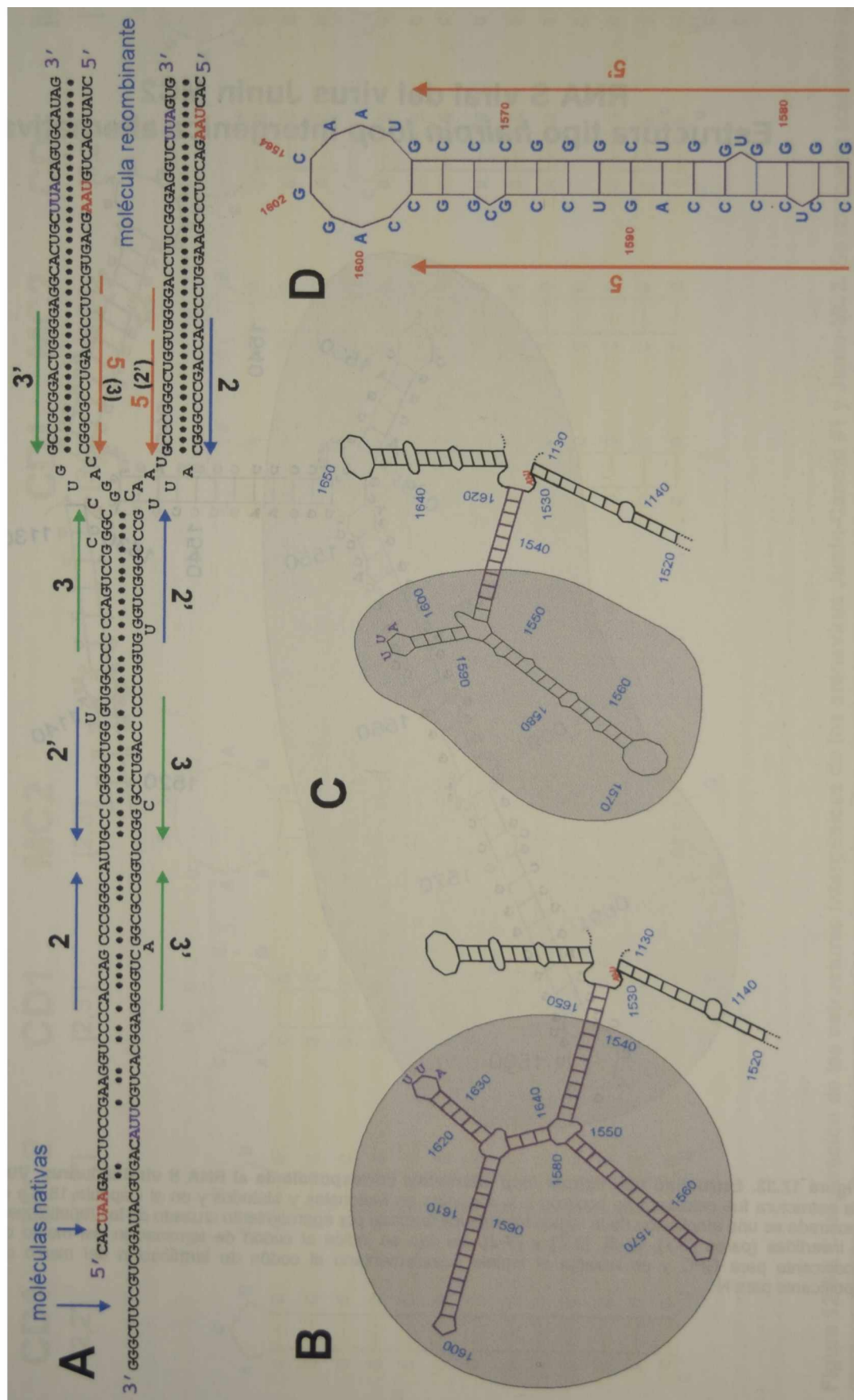


Figura 12.32. Generación de un recombinante del virus Junín-MC2 con simple estructura Intergénica. A. Secuencia nucleotídica del híbrido parcial entre dos cadenas de RNA S viral (ver Figura 12.30) y secuencia nucleotídica de la molécula recombinante. B. Detalle de las estructuras tipo *hairpin-loop* originales. C. Detalle de la estructura tipo *hairpin-loop* recombinante. D. Detalle estructural del *hairpin-loop* recombinante [5-5'] generado por duplicación de las secuencias complementarias a 2 y 3', y pérdida de las regiones originales correspondientes a 2, 2' y 3. En rojo se muestra el codón de terminación del marco de lectura codificante para GPC y en violeta el triplete complementario al marco de lectura codificante para N.

RNA S viral del virus Junín MC2 Estructura tipo *hairpin loop* intergénica alternativa

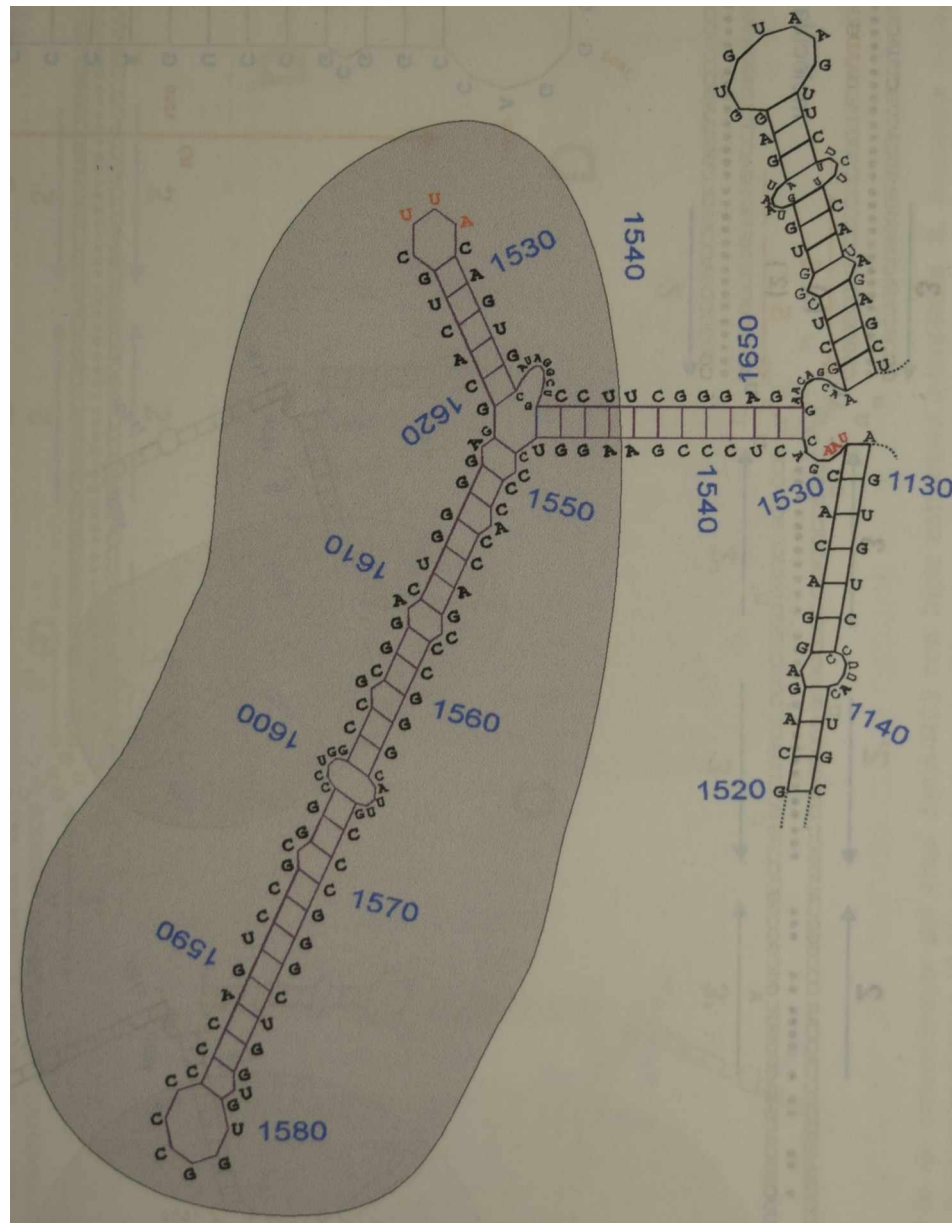


Figura 12.33. Estructura tipo *hairpin loop* alternativa correspondiente al RNA S viral del virus Junín-MC2. La estructura fue calculada de acuerdo a lo descrito en Materiales y Métodos y en el Capítulo 18. La estructura mostrada es una ampliación de la región intergénica formada por apareamiento cruzado de las repeticiones directas e invertidas (pares [2'-3'], [2'-3], [3'-2'] y [3'-2]). En rojo se indica el codón de terminación del marco de lectura codificante para GPC y en naranja el triplete complementario al codón de terminación del marco de lectura codificante para N.

RNAs S virales de las cepas Candid 1 y MC2 del virus Junín Comparación de estructuras tipo hairpin loop originales y recombinantes

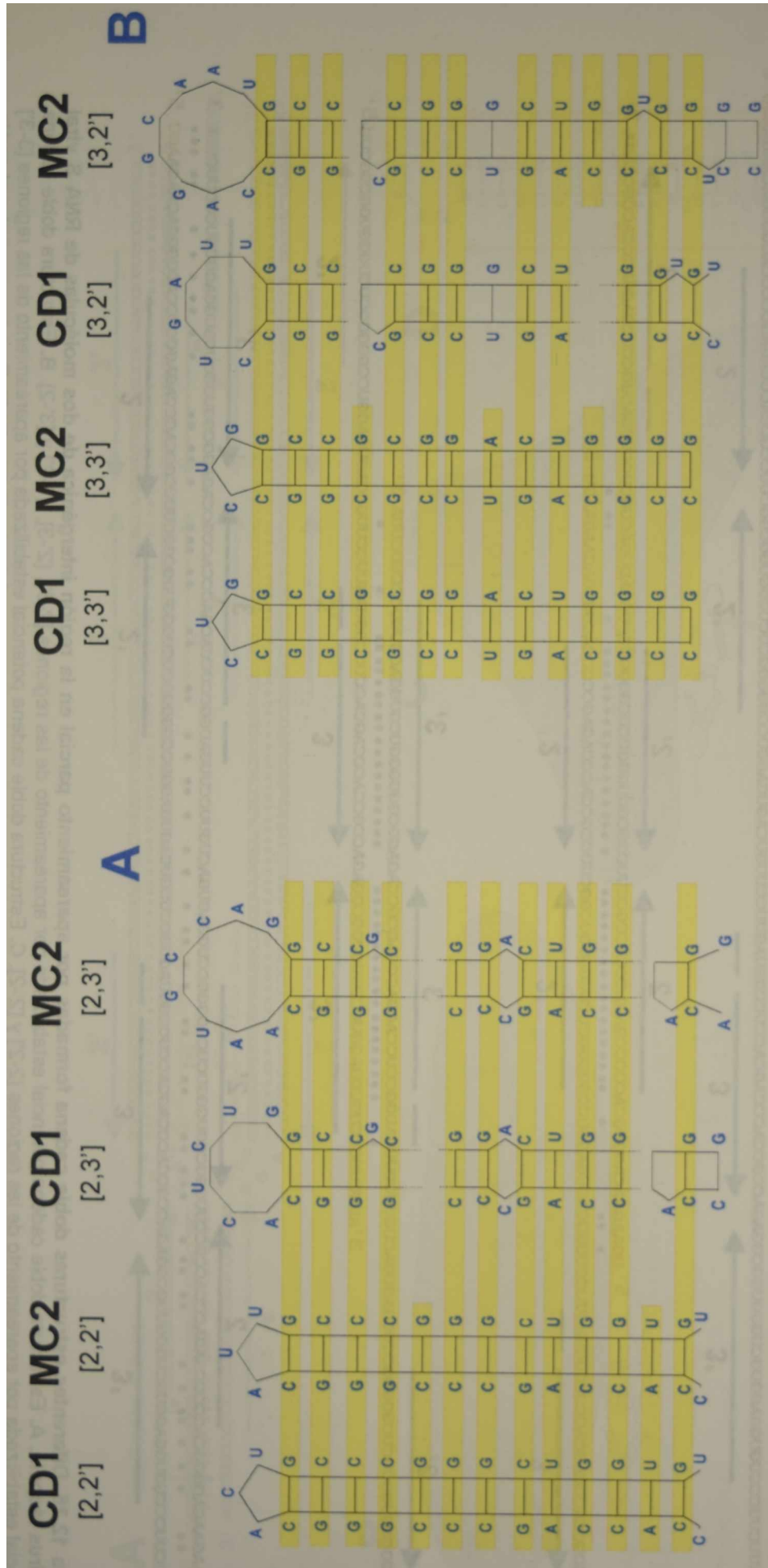


Figura 12.34. Comparación de las estructuras intergénicas de los arnavirus JunIn-Candid #1 y JunIn-MC2. Se comparan solamente las estructuras correspondientes a las repeticiones directas e invertidas. Para facilitar la comparación, en algunas de las estructuras se han introducido gaps. Para destacar la conservación del motivo estructural, las identidades de apareamientos nucleotídicos en las mismas posiciones dentro del grupo se indican con recuadros naranjas sombreados de amarillo pálido. **A.** Comparación de las estructuras nativas formadas por [2-2'] de CD1 y MC2, y las estructuras recombinantes [2-3'] de ambos virus. **B.** Comparación de las estructuras nativas formadas por [3-3'] de CD1 y MC2, y las estructuras recombinantes [3-2'] de ambos virus. Es notorio el alto grado de conservación dentro de cada panel. CD1: JunIn-Candid #1; MC2: JunIn-MC2.

RNA S viral del virus Mopeia
Posibles estructuras doble cadena en la región intergénica

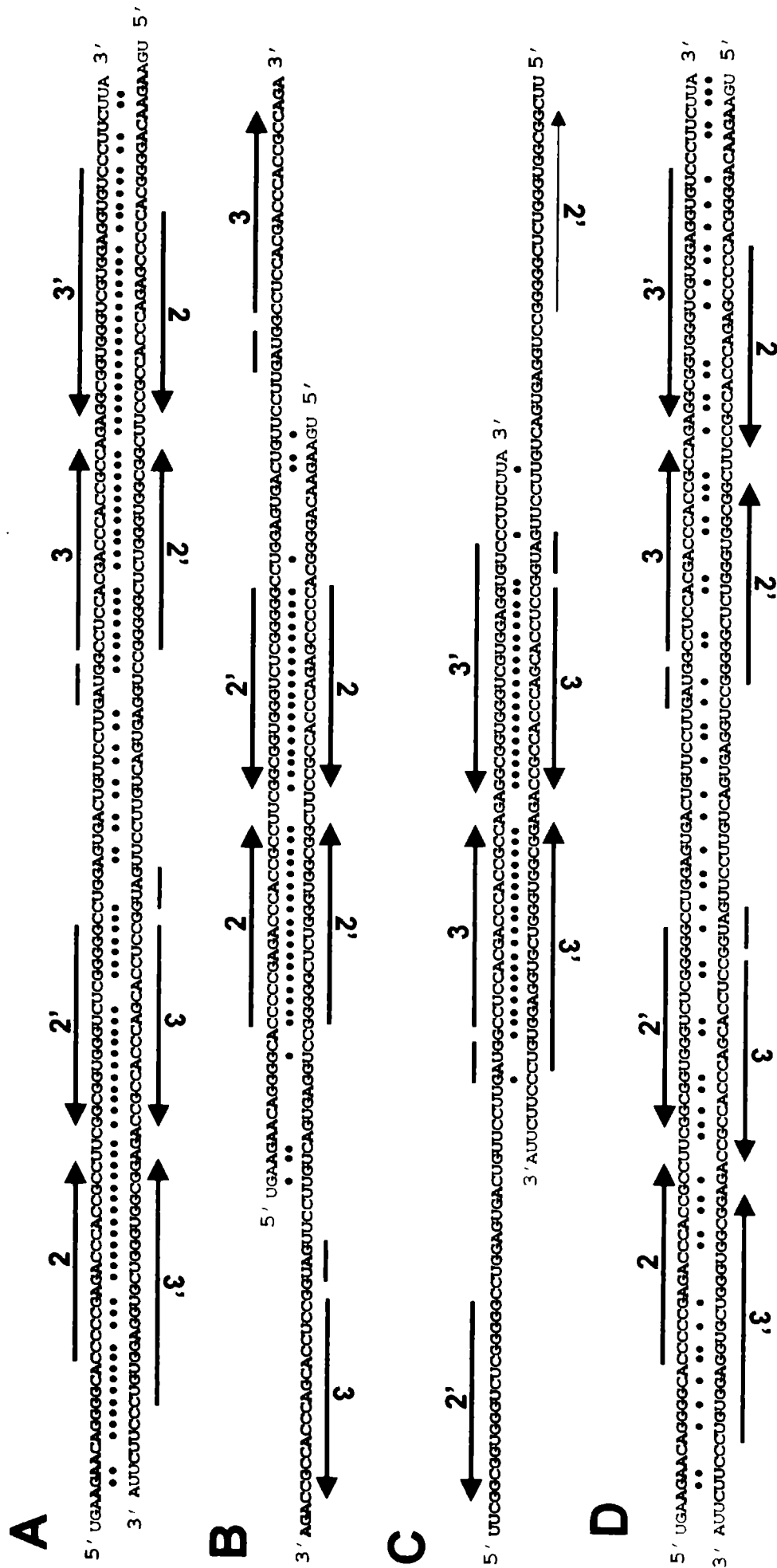


Figura 12.35. Diferentes estructuras doble cadena formadas por apareamiento parcial en la región intergénica de dos moléculas de RNA S viral del virus Mopeia. A. Estructura doble cadena potencial estabilizada por apareamiento de las regiones [2-3], [2'-3], [3-2] y [3'-2]. B. Estructura doble cadena potencial estabilizada por apareamiento de las regiones [2-2'] y [2'-2]. C. Estructura doble cadena potencial estabilizada por apareamiento de las regiones [3-3] y [3'-3]. D. Estructura doble cadena potencial estabilizada por apareamiento imperfecto de las regiones [2-3], [2'-3], [3-2] y [3'-2]. En rojo se muestra el codón de terminación del marco de lectura complementario para GPC y en violeta el triplete codificante para N.

RNA S viral del arnavirus Mopeia Comparación de estructuras tipo hairpin loop originales y recombinantes

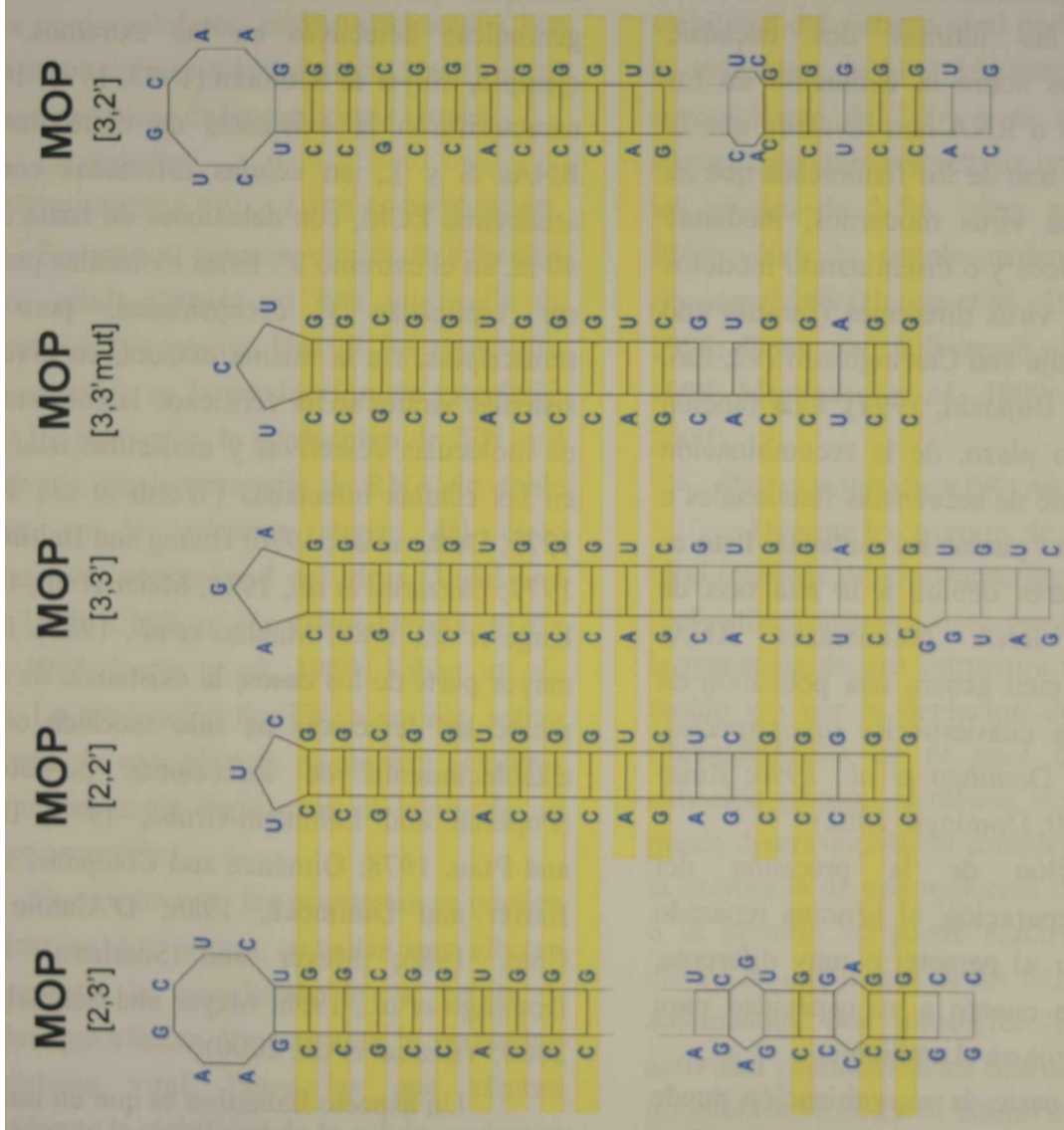


Figura 12.38. Comparación de las estructuras intergénicas del arnavirus Mopeia. Se comparan solamente las estructuras correspondientes a las repeticiones directas e invertidas. Para facilitar la comparación, en algunas de las estructuras se han introducido gaps. Para destacar la conservación del motivo estructural, las identidades de apareamientos nucleotídicos en las mismas posiciones dentro del grupo se indican con recuadros naranjas sombreados de amarillo pálido. Es notorio el alto grado de conservación entre las estructuras nativas formadas por [2-2'] y [3-3'], y las estructuras recombinantes [2-3'] y [3-2'], como también con la estructura [3-3' mut] que puede aparecer durante uno de los procesos de recombinación. MOP: Mopeia.

Discusión sobre la estructura secundaria de los RNAs S, la existencia de repeticiones invertidas, perfectas o imperfectas, y los posibles mecanismos de recombinación

La recombinación de RNA, un proceso que lleva a la unión covalente de dos o más moléculas de RNA, es considerado uno de los mecanismos más importantes en la reparación y evolución de los genomas virales a RNA.

Durante las últimas dos décadas, diferentes estudios sobre la evolución de los virus con genoma a RNA han revelado que la recombinación es uno de los fenómenos que ha dado forma a los virus modernos, mediante rearrreglos genómicos y/o diseminando módulos funcionales entre virus diferentes (Strauss and Strauss, 1988; Dolja and Carrington, 1992; Lai, 1992; Simon and Bujarski, 1994). Una función adicional, a corto plazo, de la recombinación puede ser el rescate de secuencias funcionales a partir de moléculas parentales mutadas. Esto es de particular interés debido a la alta tasa de error de las RNA polimerasas RNA-dependientes, lo cual genera una población de mutantes llamada cuasiespecies (*quasispecies*, Domingo, 1992; Domingo *et al.*, 1996; Ruiz-Jarabo *et al.*, 2000; Domingo, 2002).

En función de la precisión del mecanismo de reparación, el genoma reparado puede ser similar al parental o muy diferente, especialmente en cuanto a su capacidad para superar los cambios en el ambiente.

Por otra parte, la recombinación puede ser mucho más frecuente de lo esperado. King (1988) ha estimado que alrededor del 10-20 % de los genomas de poliovirus están sometidos a algún evento de recombinación durante cada ciclo replicativo. Chetverin (1999) estima que la recombinación no homóloga es tanto o más frecuente que la recombinación homóloga en coronavirus, picornavirus, alphavirus y fagos con genoma de RNA.

La necesidad, la oportunidad

Una particularidad de los sistemas virales es la generación de partículas defectivas,

que, en algún momento, se pueden transformar en partículas que interfieren con la replicación de las viables tipo *wild type* (partículas defectivas interferentes, DI). Probablemente estas DI forman parte de la población de cuasiespecies (Domingo, 1992; Domingo *et al.*, 1998).

En la literatura científica existen múltiples publicaciones acerca de moléculas genómicas defectivas en los extremos. Por ejemplo, Meyer & Southern (1993, 1994, 1997) caracterizaron la existencia de moléculas de RNAs S y L, en células infectadas con el arnavirus LCM, con deleciones de hasta 30 o 40 nt. en el extremo 3'. Estas moléculas pueden ser replicadas y encapsidadas, pero no transcriptas. De la misma manera, en diversos sistemas virales se ha verificado la coexistencia de moléculas defectivas y moléculas *wild type* en las células infectadas (Welsh *et al.*, 1972, 1975; Dutko *et al.*, 1976; Huang and Baltimore, 1977; Lazzarini *et al.*, 1981; Meier *et al.*, 1984; Kuge *et al.*, 1986; Makino *et al.*, 1988). En la mayor parte de los casos, la existencia de estas moléculas defectivas ha sido asociada con el establecimiento de infecciones persistentes (Popescu and Lehmann-Grube, 1977; Dutko and Pfau, 1978; Giménez and Compans, 1980; Barret and Dimmock, 1986; D'Aiutolo and Coto, 1986; Meyer and Southern, 1997; Domingo *et al.*, 1998; Meyer and Schmaljohn, 2000; Villareal *et al.*, 2000).

Un aspecto llamativo es que en ninguna de las publicaciones mencionadas anteriormente se describe la existencia de moléculas defectivas con deleciones importantes en los extremos 3' (100 o más nt). Probablemente, esto se debe a que las metodologías utilizadas para detectar los RNAs defectivos sólo permiten observar una pequeña fracción de la población existente.

No obstante esto, dado que existen múltiples mecanismos que podrían asegurar la existencia de moléculas defectivas y/o fragmentos de RNAs, tanto virales como

celulares, hay grandes chances de que la población de cuasiespecies esté formada no sólo por moléculas defectivas derivadas de un evento replicativo o transcripcional abortivo, sino también por fragmentos de RNA derivados de procesos degradativos, asociados con la infección viral. Algunos de los mecanismos que posibilitarían la existencia de moléculas defectivas y/o fragmentos de RNAs, tanto virales como celulares, están relacionados con respuestas celulares a la infección viral y otros con características de las replicasas virales.

En síntesis:

➤ RESPUESTAS CELULARES APOPTÓTICAS

En muchos sistemas virales, la infección de una célula dispara en ésta una serie de respuestas a la agresión. Una de las respuestas más conocidas es la producción de interferón (IFN). En particular, la producción de IFN α/β es activada por la presencia de RNA de doble cadena, en las primeras etapas del ciclo replicativo/transcripcional del virus (Vidal and Coto, 1980; Clark *et al.*, 1986; Dejean *et al.*, 1987, 1988; Lucia *et al.*, 1989; Kibler *et al.*, 1997). La producción de IFN, a su vez, activa una serie de cascadas de respuestas celulares que, normalmente, derivan en el inicio de un proceso apoptótico.

Esencialmente, las respuestas se pueden clasificar en dos grupos, aquellas que afectan directamente la viabilidad de la célula y aquellas que afectan directamente la viabilidad del sistema viral. Entre las que afectan directamente la viabilidad de la célula, podemos mencionar la activación de PKR (*protein kinase R*) que participa del inicio de la apoptosis celular mediante la inducción de DISC (*death-induced signaling complexes*), que comprenden una compleja sucesión de cascadas de activación de factores proteicos que llevan a la apoptosis celular; y la reubicación de la proteína nuclear PML (*promyelocyte leukemia protein*) al citoplasma, mediada por la proteína Z (en el caso de arnavirus). En cualquiera de estos casos, uno de los resultados principales es

una inhibición de los eventos traduccionales, mediada por interacción directa entre una proteína celular y/o viral con un factor traduccional (por ejemplo, PML/eIF4E y/o Z/eIF4E) o mediante un clivaje selectivo del RNA ribosomal 28S (Borden *et al.*, 1998a, 1998b; Campbell Dwyer *et al.*, 2000; Lee *et al.*, 2000; Balachandran *et al.*, 2000; Kentsis *et al.*, 2001). Entre las que afectan directamente la viabilidad del sistema viral podemos mencionar la inducción de la 2-5A sintetasa, la síntesis concomitante de 2-5A y la activación de la RNasa L (endorribonucleasa latente) que, según los niveles de 2-5A, cliva preferencialmente RNA viral de simple cadena y/o el RNA ribosomal 28S (Houge *et al.*, 1995; Mitra *et al.*, 1996; Dong and Silverman, 1997; Li *et al.*, 1998; Martinand *et al.*, 1999; Banerjee *et al.*, 2000).

➤ PECULIARIDADES DE LAS REPLICASAS

Durante los eventos de síntesis de RNA viral (replicación y/o transcripción) el complejo replicativo/transcripcional puede detenerse por la presencia de una estructura secundaria en el molde y/o por la depleción de alguno de los ribonucleótidos en el *pool* de rNTPs. Esta situación plantea, al menos, dos soluciones: se puede desensamblar el complejo dando lugar a la existencia de una molécula de RNA abortiva o el proceso se puede reactivar mediante el clivaje de uno o más de los últimos nucleótidos adicionados por pirofosforólisis o por una actividad endonucleásica celular o asociada a la polimerasa viral. En el último caso, dado que la reactivación depende de un balance general entre la disponibilidad de precursores y las actividades pirofosforolítica/endonucleolítica y polimerásica, también puede derivar en el desensamblaje de los complejos posteriormente a una pirofosforólisis/endonucleólisis extensiva; dando lugar, por lo tanto, a la existencia de moléculas de RNA abortivas (Rozovskaya *et al.*, 1982; Rozovskaya *et al.*, 1984; Wang and Hawley, 1993; Rudd *et al.*, 1994; Orlova *et al.*, 1995; Sastry and Ross, 1997; Li *et al.*, 1998; Urban *et al.*, 2001).

➤ CLIVAJE MEDIADO POR OTRAS RNASAS

Tampoco se puede descartar que la generación de moléculas abortivas y/o fragmentos de RNA, viral o celular, puedan derivar de actividades endo y exonucleolíticas asociadas con RNAsas similares a la RNasa M (Wreschner *et al.*, 1978) y otras no descriptas todavía.

En función de las disquisiciones anteriores, debemos pensar que es posible que la mayor parte de las moléculas que componen la población de cuasiespecies no pueda ser replicada ni transcripta. Sin embargo, las moléculas defectivas y/o fragmentos están presentes y pueden participar en eventos de recombinación en, al menos, dos situaciones diferentes:

- (1) En el caso en que en la población haya alguna molécula tipo *wild type*, las defectivas pueden recombinar durante el ciclo replicativo viral, con la participación activa de la polimerasa viral.
- (2) En el caso en que en la población no haya ninguna molécula tipo *wild type*, las únicas posibilidades para que se generen revertantes radican en la recombinación entre moléculas defectivas (o una molécula defectiva y un RNA celular), mediada por alguna proteína celular o por un mecanismo no proteico, por ej. la transesterificación.

La recombinación en sí misma

La recombinación *in vivo* de RNA:

- (1) Puede ocurrir cuando una de las moléculas o ambas moléculas de RNA no son capaces de replicar (Kirkegaard and Baltimore, 1986; Munishkin *et al.*, 1988; Munishkin *et al.*, 1991; Moody *et al.*, 1994; Hajjou *et al.*, 1996; Bos *et al.*, 1996; Brian and Spaan, 1997).
- (2) Puede ser guiada por las estructuras secundarias formadas en cada molécula de RNA individual, o entre las dos moléculas de RNA (Kuge *et al.*, 1986; Romanova *et al.*, 1986; Nagy and

Bujarski, 1993; Carpenter *et al.*, 1995; Rowe *et al.*, 1997).

- (3) Puede ocurrir predominantemente en regiones de los RNAs que carecen de estructura secundaria (Romanova *et al.*, 1986; Tolskaya *et al.*, 1987; Carpenter *et al.*, 1995).
- (4) Es promovida por la hibridación local de las moléculas de RNA en proceso de recombinación (Dzianott *et al.*, 1995).
- (5) Presenta una dependencia de la concentración de RNA similar a la recombinación *in vitro* (Jarvis and Kirkegaard, 1992).

Actualmente existen varios modelos de recombinación de RNA: el modelo de *template-switching* o *copy choice* (soportado por la mayor parte de los trabajos publicados sobre distintos virus de plantas y animales), el modelo de *dynamic copy choice*, inducido por ruptura, (Hajjar and Linial, 1993; Zhang and Sapp, 1999; Svarovskaia *et al.*, 2000; Hwang *et al.*, 2000) y el modelo de ruptura y ligación (similar al mecanismo de una ribozima; Chetverin *et al.*, 1997, 1999; Chetverina *et al.*, 1999; Gmyl *et al.*, 1999). Un mecanismo adicional postulado es el de *primer alignment and extension*. Este mecanismo implica la hibridación local entre una región de una hebra de RNA de polaridad positiva y una hebra de RNA de polaridad negativa, que sirve como molde para la continuidad de la síntesis de la nueva hebra positiva. De esta manera, se ha demostrado el rescate de clones inactivados de poliovirus (Pierangeli *et al.*, 1999).

Ninguna de las 5 propiedades básicas mencionadas anteriormente permite distinguir, con total certeza, cual de estos mecanismos ha generado un determinado recombinante. Es factible que muchas de las recombinaciones homólogas ocurran mediante el mecanismo de *copy choice*, y su frecuencia sea un parámetro específico del virus en cuestión, posiblemente asociado con alguna de las propiedades de la replicasa, tal como la procesividad y/o la

presencia de una actividad de RNA helicasa asociada. Al mismo tiempo, la recombinación no homóloga puede ocurrir, principalmente mediante un mecanismo similar al *splicing*, probablemente no dependiente de una maquinaria proteica, y su frecuencia sería independiente del virus en cuestión, ya que estaría determinada por las propiedades generales del RNA, tales como la estructura secundaria y/o terciaria (Chetverin *et al.*, 1997).

Por otra parte, tradicionalmente, la recombinación de RNA ha sido clasificada como homóloga y no homóloga, sin tener en cuenta el mecanismo involucrado. Sin embargo, existe mucha confusión acerca del uso de estas denominaciones. Por ello, Nagy and Simon (1997) han propuesto una reclasificación de estas denominaciones en tres clases: **Clase 1** recombinación donde la similitud de secuencia nucleotídica es esencial (*similarity-essential recombination*), **Clase 2** recombinación donde la similitud no es esencial (*similarity-nonessential recombination*) y **Clase 3** donde la recombinación es asistida por la similitud (*similarity-assisted recombination*). No obstante esto, en la literatura actual todavía se sigue utilizando la terminología de recombinación homóloga y no homóloga, con criterios relativamente ambiguos.

Actualmente, en muchos virus con genoma a RNA, virus de plantas, de animales y bacteriofagos, ha sido verificada en forma experimental la ocurrencia de eventos de recombinación (Romanova *et al.*, 1986; Tolskaya *et al.*, 1987; Luytjes *et al.*, 1988; Pilipenko *et al.*, 1989a, 1989b, 1990; Weiss and Schlesinger, 1991; Lai, 1992; Li and Ball, 1993; Hajjar and Linial, 1993; Raju *et al.*, 1995; Nagy and Bujarsky, 1995, 1996, 1997; Bujarsky and Nagy, 1996; Hajjou *et al.*, 1996; Olsthoorn and van Duin, 1996; Lai, 1996; Ball, 1997; Chetverin *et al.*, 1997; Hill *et al.*, 1997; Figlerowicz *et al.*, 1997; Klovins *et al.*, 1997; Nagy and Simon, 1997; Desport *et al.*, 1998; Miller and Koev, 1998; Nagy and Bujarsky,

1998; Nagy *et al.*, 1998, 1999a, 1999b; Aaziz *et al.*, 1999; Chetverina *et al.*, 1999; Gmyl *et al.*, 1999; Pierangeli *et al.*, 1999; Negroni and Bue, 1999; Wu *et al.*, 1999; Zhang and Sapp, 1999; Bruyere *et al.*, 2000; Figlerowicz, 2000; Molenkamp *et al.*, 2000; Dzianott *et al.*, 2001; Hwang *et al.*, 2001).

Además, muchos estudios detallados, realizados sobre BMV (*brome mosaic virus*), coronavirus, poliovirus, SIN (*Sindbis virus*), tombusvirus, TCV (*turnip crinkle virus*), etc., han explorado los requerimientos estructurales y de secuencia, y la maquinaria proteica, involucrados en la recombinación de RNA (Cascone *et al.*, 1993; Carpenter *et al.*, 1995; Nagy and Bujarsky, 1995, 1996, 1997; Chetverin *et al.*, 1997; Rowe *et al.*, 1997; Desport *et al.*, 1998; Nagy *et al.*, 1999a, 1999b; Simmonds and Smith, 1999; Figlerowicz, 2000; Molenkamp *et al.*, 2000; Svarovskaia *et al.*, 2000).

En casi todas las publicaciones mencionadas anteriormente, el mecanismo de recombinación básico postulado es el de *template-switching* o *copy choice*. Este mecanismo implica, al menos, tres etapas: (a) el inicio de un evento replicativo con la subsiguiente pausa de la polimerasa (por la presencia de estructuras secundarias, la depleción de algún nucleótido, etc.), (b) la selección de molde y (c) el reinicio sobre un molde diferente. Por otra parte, los determinantes que facilitan el cambio de molde son: la existencia de una hibridación local entre los RNAs sustrato, la existencia de motivos de reconocimiento de la RNA replicasa y/o la presencia de estructuras secundarias (Nagy and Simon, 1997).

La estructura secundaria y la recombinación en los arnavirus

Un análisis detallado de la secuencia nucleotídica de los RNAs S de los arnavirus revela la posibilidad de deleciones y/o inserciones en la región intergénica de varios arnavirus. Su ocurrencia puede ser explicada

por la existencia de repeticiones en tandem (perfectas o imperfectas) que posibilitan la formación de motivos de estructura secundaria conservados. Las repeticiones en tandem se encuentran en la región intergénica de los arnavirus Junín, Oliveros, Sabiá, Tacaribe y Mopeia; mientras que los arnavirus Lassa, LCM y Pichindé carecen de la segunda copia.

La existencia de conformaciones alternativas en la estructura secundaria de las regiones intergénicas, puede generar la formación de complejos homo- o heterodúplex por apareamiento entre regiones intergénicas, idénticas o similares, de genomas *wild type* de diferentes arnavirus.

Por lo tanto, las secuencias intergénicas constituyen regiones recombinogénicas. Esta particularidad podría jugar un rol importante, no sólo en la recombinación entre miembros de la familia *arnaviridae*, sino también en la posible interacción entre genomas de RNA de virus diferentes.

Dos ejemplos particulares de la generación de moléculas de RNAs S híbridas diferentes, pueden observarse en la **Figura 12.39**. Además de lo mostrado, la generación de moléculas de RNA S híbridas interespecíficas puede pensarse en cualquiera de las dos posibilidades, mitad 3' del RNA S del virus Junín y mitad 5' del RNA S del virus Sabiá, o mitad 3' del RNA S del virus Sabiá y mitad 5' del RNA S del virus Junín. El resultado final simplemente varía en la combinación de proteínas: N del virus Junín y GPC del virus Sabiá o N del virus Sabiá y GPC del virus Junín. Es probable que las diferentes combinaciones puedan generar cambios en el hospedador y/o el tropismo del virus recombinante.

De manera equivalente a la hipótesis presentada en la **Figura 12.39**, podría especularse sobre la generación de RNAs S recombinantes entre otros pares de arnavirus, que posean la región 3' de un arnavirus y la región 5' de otro (o viceversa), derivadas de

eventos de coinfección.

Por otra parte, la existencia de repeticiones imperfectas, y de estructuras secundarias tipo *hairpin loop*, dentro del genoma de un mismo arnavirus, o entre genomas de distintos arnavirus, se extiende más allá de las regiones no codificantes intergénicas. En la **Figura 12.40** se muestran, a manera de ejemplo, todas las posibles regiones de complementariedad entre el RNA S del virus Junín-MC2 en polaridad viral *versus* los RNAs S de los arnavirus del Viejo Mundo (**Figura 12.40.A**) y del Nuevo Mundo (**Figura 12.40.B**), en la misma polaridad; cualquiera de estas posiciones es candidata para un evento de recombinación mediante los mecanismos de *copy choice* o transesterificación, con la única condición diferencial de que el segundo mecanismo requiere que al menos una de las hebras involucradas sea abortiva y presente un hidroxilo 3' libre. En la **Figura 12.41** se muestran, a manera de ejemplo, todas las posibles regiones de complementariedad entre el RNA S del virus Junín-MC2 en polaridad viral *versus* los RNAs S de los arnavirus del Viejo Mundo (**Figura 12.41.A**) y del Nuevo Mundo (**Figura 12.41.B**), en polaridad viral complementaria; cualquiera de estas posiciones es candidata para un evento de recombinación mediante el mecanismo de *primer alignment and extension*, con la única condición de que al menos una de las hebras involucradas sea abortiva y presente un hidroxilo 3' libre.

Si bien es cierto que en las regiones codificantes, existe un limitante adicional, que es la necesidad de un marco de lectura continuo, para generar un virus recombinante potencialmente viable; cuando se analiza la existencia de repeticiones, perfectas o imperfectas, asociadas con la producción de este tipo de recombinantes, es factible encontrar muchas alternativas de recombinación; las cuales podrían derivar en la aparición de nuevos genotipos (y fenotipos) virales.

En la **Figura 12.42** se muestra una de

las tantas alternativas posibles de emergencia de un nuevo arnavirus, mediante recombinación en las regiones complementarias a secuencias codificantes para las proteínas N de los arnavirus Junín-MC2 y Oliveros. En este caso teórico particular, el mecanismo de *copy choice* permitiría explicar la generación de un RNA S viral complementario que codifica para los primeros 159 aminoácidos de la proteína N *wild type* de la cepa MC2 del virus Junín y para los últimos 357 residuos de la proteína N *wild type* del virus Oliveros. Esencialmente, la proteína N recombinante podría tener las mismas propiedades que las *wild type* originales.

Por otra parte, también se podrían generar distintos recombinantes entre moléculas de RNA de partículas defectivas (no replicables por sí mismas) de un mismo arnavirus, o entre moléculas de RNA de partículas defectivas de diferentes arnavirus, que coinfecten al mismo hospedador circunstancialmente. El mecanismo involucrado en este caso podría ser también el de *copy choice*, así como el de ruptura y ligación (transesterificación).

En las Figuras 12.43, 12.44, 12.45 y 12.46, se muestran ejemplos de potenciales recombinaciones entre los RNAs S de los arnavirus Junín MC2 y Oliveros, mediante el mecanismo de transesterificación.

En la Figura 12.43, el mecanismo de transesterificación daría como resultado un RNA S recombinante de 3988 nt. de longitud, derivado de la unión de la región 3' no codificante del RNA S y parte de la región complementaria al marco de lectura para la proteína N del arnavirus Oliveros, con la región 3' no codificante del arnavirus Junín MC2. Esta recombinación, probablemente, daría lugar a un virus no viable o con fenotipo Junín, ya que el segmento de marco de lectura codificante para N de Oliveros no queda en la misma fase que el correspondiente marco codificante para la proteína N de Junín MC2. También es posible que, si este genoma queda en la población de cuasiespecies, eventos

posteriores de mutagénesis (transiciones y transversiones) podrían permitir el rescate de un RNA S arnaviral viable.

En la Figura 12.44, el mecanismo de transesterificación daría como resultado un RNA S recombinante de 4076 nt. de longitud, derivado de la unión de parte de la región complementaria al marco de lectura para la proteína N del arnavirus Oliveros, con la región complementaria al marco de lectura para la proteína N del arnavirus Junín-MC2. Esta recombinación, probablemente, daría lugar a un virus viable, con fenotipo de tipo emergente, ya que el segmento de marco de lectura codificante para N de Oliveros queda en la misma fase que el correspondiente marco codificante para la proteína N de Junín MC2, permitiendo la traducción de una proteína N híbrida.

Las Figuras 12.45 y 12.46 muestran otras dos alternativas de generación de arnavirus emergentes, con proteínas N recombinantes, derivadas de la unión de diferentes zonas de la región codificante para la proteína N del arnavirus Oliveros y la misma zona de la región codificante para la proteína N del arnavirus Junín-MC2.

Durante el año 2001 se han publicado las caracterizaciones moleculares de los RNAs S de otros 3 arnavirus del Nuevo Mundo, Allpahuayo (3343 nt., Moncayo *et al.*, 2001), Pirital (3393 nt., Charrel *et al.*, 2001a) y Whitewater Arroyo (3316 nt., Charrel *et al.*, 2001b). Un aspecto llamativo de estos RNAs S es su tamaño, más pequeño que el promedio de los anteriormente descritos. Por otra parte, en Pirital se ha encontrado una gran diversidad genética entre diferentes aislamientos (Weaver *et al.*, 2001); mientras que en Whitewater Arroyo, un análisis filogenético revela que el gen N deriva del linaje A y el gen GPC deriva del linaje B de un arnavirus del complejo Tacaribe (Charrel *et al.*, 2001b). En las Figuras 12.47 a 12.52 se muestran diferentes alternativas que permitirían la aparición de un recombinante con GPC derivada de un linaje y

N derivada de otro. En todas las figuras mencionadas, la región recombinogénica involucrada en ambos virus es la región intergénica.

Desde ya que todas las hipótesis anteriores son sólo eso y deberían ser confirmadas experimentalmente, diseñando los experimentos *ad hoc* para ello.

Un caso particular, el clon viral 6₂ de la cepa MC2 del virus Junín

En 1996, cuando se envió a publicar la secuencia de la GPC de la cepa Candid #1 del virus Junín (Ghiringhelli *et al.*, 1997) la única secuencia de referencia disponible era la de la cepa MC2 del mismo virus. Una de las interpretaciones posibles sobre el origen de las diferencias entre ambas cepas es la presentada en el **Capítulo 13 (Tabla 13.a y Figura 13.1)**, asociando las diferencias con la atenuación de Candid #1. Posteriormente, esto fue descartado en base a las secuencias de los antecesores filogenéticos (Albariño *et al.*, 1997; Albariño, 1997). No obstante, se mantuvo la presunción de que las diferencias estaban relacionadas con las diferencias de origen de los aislamientos; la cepa MC2 fue aislada de un roedor mientras que la cepa Candid #1 deriva en su origen de un aislamiento a partir de un caso humano terminal.

Sin embargo, en 2000, García *et al.* publicaron los resultados obtenidos en un estudio de diversidad poblacional utilizando 39 aislamientos (26 aislamientos de humanos y 13 de roedores). El análisis conjunto de todas las secuencias del virus Junín existentes en el GenBank actualmente (**Figura 12.53**), revela que el clon viral 6₂ de la cepa MC2 es único en cuanto a las secuencias que presenta en las regiones identificadas como **a**, **b**, **c** y **d** en la **Figura 13.1** y en la **Tabla 13.a**, codificantes para la zona C-terminal del péptido señal y N-terminal de la glicoproteína G1 (ver **Figura 12.53** y **Apéndice 12**).

Además, en 1996, en el congreso argentino de virología, Padula *et al.* presentaron

un póster donde mostraron las secuencias N-terminales de las glicoproteínas G1 y G2 de la cepa MC2 del virus Junín, y la secuencia aminoacídica deducida a partir de un fragmento de PCR que comprendía desde el AUG del precursor de las glicoproteínas hasta el primer tercio de la glicoproteína G1. Tanto la secuencia N-terminal de la glicoproteína G1, como la secuencia deducida a partir del fragmento de PCR resultaron iguales a la de la cepa Candid #1 y a las del resto de los aislamientos secuenciados por García *et al.* (2000). La diferencia principal entre el trabajo de Padula *et al.* (1996) y el nuestro (Ghiringhelli *et al.*, 1991) es el material de partida. El RNA S de la cepa MC2 clonado y secuenciado en el IBBM deriva de un proceso de clonado viral y aislamiento por plaqueo (ver **Capítulo 5, Figura 5.1**), mientras que en el trabajo de Padula *et al.* (1996) partieron del inóculo inicial del aislamiento no clonado, derivado del Instituto Malbrán.

En función de la información actual, se puede hipotetizar acerca de que el clon viral 6₂ de la cepa MC2, clonado y secuenciado, es un recombinante.

En función de esto, es necesario analizar nuevamente las diferencias encontradas en la región codificante para el precursor de las glicoproteínas (GPC) en la cepa MC2.

Asumiendo como secuencia genómica ancestral la de Candid #1, que coincide totalmente con la mayoritaria, se pueden considerar varias interpretaciones para las diferencias encontradas.

Interpretación 1

Teniendo en cuenta exclusivamente la tasa de error de la RNA polimerasa viral, se requieren 9 eventos de delección (7 puntuales, 1 bloque de dos residuos y 1 bloque de seis nucleótidos) y 1 transversión para convertir una secuencia genómica del tipo de Candid #1 en la del clon 6₂ de MC2, en las regiones marcadas como **a**, **b**, y **c** en la **Figura 12.54**.

Además, se requieren 3 inserciones

puntuales en la región **d** (**Figura 12.54**). Dado que en este último caso todas las inserciones implican la incorporación de un residuo idéntico al anterior, los cambios de marco de lectura que han transformado la secuencia del tipo de Candid #1 en la de MC2 se pueden explicar mediante el mecanismo de detención, retroceso y reinicio de la polimerasa, característica de las RNAs polimerasas ya descrita en muchas publicaciones.

Sin embargo, aún cuando todo lo anterior es factible, y dejando de lado cualquier consideración temporal, subyacen algunos cuestionamientos y requisitos claves.

Si las deleciones señaladas en la **Figura 12.54** (regiones **a**, **b** y **c**) ocurrieron en forma independiente, una por una:

⇒ ¿Cómo “subsistió” la molécula de RNA S defectiva en los sucesivos ciclos replicativos?

Una respuesta probable es que se mantuvo en la población de cuasiespecies hasta la conversión final (Domingo, 1992; Domingo *et al.*, 1996; Ruiz-Jarabo *et al.*, 2000; Domingo, 2002)

Por otra parte, si las deleciones ocurrieron en conjuntos tales que se mantuvo el requisito de un marco de lectura abierto:

⇒ ¿Por qué los errores “aleatorios” de la replicasa viral se concentran en esta región, mientras que en el resto de la GPC se observan sólo cambios puntuales, la mayor parte de los cuales son sólo transiciones o transversiones?

Hasta el momento no hay más respuesta a estos interrogantes que la percepción de que es un conjunto de eventos poco probables.

Interpretación 2

Teniendo en cuenta las estimaciones de King (1988) sobre la frecuencia de eventos de recombinación en poliovirus y la teoría de las cuasiespecies, se puede pensar en una alternativa para los cambios señalados en la

Figura 12.54 (regiones **a**, **b** y **c**).

El punto de partida es considerar la existencia en la población viral de una molécula defectiva por deleción de las regiones **a+b** o **a+b+c** que puede ser replicada y transcripta (los extremos 5' y 3' del RNA S genómico se conservan intactos), pero que no generaría una GPC funcional.

A partir de esta molécula de RNA S se puede haber recuperado un genoma funcional mutado mediante la inserción de fragmentos de RNA derivados de mRNAs celulares (**Figura 12.54**).

Esta hipótesis permite explicar los cambios en las regiones **a**, **b** y **c** mediante la recombinación en 1 a 3 eventos independientes.

Luego de una búsqueda exhaustiva en el banco de datos de EST humanos y de roedor (GenBank), se detectaron múltiples fragmentos de mRNAs que cumplen las condiciones necesarias y constituyen el principal sustento de esta interpretación (**Figuras 12.55** y **12.56**).

Una consideración adicional necesaria está en relación con la gran conservación aminoacídica detectada en la región C-terminal del péptido señal de GPC (ver **Apéndice 6**, **Figura Ap6.1** y **Apéndice 12**). En todos los aislamientos del virus Junín la secuencia C-terminal es L-[AD]-L-A-G-R-S-C-X₁₋₃, mientras que cuando se extiende el análisis a todos los arnavirus, la secuencia consenso general es L-[AIDFLTV]-L-[ACS]-G-R-S-C-X₁₋₃. Esto sugiere que esta región del péptido señal resulta crucial para la función del mismo y/o es la secuencia de reconocimiento para el clivaje proteolítico en el retículo endoplásmico mediado por una peptidasa celular.

Todos los eventos de recombinación considerados en esta interpretación cumplen con la condición de mantener constante esta región del marco de lectura de GPC, derivado del genoma ancestral original o derivado del fragmento insertado.

En síntesis, cualquiera de los mecanismos de recombinación planteados en

los títulos anteriores de este capítulo permitiría explicar el/los eventos que habrían posibilitado el rescate de un genoma viral viable a partir de una molécula defectiva.

Por último, también se debe mencionar que los cambios aquí descritos pueden haber ocurrido en el reservorio natural (el roedor) o durante el proceso de clonado viral realizado en el IBBM. Experimentos adicionales con los inóculos originales e intermedios permitirían, tal vez, resolver esta cuestión.

Conclusiones generales sobre recombinación

Así como la recombinación entre RNAs genómicos y subgenómicos de un mismo virus puede generar partículas DI y/o nuevas variantes virales (Bar-Joseph *et al.*, 1997), la recombinación intertípica entre genomas funcionales y la recombinación entre partículas DI también pueden derivar en genomas híbridos funcionales (White and Morris, 1994). Las características biológicas de los recombinantes son impredecibles y, algunas veces, derivan en la generación de virus híbridos más virulentos que cualquiera de los parentales, como han demostrado Ding *et al.* (1996) en un ensayo de laboratorio. Estas pruebas replantean una tradicional pregunta acerca del uso de cepas virales atenuadas como vacunas: ¿Cuán seguras son?. Recientemente, se han publicado dos trabajos que permiten cuestionar severamente la seguridad del empleo de virus atenuados: Cuervo *et al.* (2001) publican el hallazgo de poliovirus viables excretados por individuos vacunados con cepas atenuadas generados mediante recombinación intertípica y Becher *et al.* (2001) presentan el hallazgo de un pestivirus citopatogénico letal generado mediante recombinación entre un genoma presente en una infección persistente y el genoma de una cepa vacunal.

Además, también existen publicaciones que sustentan la recombinación entre virus diferentes: Luytjes *et al.* (1988) sugieren que el mRNA 2 del *Mouse Hepatitis Virus A59* deriva de una recombinación entre coronavirus y el

virus de influenza C y Gibbs & Weiller (1999) especulan filogenéticamente acerca de la transmisión de un nanovirus (virus a DNA de plantas) a un vertebrado, transformándose en un circovirus (virus a DNA de animales) mediante la adquisición de la mitad C-terminal de la proteína Rep (*replication protein*) de un calicivirus (virus a RNA de animales).

Por otra parte, de ninguna manera se pueden descartar otras posibilidades mucho más aleatorias que las planteadas en toda la discusión precedente.

En muchos sistemas virales, los ensayos de *Northern* sobre RNA total de células infectadas, hibridados con una sonda viral genómica, revelan homología con rRNAs celulares (especialmente el rRNA 28S), aún en condiciones de alta estrictez (McClure and Perrault, 1986). Esto puede estar asociado con las respuestas celulares apoptóticas y el empleo de fragmentos de RNAs celulares para reparar moléculas virales defectivas (Carpenter and Simon, 1996) y/o facilitar la recombinación entre moléculas virales (Dzianott *et al.*, 1995). Charini *et al.* (1994) demuestran la generación de revertantes de poliovirus mediante la incorporación de un segmento de rRNA 28S celular y Becher *et al.* (1998) publican el desarrollo de citopatogenicidad en un pestivirus mediante la incorporación de un segmento de un mRNA celular híbrido codificante para la ubiquitina y para la proteína ribosomal S27a (NH₂-ubiquitina-S27a-COOH).

Estos hallazgos plantean la posibilidad de una evolución viral mediada, en parte, por rearrreglos en el propio genoma viral y, en parte, por la adquisición de segmentos de RNAs o DNAs celulares (Becker, 2000). Algunos de los ejemplos más llamativos son aquellos donde se ha demostrado la recuperación de viabilidad mediante la reconstrucción de estructuras secundarias cruciales en la regulación de la replicación, transcripción o traducción (Klaver and Berkhoult, 1994; Olsthoorn and van Duin, 1996; Klovins *et al.*, 1997).

Recombinación entre los RNAs S virales del virus Junín MC2 y del virus Sabiá

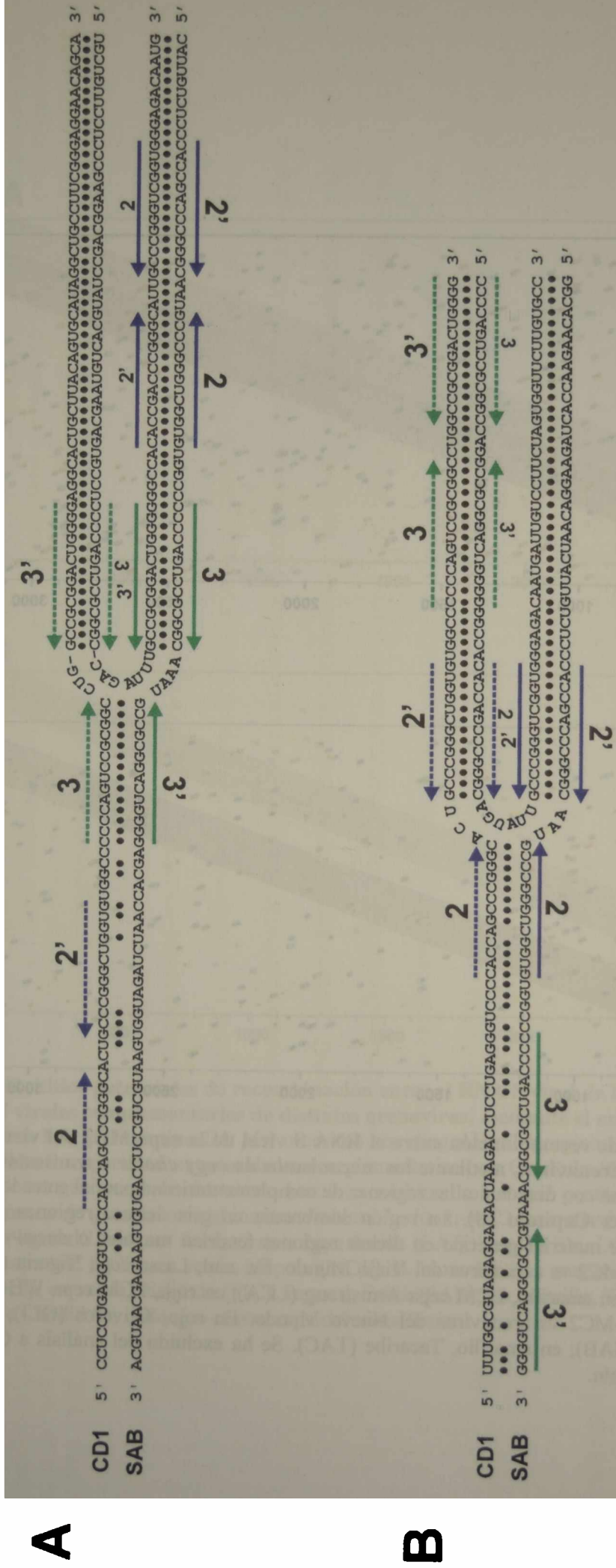


Figura 12.39. Generación de un recombinante interespecífico entre el virus Junín-MC2 y el virus Sabiá. Las flechas y los números corresponden a las estructuras tipo *hairpin loop* indicadas en la Figura 12.26. A. Secuencia nucleotídica de una de las alternativas de heteroduplex híbrido parcial entre las dos cadenas de RNA S viral y secuencia nucleotídica de la molécula recombinante. B. Secuencia nucleotídica de otra alternativa de heteroduplex híbrido parcial entre las dos cadenas de RNA S viral y secuencia nucleotídica de la molécula recombinante.

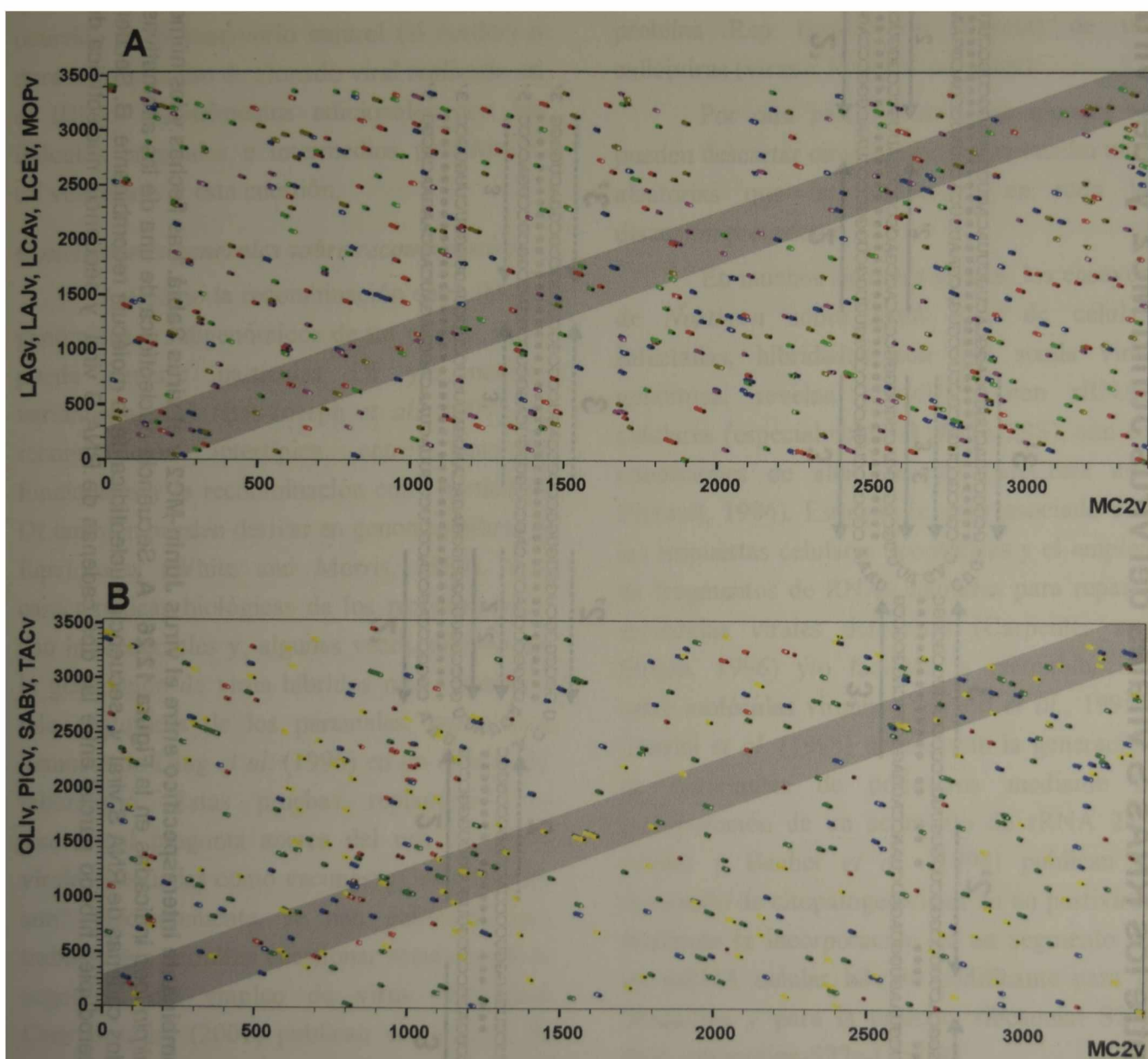


Figura 12.40. Sitios potenciales de recombinación entre el RNA S viral de la cepa MC2 del virus Junín y los RNAs S virales de distintos arnavirus, mediante los mecanismos de *copy choice* o transesterificación. Los gráficos tipo *Dot plot* muestran con diagonales las regiones de complementariedad parcial entre los RNAs S respectivos en polaridad viral (ver **Capítulo 18**). La región sombreada en gris destaca regiones genómicas equivalentes. Los intercambios de material genético en dichas regiones tendrían mayores chances de generar recombinantes viables. **A.** Junín-MC2 vs arnavirus del Viejo Mundo. En azul, Lassa cepa Nigeria (LAG); en fucsia, Lassa cepa Josiah (LAJ); en amarillo, LCM cepa Armstrong (LCA); en rojo, LCM cepa WE (LCE); en verde, Mopeia (MOP). **B.** Junín-MC2 vs arnavirus del Nuevo Mundo. En rojo, Oliveros (OLI); en verde, Pichindé (PIC); en azul, Sabiá (SAB); en amarillo, Tacaribe (TAC). Se ha excluido del análisis a Candid #1, dado que es otra cepa del virus Junín.

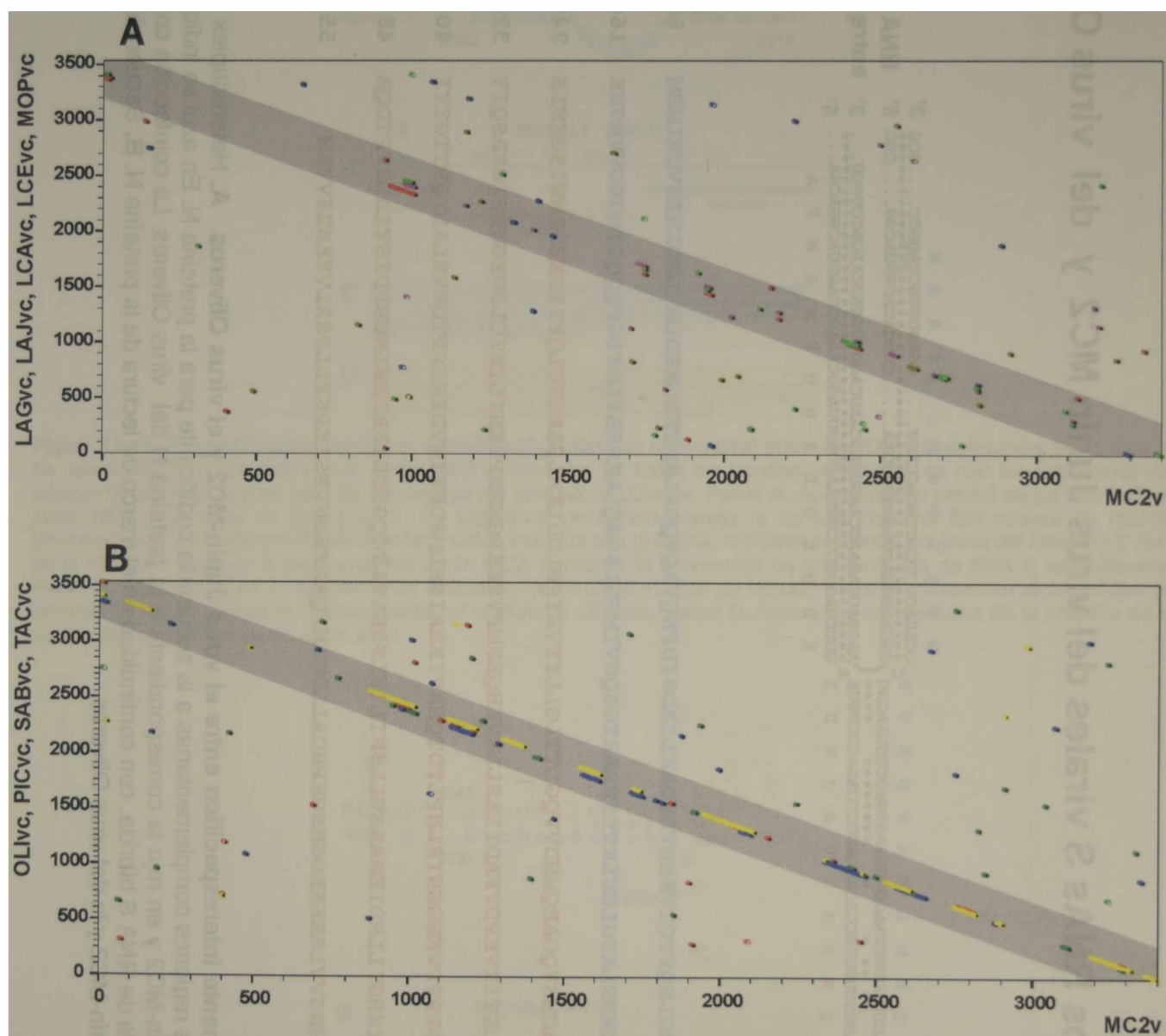


Figura 12.41. Sitios potenciales de recombinación entre el RNA S viral de la cepa MC2 del virus Junín y los RNAs S virales complementarios de distintos arnavirus, mediante el mecanismo de *primer alignment and extension*. Los gráficos tipo *Dot plot* muestran con diagonales las regiones de complementariedad parcial entre los RNAs S respectivos en polaridades complementarias (ver **Capítulo 18). La región sombreada en gris destaca regiones genómicas equivalentes. Los intercambios de material genético en dichas regiones tendrían mayores chances de generar recombinantes viables. **A.** Junín-MC2 vs arnavirus del Viejo Mundo. En azul, Lassa cepa Nigeria (LAG); en fucsia, Lassa cepa Josiah (LAJ); en amarillo, LCM cepa Armstrong (LCA); en rojo, LCM cepa WE (LCE); en verde, Mopeia (MOP). **B.** Junín-MC2 vs arnavirus del Nuevo Mundo. En rojo, Oliveros (OLI); en verde, Pichindé (PIC); en azul, Sabiá (SAB); en amarillo, Tacaribe (TAC). Se ha excluido del análisis a Candid #1, dado que es otra cepa del virus Junín.**

Recombinación entre los RNAs S virales del virus Junín MC2 y del virus Oliveros

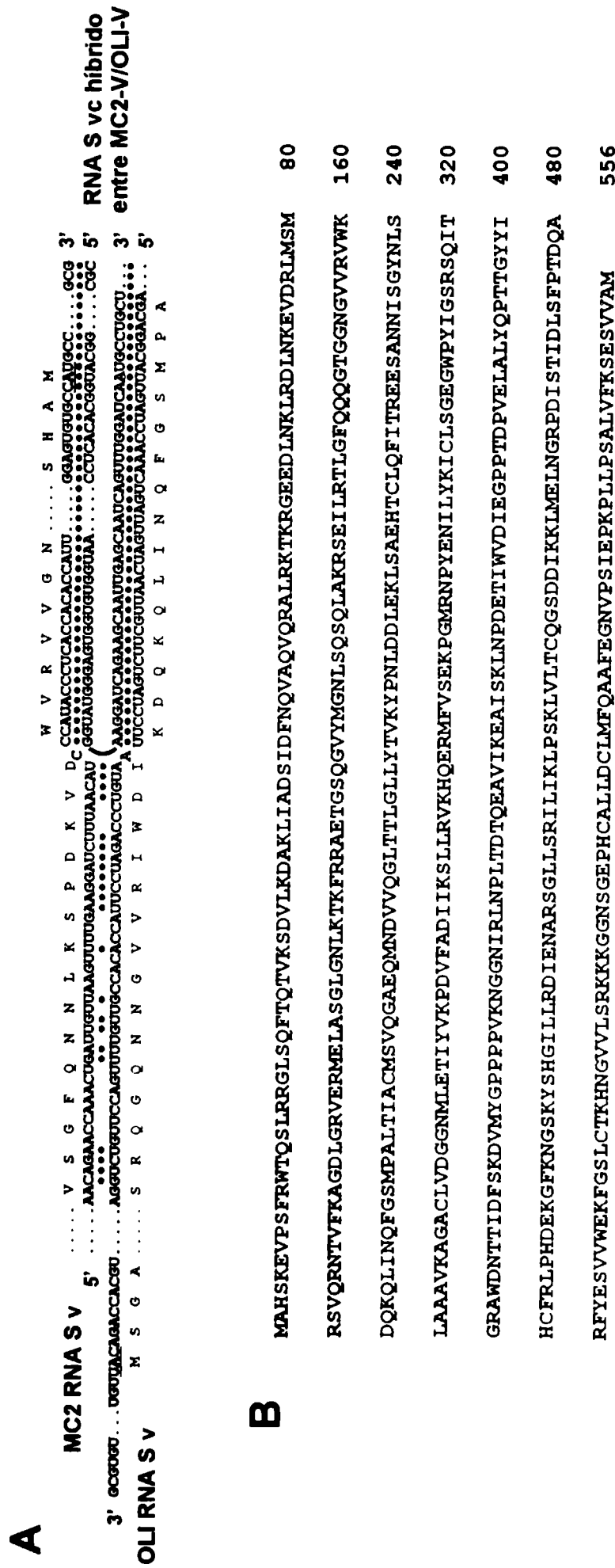


Figura 12.42. Generación de un recombinante interespecífico entre el virus Junín-MC2 y el virus Oliveros. A. Heteroduplex híbrido parcial entre las dos cadenas de RNA S viral en las regiones complementarias a la secuencia codificante para la proteína N. En azul se indica la secuencia aminoacídica de la proteína N del virus Junín-MC2 y en rojo la correspondiente a la proteína N del virus Oliveros. La coinfección con los dos virus podría permitir la generación de una molécula de RNA S híbrida, con continuidad en el marco de lectura de la proteína N. B. Secuencia aminoacídica de la proteína N híbrida entre la del virus Junín-MC2 y la del virus Oliveros

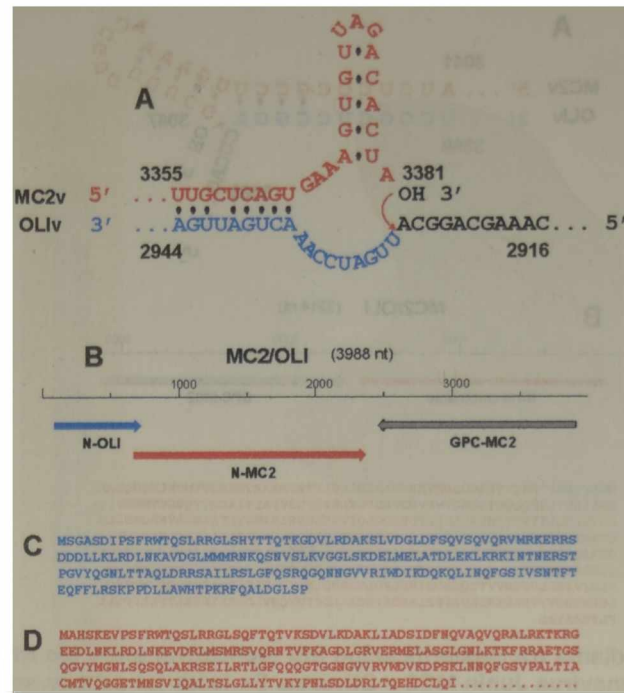


Figura 12.43. Recombinación mediante transesterificación (no replicativa) entre los RNAs S defectivos (al menos el de Junín MC2) de los arnavirus Junín MC2 y Oliveros. En todos los paneles se indican, en rojo las secuencias del arnavirus Junín MC2 y en azul las secuencias del arnavirus Oliveros. **Panel A.** Apareamiento parcial de los RNAs S de Junín MC2 y Oliveros en la región 3'. La estructura secundaria semeja la de una ribozima tipo cabeza de martillo (*hammerhead*). La transesterificación en la posición indicada con la flecha, mediante el ataque por parte del hidroxilo 3' libre de la molécula de RNA S del arnavirus Junín MC2, permitiría la generación de una molécula de RNA S recombinante. **Panel B.** Mapa del RNA S recombinante, indicando los distintos marcos de lectura. **Panel C.** Secuencia aminoacídica N-terminal de la proteína de la nucleocápside del arnavirus Oliveros. **Panel D.** Secuencia aminoacídica de la proteína de la nucleocápside del arnavirus Junín MC2.

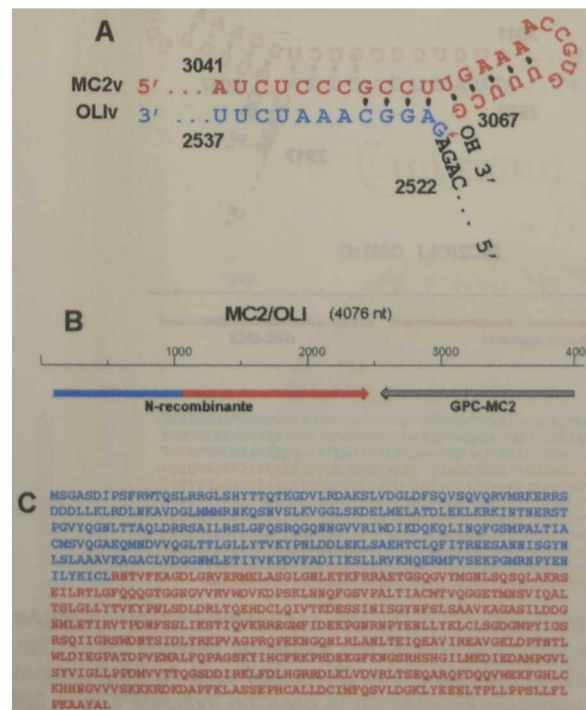


Figura 12.44. Recombinación mediante transesterificación (no replicativa) entre los RNAs S defectivos (al menos el de Junín MC2) de los arnavirus Junín MC2 y Oliveros. En todos los paneles se indican, en rojo las secuencias del arnavirus Junín MC2 y en azul las secuencias del arnavirus Oliveros. **Panel A.** Apareamiento parcial de los RNAs S de Junín MC2 y Oliveros en la región 3'. La transesterificación en la posición indicada con la flecha, mediante el ataque por parte del hidroxilo 3' libre de la molécula de RNA S del arnavirus Junín MC2, permitiría la generación de una molécula de RNA S recombinante. **Panel B.** Mapa del RNA S recombinante, indicando los distintos marcos de lectura. **Panel C.** Secuencia aminoacídica de la proteína de la nucleocápside recombinante entre la del arnavirus Oliveros y la del arnavirus Junín MC2.

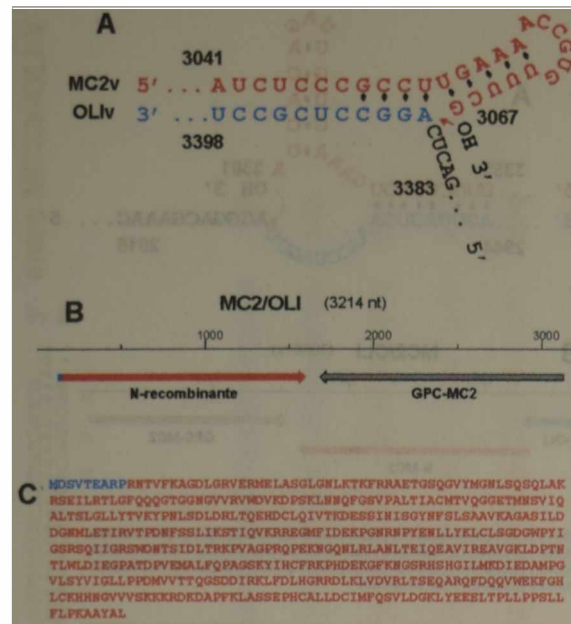


Figura 12.45. Recombinación mediante transesterificación (no replicativa) entre los RNAs S defectivos (al menos el de Junin MC2) RNAs S de los arnavirus Junin MC2 y Oliveros. En todos los paneles se indican, en rojo las secuencias del arnavirus Junin MC2 y en azul las secuencias del arnavirus Oliveros. **Panel A.** Apareamiento parcial de los RNAs S de Junin MC2 y Oliveros en la región 3'. La transesterificación en la posición indicada con la flecha, mediante el ataque por parte del hidroxilo 3' libre de la molécula de RNA S del arnavirus Junin MC2, permitiría la generación de una molécula de RNA S recombinante. **Panel B.** Mapa del RNA S recombinante, indicando los distintos marcos de lectura. **Panel C.** Secuencia aminoacídica de la proteína de la nucleocápside recombinante entre la del arnavirus Oliveros y la del arnavirus Junin MC2.

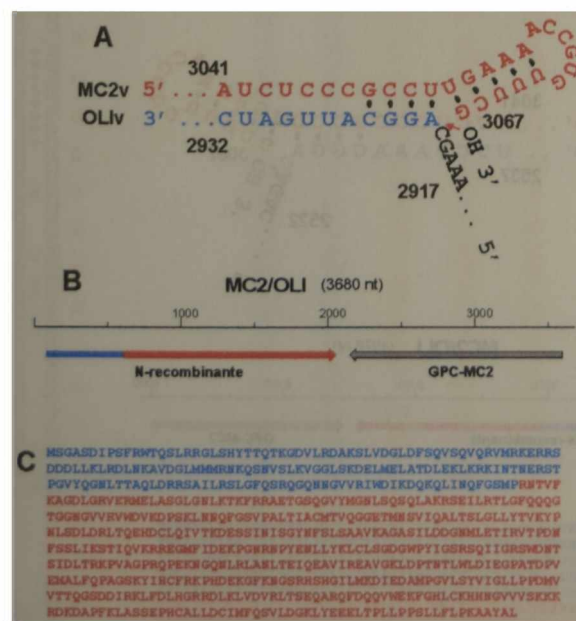


Figura 12.46. Recombinación mediante transesterificación (no replicativa) entre los RNAs S defectivos (al menos el de Junin MC2) de los arnavirus Junin MC2 y Oliveros. En todos los paneles se indican, en rojo las secuencias del arnavirus Junin MC2 y en azul las secuencias del arnavirus Oliveros. **Panel A.** Apareamiento parcial de los RNAs S de Junin MC2 y Oliveros en la región 3'. La transesterificación en la posición indicada con la flecha, mediante el ataque por parte del hidroxilo 3' libre de la molécula de RNA S del arnavirus Junin MC2, permitiría la generación de una molécula de RNA S recombinante. **Panel B.** Mapa del RNA S recombinante, indicando los distintos marcos de lectura. **Panel C.** Secuencia aminoacídica de la proteína de la nucleocápside recombinante entre la del arnavirus Oliveros y la del arnavirus Junin MC2.

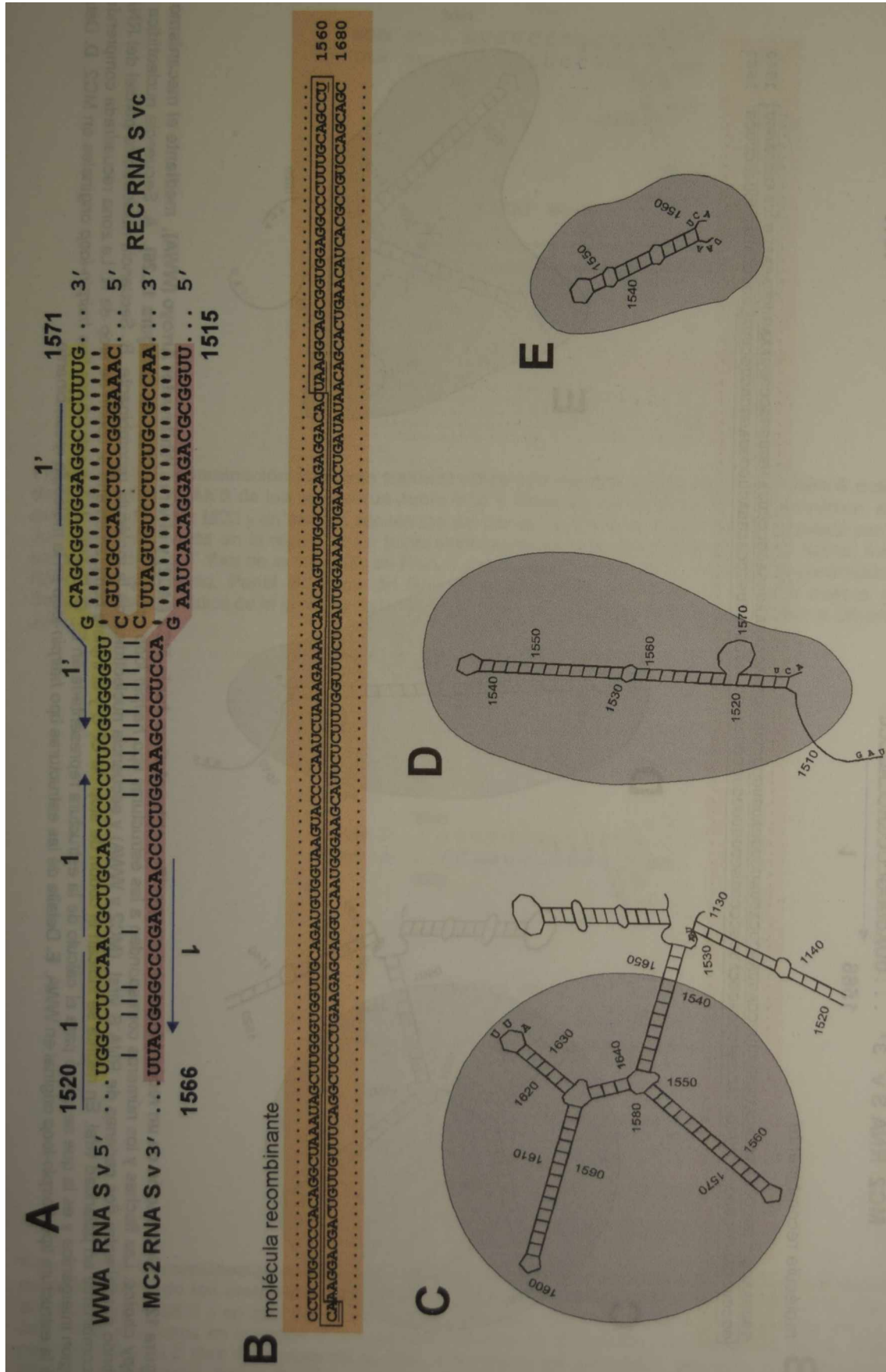


Figura 12.48. Generación de un recombinante Interes específico entre el virus Junin-MC2 (MC2) y el virus Whitewater Arroyo (WWA), mediante el mecanismo de copy choice. Las flechas y los números corresponden a las estructuras tipo hairpin loop intergénicas (para MC2, ver la Figura 12.26). A. Secuencia nucleotídica del híbrido parcial entre dos cadenas de RNA S viral (MC2 y WWA) y secuencia nucleotídica de la molécula recombinante. B. Secuencia nucleotídica parcial del RNA S recombinante en polaridad viral. En negrita se indican el codón de stop de GPC y el triplete complementario al codón de stop de N. La zona recuadrada comprende la región intergénica y es la que se utilizó para el cálculo de la estructura representada en E. C. Detalle de las estructuras tipo hairpin-loop originales en MC2. D. Detalle de la estructura tipo hairpin-loop original en WWA. E. Detalle de las estructuras tipo hairpin-loop en la molécula recombinante.

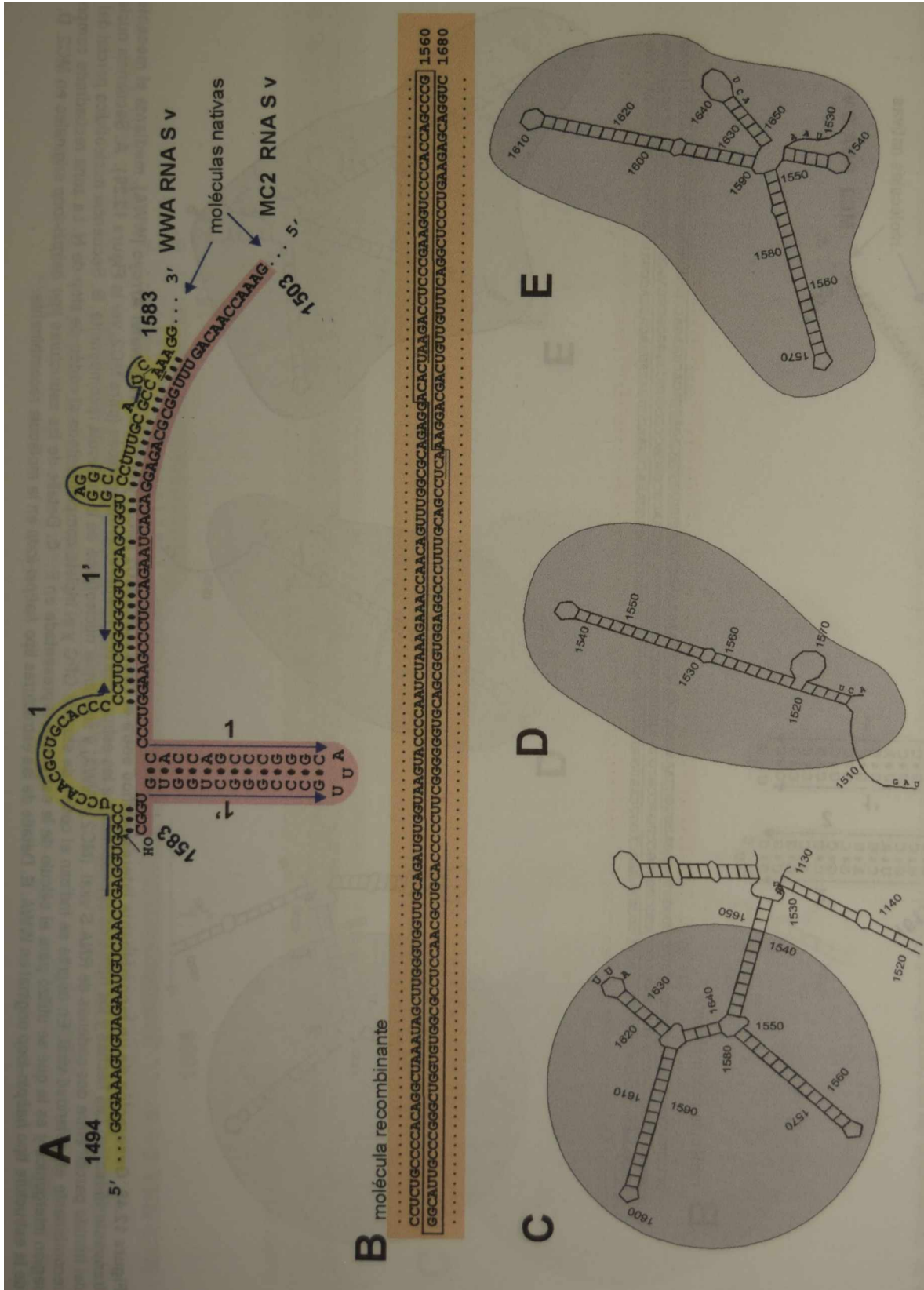


Figura 12.50. Generación de un recombinante interespecífico entre el virus Junin-MC2 (MC2) y el virus Whitewater Arroyo (WWA), mediante el mecanismo de transesierificación. Las flechas y los números corresponden a las estructuras tipo *hairpin loop* intergénicas (para MC2, ver la Figura 12.26). A. Secuencia nucleotídica del híbrido parcial entre dos cadenas de RNA S viral (MC2 y WWA) y secuencia nucleotídica de la molécula recombinante. B. Secuencia nucleotídica parcial del RNA S recombinante en polaridad viral. En negrita se indican el codón de stop de GPC y el triplete complementario al codón de stop de N. La zona recuadrada comprende la región intergénica y es la que se utilizó para el cálculo de la estructura representada en E. C. Detalle de las estructuras tipo *hairpin-loop* originales en MC2. D. Detalle de la estructura tipo *hairpin-loop* original en WWA. E. Detalle de las estructuras tipo *hairpin-loop* en la molécula recombinante.

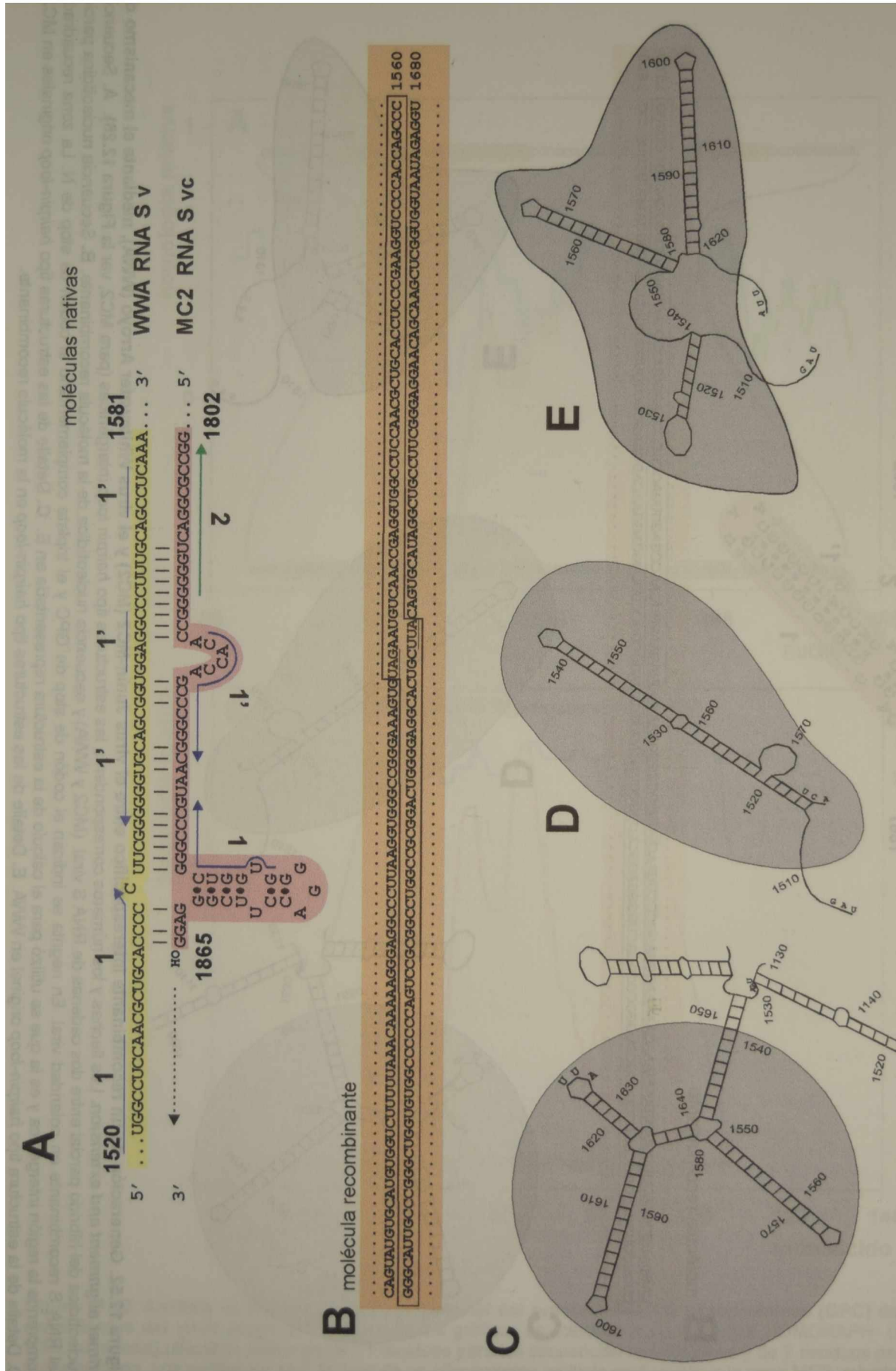


Figura 12.51. Generación de un recombinante interés específico entre el virus Junin-MC2 (MC2) y el virus Whitewater Arroyo (WWA), mediante el mecanismo de primer alignment and extension. Las flechas y los números corresponden a las estructuras tipo hairpin loop intergénicas (para MC2, ver la Figura 12.26). A. Secuencia nucleotídica del híbrido parcial entre dos cadenas de RNA S viral (MC2 y WWA) y secuencia nucleotídica de la molécula recombinante. B. Secuencia nucleotídica parcial del RNA S recombinante en polaridad viral. En negrita se indican el codón de stop de GPC y el triplete complementario al codón de stop de N. La zona recuadrada comprende la región intergénica y es la que se utilizó para el cálculo de la estructura representada en E. C. Detalle de las estructuras tipo hairpin-loop originales en MC2. D. Detalle de la estructura tipo hairpin-loop original en WWA. E. Detalle de las estructuras tipo hairpin-loop en la molécula recombinante.

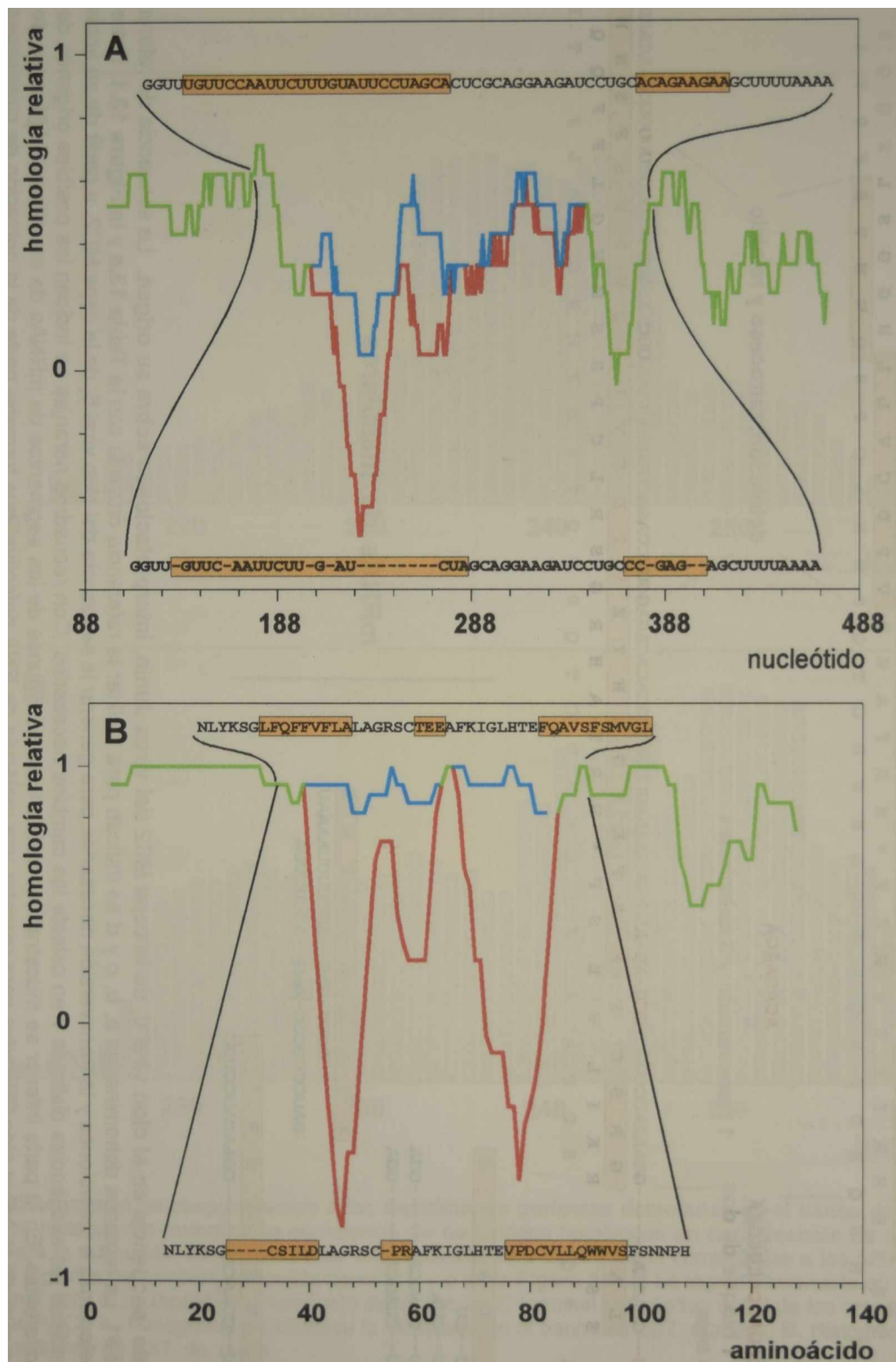


Figura 12.53. Análisis de homología del tercio inicial del precursor de las glicoproteínas (GPC) de los diferentes aislamientos del virus Junín. Los datos para los gráficos se calcularon con la rutina HOMGRAPH (P. D. Ghiringhelli, 2002, esta tesis) utilizando ventanas de 21 residuos para las secuencias nucleotídicas y de 7 residuos para las secuencias aminoácidas. Los perfiles en azul derivan de un alineamiento múltiple excluyendo a la cepa MC2, mientras que el perfil en rojo lo incluye. Las zonas verdes corresponden a regiones donde ambos perfiles son idénticos. Las secuencias ubicadas arriba de los gráficos corresponden a la cepa Candid #1, mientras que las ubicadas debajo corresponden a la cepa MC2 (en rojo se indican las diferencias). Con recuadros naranjas se señalan las diferencias de secuencia, que se corresponden con las disminuciones de homología relativa. **A.** Gráfico correspondiente a las secuencias nucleotídicas. **B.** Gráfico correspondiente a las secuencias aminoácidas.

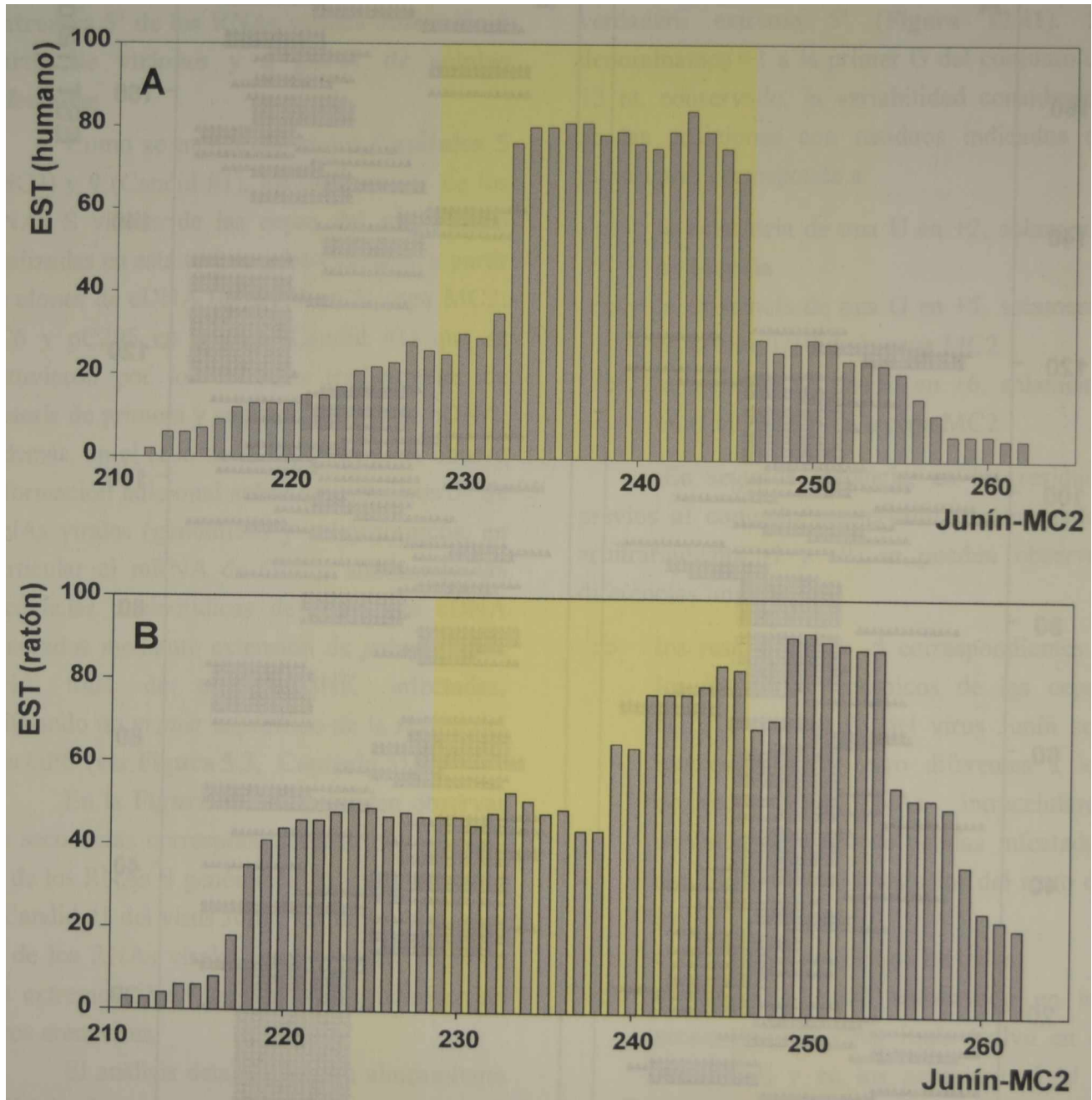


Figura 12.55. Histogramas correspondientes a las repeticiones perfectas detectadas en el banco de datos de EST (GenBank). Los histogramas muestran la abundancia de nucleótidos homólogos en cada posición de la secuencia de MC2 utilizada para la búsqueda. La secuencia de MC2 utilizada para la búsqueda corresponde a los 52 nucleótidos que comprenden las regiones de cambios marcadas como a, b y c en la Figura 12.54. La región sombreada en amarillo señala la zona codificante para el hexapéptido conservado en la porción C-terminal del péptido señal de los arnavirus (en MC2 y Candid #1: LAGRSC). A. Histograma derivado de la búsqueda en el banco de EST humano. B. Histograma derivado de la búsqueda en el banco de EST de ratón.

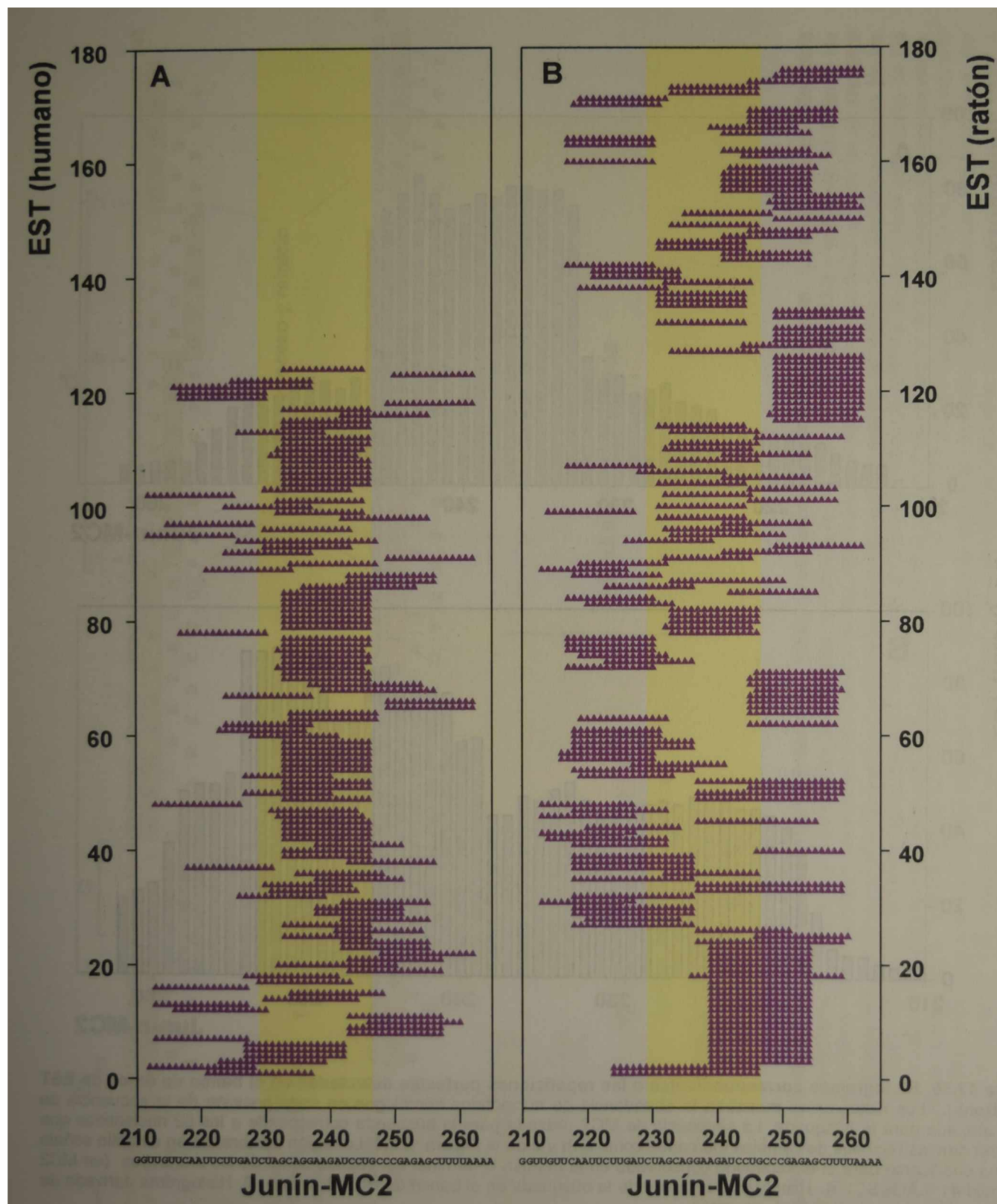


Figura 12.56. Gráficos correspondientes a las repeticiones perfectas detectadas en el banco de datos de EST (GenBank). Las sucesiones de triángulos representan los diferentes oligonucleótidos homólogos detectados. El rango de tamaños de las repeticiones perfectas halladas abarca desde 14 hasta 24 nucleótidos, tolerando diferencias en sólo un residuo y no aceptando la inserción de *gaps*. La secuencia de MC2 utilizada para la búsqueda corresponde a los 52 nucleótidos que comprenden las regiones de cambios marcadas como a, b y c en la Figura 12.54, y se indica debajo de cada gráfico. La región sombreada en amarillo señala la zona codificante para el hexapéptido conservado en la porción C-terminal del péptido señal de los arnavirus (en MC2 y *Candid* #1: LAGRSC). **A.** Gráfico derivado de la búsqueda en el banco de EST humano. **B.** Gráfico derivado de la búsqueda en el banco de EST de ratón.

Sección II - Extremos 5' de los RNAs virales

Extremos 5' de los RNAs virales obtenidos a partir de viriones y a partir de células infectadas

Como se mencionó en los **Capítulos 5** (MC2) y **9** (Candid #1), los extremos 5' de los RNAs S virales de las cepas del virus Junín analizadas en esta tesis, se determinaron a partir de clones de cDNA (pJUN58 en la cepa MC2; pC6 y pC295 en la cepa Candid #1) que se obtuvieron por los métodos tradicionales de síntesis de primera y segunda cadena de cDNA. Además, en el caso de la cepa MC2, se obtuvo información adicional sobre los extremos 5' de RNAs virales (genómicos y subgenómicos, en particular el mRNA de GPC), analizando las secuencias nucleotídicas de clones de cDNA generados mediante extensión de *primer* sobre RNA total de células BHK infectadas, utilizando un *primer* específico de la región del gen GPC (ver **Figura 5.3**, **Capítulo 5**).

En la **Figura 12.57** se pueden observar las secuencias correspondientes a los extremos 5' de los RNAs S genómicos de las cepas MC2 y Candid #1 del virus Junín; las de los extremos 5' de los RNAs virales intracelulares; y las de los extremos 5' de los RNAs S genómicos de otros arnavirus.

El análisis detallado de un alineamiento múltiple de estas secuencias permite observar aspectos que resultan llamativos.

En primera instancia, todas las secuencias presentan un conjunto de 13 nucleótidos muy conservados: 5' GgGGauCCU AGGC 3' (en mayúsculas los residuos presentes en todas las secuencias y en minúsculas los presentes en la mayoría), ubicado a una distancia variable (5 a 9 residuos) desde el extremo 5' descrito, salvo en el caso de Mopeia en el cual ha sido descrito como el

verdadero extremo 5' (**Figura 12.41**). Si denominamos +1 a la primer G del conjunto de 13 nt. conservado, la variabilidad considerada en las posiciones con residuos indicados en minúsculas corresponde a:

- ⇒ la existencia de una U en +2, solamente en Mopeia
- ⇒ la existencia de una G en +5, solamente en el clon 1/10 de la cepa MC2
- ⇒ la ausencia de una U en +6, solamente en el clon 2/59 de la cepa MC2

En segunda instancia, en los residuos previos al conjunto conservado (denominados arbitrariamente -1 a -9) se pueden observar diferencias importantes:

- ⇒ los residuos -1 a -5 correspondientes a los RNAs S genómicos de las cepas MC2 y Candid #1 del virus Junín son iguales entre sí, pero diferentes a los derivados de RNAs intracelulares obtenidos a partir de células infectadas con MC2, y diferentes a los del resto de los arnavirus.
- ⇒ el residuo -6 es idéntico en todos.
- ⇒ el residuo -7 está conservado en las secuencias del virus Junín (salvo en el clon 2/59) y en los arnavirus LAJ y SAB.
- ⇒ el residuo -8 sólo se ha detectado en el virus Junín y es igual en los RNAs genómicos de MC2 y Candid #1, pero diferente a los derivados de RNAs intracelulares obtenidos a partir de células infectadas con MC2.
- ⇒ el residuo -9 sólo se ha detectado en el RNA genómico de Candid #1 y en el clon 2/52 derivado de MC2, y es diferente en ambos.

Discusión sobre los extremos 5' de los RNAs genómico, antígenómico y mRNAs

En 1990, Raju *et al.* describen la presencia de 3 a 5 nucleótidos adicionales en el extremo 5' del mRNA codificante para GPC de Tacaribe. Dichos nucleótidos, no han podido ser detectados en el extremo 3' de la forma antígenómica del RNA S de Tacaribe, que es el molde para la síntesis del mRNA mencionado. Por otra parte, también detectan una G adicional en el extremo 5' de la forma genómica del RNA S de Tacaribe, no descrita en la publicación original. La presencia de esta G es coincidente con lo descrito, hasta ese momento, para Lassa-Josiah y para Pichindé. En otra publicación de 1990, Garcín & Kolakofsky confirman los datos anteriores.

Posteriormente, en 1992, Garcín & Kolakofsky describen un modelo, llamado *cap-snatching*, para la iniciación de la replicación (síntesis de moléculas genómicas y antígenómicas). Este modelo es dependiente de un *primer*, y es similar al descrito para el virus de influenza (Bouloy *et al.*, 1978, 1979; Caton and Robertson, 1980; Plotch *et al.*, 1981; Krug, 1981) y para los bunyavirus (Bishop *et al.*, 1983; Patterson *et al.*, 1983, 1984; Collett, 1986; Kolakofsky and Hacker, 1991).

En 1993, Meyer y Southern, utilizando una estrategia de circularización de los RNAs (ligación de los extremos 5' y 3') y posterior amplificación por PCR y secuenciación, demostraron la presencia de 0 a 7 nucleótidos adicionales en los extremos 5' de los mRNAs codificantes para GPC y N del virus LCM. Además, empleando ensayos de cromatografía de afinidad en boronato y tratamiento con pirofosfatasa ácida de tabaco (TAP) verificaron experimentalmente que los extremos 5' de los mRNAs de LCM presentan un residuo m⁷G unido por un trifosfato al siguiente nucleótido (estructura tipo cap).

En 1994, Meyer & Southern, empleando la estrategia de circularización de RNAs, verifican la existencia de heterogeneidad

en los extremos de los RNAs genómico y antígenómico del mismo arnavirus. Además, también verifican la necesidad de realizar el tratamiento con TAP para obtener un producto de amplificación definido. Esto último evidencia la presencia de un extremo 5' trifosfato (o 5' cap, no demostrado en forma directa). A diferencia del caso de los mensajeros, la heterogeneidad encontrada en los extremos 5' de las moléculas de RNA S corresponde a la presencia de 0 a 3 nucleótidos adicionales, deleciones y/o cambios puntuales en alguno de los primeros nucleótidos. Por otra parte, en los extremos 5' de las moléculas de RNA L, genómico y antígenómico, encuentran dos tipos de extremos, con 5' trifosfato y con 5' monofosfato. En el último caso, las moléculas de RNA L presentan deleciones terminales (desde 2 hasta 38 nt.) y la adición de residuos no codificados en el molde (1 a 6 nt.).

Por último, en un trabajo publicado en 1997, Meyer & Southern presentan resultados que demuestran la capacidad replicativa de las moléculas de RNA S genómica y antígenómica que poseen deleciones, tanto en el extremo 5' como 3', y la persistencia de los cambios en las moléculas sintetizadas sobre esos moldes defectivos. Es de destacar que las moléculas de RNA S con extremos modificados (deleciones y/o adiciones) en los extremos son encapsidadas y, en algunos casos, pueden ser reparadas mediante mecanismos desconocidos. Por otra parte, también demuestran que las moléculas de RNA S genómico y antígenómico con extremos 5' con deleciones importantes (ausencia de la mayor parte de los 19 nt. conservados en todos los arnavirus), no son utilizadas como moldes para la transcripción.

El conjunto de datos analizado en los párrafos precedentes sugiere que, en el caso de los arnavirus, los mecanismos de replicación y encapsidación no son estrictamente dependientes de la integridad de los extremos 5' y/o 3'; mientras que la integridad de los extremos 5' sería crucial para la transcripción.

En cualquiera de los dos casos, la presencia de nucleótidos adicionales en el extremo 5', no codificados en el molde, no parece afectar a ninguno de los procesos.

Por otra parte, el origen de los nucleótidos adicionales en los extremos 5' es desconocido, así como los mecanismos que llevan a la existencia de extremos 5' trifosfato o 5' cap. En función de los datos obtenidos por los autores precedentemente mencionados, se pueden elaborar, al menos, tres hipótesis:

1. La existencia de un mecanismo idéntico o equivalente al *cap stealing* descrito para el virus de influenza y para algunos bunyavirus, o el *cap-snatching* postulado por Garcin & Kolakosfsky (1992).
2. La polimerasa viral puede utilizar como *primers* para iniciar la transcripción oligorribonucleótidos generados al azar durante el procesamiento natural de recambio de los mRNAs celulares y agregar el 5' cap en el citoplasma celular mediante una actividad asociada con el complejo replicativo viral.
3. La polimerasa viral puede sintetizar *primers* cortos en la aparente ausencia de un molde y agregar el 5' cap mediante una actividad asociada con el complejo replicativo viral.

Respecto de la hipótesis 1, desde 1990 a la fecha, se ha publicado información adicional que sustenta una cierta generalidad de este mecanismo en diferentes virus con genoma de RNA. En los mRNAs purificados a partir de células infectadas con un tenuivirus, el virus Rice Hoja Blanca (RHBV), se han encontrado entre 10 a 17 nucleótidos adicionales (de secuencia heterogénea) y una estructura tipo cap en el extremo 5' (Ramirez *et al.*, 1995). Además, en ensayos *in vitro*, se ha demostrado que la RNA polimerasa RNA dependiente de RHBV tiene la capacidad de incorporar en el extremo 5' sintetizado *de novo*, los primeros 10 a 16 nt. derivados de un cap-RNA heterólogo agregado a la reacción (Nguyen *et al.*, 1997).

Este último resultado es equivalente al obtenido por Bouloy *et al.* en 1978: al agregar cap-mRNAs de globina, en ensayos de transcripción *in vitro* con extractos del virus de influenza, detectaron la incorporación de los primeros 10 a 15 nt. en el extremo 5' sintetizado *de novo*.

Sin embargo, los resultados de algunos trabajos recientes sobre el virus de influenza A (Cianci *et al.*, 1995; Galarza *et al.*, 1996), demuestran que, al menos para este virus, la síntesis *in vitro* de RNAs virales es muy dependiente de la manera en que se realizan los experimentos. En experimentos de transcripción /replicación *in vitro*, utilizando la RNA polimerasa viral, purificada a partir de complejos ribonucleoproteicos virales liberados de RNA endógeno por tratamiento con nucleasa micrococcal, se han obtenido resultados en parte coincidentes con la información preexistente y en parte inesperados. La síntesis de RNA viral utilizando nucleoproteínas virales como molde y el dinucleótido ApG como *primer* permite obtener copias completas del vRNA. Mientras que la síntesis de RNA viral utilizando nucleoproteínas virales como molde y el cap-RNA4 del virus del mosaico de la alfalfa (AIMV) como fuente del *primer* con cap, permite detectar la presencia de 10 a 11 nt. adicionales en el extremo 5' de los RNAs sintetizados *de novo*. Sorprendentemente, estos nucleótidos adicionales no corresponden al extremo 5' del cap-RNA4 de AIMV sino a una región cercana al extremo 3' del propio RNA de influenza. En síntesis, la RNA polimerasa viral purificada protege de la degradación con nucleasa micrococcal un segmento de RNA viral original, cuando se agrega un cap-RNA heterólogo la síntesis de RNA se activa; sin embargo, el *primer* utilizado mayoritariamente es el fragmento protegido y no el extremo 5' del RNA adicionado.

Por otra parte, para el virus Hantaan (*Bunyaviridae*), Garcin *et al.* (1995) han propuesto un mecanismo de *Prime-and-Realign* para explicar las evidencias experimentales. Los

extremos 5' y 3' de los tres RNAs virales (S, M y L) de todos los bunyavirus descritos hasta el momento presentan un conjunto de 9 residuos complementarios entre sí (5' UAGUAGUAG....CUACUACUA 3'). Los mRNAs del virus Hantaan presentan un extremo 5' cap con 8 a 17 nt. adicionales y una gran predominancia de G como nucleótido -1, mientras que los RNAs genómicos y antígenómicos presentan un extremo 5' pU (monofosfato) y carecen de nucleótidos adicionales. El mecanismo propuesto se basa principalmente en la predominancia de Gs en la posición -1. En el mismo se postula el apareamiento de una G (derivada de alguna región próxima al extremo 5' de un RNA celular) con la primer C que aparece en el molde (posición +3), el comienzo de la síntesis (3 o 4 nt.), el desapareamiento y reapareamiento en posiciones previas, con la ulterior continuación de la síntesis. En el caso de los mRNAs, el primer utilizado se mantiene, mientras que en el caso de los RNAs genómicos y antígenómicos se produciría un clivaje endonucleolítico entre la G de -1 y la U de +1, quedando un extremo 5' monofosfato. Sin embargo, el análisis de las secuencias de los extremos 5' de los mensajeros, presentadas en el trabajo citado, no permite discernir por qué los autores centran su hipótesis en el apareamiento inicial de un único nucleótido (G), cuando es igualmente probable que este apareamiento inicial involucre más nucleótidos derivados del extremo 5' de un RNA celular. La existencia de extremos 5' de mensajeros donde el nucleótido -1 es diferente de G, o donde falta alguna de las repeticiones UAG, sustenta mejor la última hipótesis (**Figura 12.58.A**).

En el virus Thogoto, un miembro de la familia *Orthomyxoviridae*, Leahy *et al.* (1997) han postulado la generación de mRNAs mediante una variante del mecanismo de *cap-snatching*. En ensayos *in vitro*, utilizando análogos de cap o mRNAs de α -globina y

β -globina, los autores muestran una gran preferencia por la incorporación de estructuras tipo cap-0 (m⁷GpppN) o cap-1 (m⁷GpppNm: 7'-metil-GpppN-2'-O-metil) en el extremo 5' de los mRNAs sintetizados. Un aspecto distintivo en este sistema es la ausencia de nucleótidos adicionales entre el cap y el extremo 5' del mRNA, de acuerdo con la secuencia del molde. Sin embargo, no se puede descartar que, en la síntesis *in vivo*, el extremo 5' cap de los mRNAs virales, derivado del conjunto de hnRNAs, presente secuencias compatibles con la posibilidad de un apareamiento perfecto con el extremo 3' del molde en una cantidad de nucleótidos mayor que 1.

Por último, en 1998, Estabrook *et al.* publican un trabajo en el cual muestran la transferencia *in vivo* de los extremos 5' cap de los RNAs α , β y γ del virus *Barley Stripe Mosaic Hordeivirus* (BSMV) al extremo 5' del mRNA4 del virus *Maize Stripe Virus* (MStV). En este trabajo, los autores coinfectan plantas con ambos virus y buscan específicamente la existencia de mRNAs quiméricos mediante RT-PCR, utilizando *primers* específicos para los 10 primeros nucleótidos de los RNAs α , β y γ de BSMV y un *primer* interno específico para el RNA4 de MStV.

Aún cuando este esquema experimental no permite observar toda la diversidad de extremos generada *in vivo*, resulta enriquecedor en varios aspectos. En primera instancia, se puede observar que el inicio de la síntesis del mRNA4 de MStV puede derivar del apareamiento de 1 a 4 nt. provenientes de alguno de los RNAs de BSMV (**Figura 12.58.B1.1, 1.2, 1.4, 2.1, 2.2, 3.1, 3.2 y 3.3**). Por otra parte, también se puede observar la aparente existencia de mecanismos equivalentes al *Prime-and-Realign* (**Figura 12.58.B2.3 y 2.4**) y al *RNA-editing* (**Figura 12.58.B1.3**).

A

RNA S		+1
		UAGUAGUAGACUCC...
uaaugacuca		UAGUAGUAGACUCC...
ucuuuguucuc		UAGUAGUAGACUCC...
aucgcuucucg		UAGUAGUAGACUCC...
aggauugggagguag		UAGUAGUAGACUCC...
agauaagcagcg		--- UAGUAGACUCC...
agguuggucucg		--- UAGUAGACUCC...

RNA M		+1
		UAGUAGUAGACUCC...
aguguuacagcucuuu		UAGUAGUAGACUCC...
cccugucucccccg		UAGUAGUAGACUCC...
augcuuccucug		UAGUAGUAGACUCC...
agucccgcacccg		--- UAGUAGACUCC...
ugucuauggucgg		--- UAGUAGACUCC...

B

	MStV RNA4		+1
			ACACAAAGUCCAGGGCAUUU...
1	αBSMV	+1	GUAUGUAAGUUGCC...
1.1	αBSMV/MStV RNA4	<u>quauquaagu</u> gc-----CACAAAGUCCA...	
1.2	αBSMV/MStV RNA4	<u>quauquaagu</u> g-----CACAAAGUCCA...	
1.3	αBSMV/MStV RNA4	<u>quauquaagu</u> u--- ACCG ---ACACAAAGUCCA...	
1.4	αBSMV/MStV RNA4	<u>quauqu</u> ----- AAGUCCA ...	
2	βBSMV	+1	GUAAAAGAAAAGGAACA...
2.1	βBSMV/MStV RNA4	<u>quaaaagaaa</u> agga-----ACACAAAGUCCA...	
2.2	βBSMV/MStV RNA4	<u>quaaaagaaa</u> -----ACACAAAGUCCA...	
2.3	βBSMV/MStV RNA4	<u>quaaaagaaa</u> ----- C ---ACACAAAGUCCA...	
2.4	βBSMV/MStV RNA4	<u>quaaaagaaa</u> ----- CACAC ---ACACAAAGUCCA...	
3	γBSMV	+1	GUAUAGCUUGAGCAUUA...
3.1	γBSMV/MStV RNA4	<u>quauagcuu</u> gag-----CACAAAGUCCA...	
3.2	γBSMV/MStV RNA4	<u>quauagcuu</u> g-----ACACAAAGUCCA...	
3.3	γBSMV/MStV RNA4	<u>quauagcuu</u> gag-----CAAAGUCCA...	
3.4	γBSMV/MStV RNA4	<u>quauagcuu</u> ga-----ACACAAAGUCCA...	

Figura 12.58. Extremos 5' de mRNAs de virus con genoma de RNA. **A.** Extremos 5' de los mRNAs del virus Hantaan (*Bunyaviridae*), tomado de Garcin *et al.* (1995). En negrita y mayúsculas se indican las secuencias originales de los extremos 5' de los RNAs antígenicos. En minúsculas se indican las secuencias heterogéneas de nucleótidos adicionales y en mayúsculas las correspondientes a los RNAs de Hantaan. En rojo se destacan secuencias que no pueden ser explicadas por el mecanismo postulado por los autores. **B.** Extremos 5' de mRNAs quiméricos generados *in vivo* entre los virus *Maize Stripe Virus* (MStV) y *Barley Stripe Mosaic Hordeivirus* (BSMV), tomado de Estabrook *et al.* (1998). En negrita y mayúsculas se indican las secuencias originales del extremo 5' del RNA4 de MStV y de los extremos 5' de los cap-RNAs α, β y γ de BSMV. En minúsculas se indican las secuencias derivadas de los extremos 5' cap-RNA de BSMV (subrayados los 10 nucleótidos correspondientes al primer específico de BSMV) y en mayúsculas las correspondientes al RNA4 de MStV. En rojo se destacan las secuencias cuyo origen (BSMV o MStV) no puede ser determinado con precisión. En negrita e itálica se indican secuencias intercaladas de origen desconocido.

En función de todas las consideraciones anteriores, un nuevo análisis de las secuencias de los extremos 5' de los RNAs genómicos y subgenómicos de los arnavirus permite elaborar hipótesis adicionales sobre los mecanismos de transcripción y replicación en arnavirus.

En la **Figura 12.59.A.** se puede observar que en los extremos 5' subgenómicos (mRNAs) de LCM-Armstrong se ha detectado una gran conservación en las secuencias complementarias al extremo 3' del molde (RNA S antigenómico), y una gran variabilidad en el número y tipo de nucleótidos adicionales en el 5' (0 a 7 nt.; Meyer & Southern, 1993). Mientras que en los extremos 5' genómicos (RNA S) del mismo arnavirus se observa cierta variabilidad dentro de los primeros 10 nt. y un número menor de nucleótidos adicionales en el 5' (0 a 3; Meyer & Southern, 1994). Además, en la **Figura 12.59.B.** se puede observar la gran conservación existente entre todas las secuencias, incluyendo el extremo 5' antigenómico, a partir del conjunto de 4 Gs (9 de 13 nt. están totalmente conservados). Desde este punto de vista, lo más destacable es el pentanucleótido diferencial destacado en las secuencias de RNA S genómico de las cepas Candid #1 y MC2 del virus Junín.

Por otra parte, se debe mencionar que la integridad del extremo 3' del RNA molde parece ser un requerimiento bastante estricto, asociado con los eventos transcripcionales, en diferentes sistemas virales (Meyer & Southern, 1997; Whelan & Wertz, 1999). Mientras que las capacidades replicativa y de encapsidación no parecen depender de la integridad de los extremos (Meyer & Southern, 1997).

En síntesis, el universo de conocimiento acerca de los extremos 5' de los RNAs genómicos y subgenómicos de los arnavirus es bastante restringido, en particular, en lo que respecta a las regiones de mayor y menor conservación. Además, los resultados obtenidos por Meyer & Southern (1993, 1994, 1997)

abren ciertas dudas acerca de la utilización *in vivo* del mecanismo postulado por Garcín & Kolakofsky (1992). La mayor parte de los cambios de secuencia interna detectados no se pueden explicar mediante la iniciación de la síntesis de RNA por apareamiento de únicamente un dinucleótido con extremo 5' trifosfato (pppGC; Garcín & Kolakofsky, 1992). En la **Figura 12.60.** se muestran distintas alternativas de orígenes posibles de los extremos 5' genómicos y subgenómicos de los arnavirus.

Otros aspectos poco explorados de la bioquímica de los arnavirus están en relación con la/las actividades y sitios de unión de ligandos de la RNA polimerasa viral (proteína L), así como también con su capacidad de interactuar con diferentes proteínas virales (N y/o Z) o celulares para formar complejos transcripcionales o replicativos con capacidades diferenciales.

Desde este punto de vista, es necesario mencionar que, para el virus de influenza, una serie de trabajos relativamente recientes han permitido descubrir nuevas capacidades del complejo replicativo/transcripcional (Cianci *et al.*, 1995; Galarza *et al.*, 1996, Li *et al.*, 1998). De acuerdo a estas publicaciones, el complejo replicativo/transcripcional de influenza posee sitios de unión independientes para los extremos 5' y 3' del RNA viral, un sitio de unión para segmentos de RNA (viral o no), un sitio de unión para extremos 5'-cap derivados de mRNAs celulares y actividades de *capping* y de endonucleasa de RNA en adición a la actividad polimerásica.

Por otra parte, también es necesario mencionar que la RNA pol II de eucariotas, en muchas circunstancias, permanece detenida en el complejo transcripcional en alguna posición y la transcripción no progresa. El complejo transcripcional detenido es reactivado con la ayuda del factor de transcripción SII, el cual facilita el clivaje del transcrito 7 a 17 nt. antes del extremo 3' actual, generando un nuevo

extremo 3'-OH a partir del cual se reinicia la transcripción. El segmento de 7 a 17 nt. es liberado con un extremo 5'-p. En 1994, Rudd *et al.* demostraron que la RNA pol II por sí misma (sin el auxilio del factor SII) puede reactivar el complejo transcripcional mediante hidrólisis con PPI (pirofosforólisis). Esta hidrólisis ocurre en las mismas posiciones (7 a 17 nt. antes) en las cuales ocurre la hidrólisis mediada por SII. La principal diferencia es que los extremos 5' de los oligorribonucleótidos liberados son 5'-ppp.

Dado que la reacción de pirofosforólisis es exactamente la reversa de la síntesis, cualquier RNA polimerasa, incluyendo la de los arnavirus, podría tener esta capacidad.

Teniendo en cuenta las últimas consideraciones y el dato de que 6 de los 7 nucleótidos del extremo 5' del RNA 28 S de mamíferos (5' CGCGACC) son idénticos a los 6 primeros nucleótidos del extremo 5' consenso de los arnavirus (5' CGCACC), en la **Figura 12.61.** se postulan dos modelos alternativos para la iniciación de la transcripción en los arnavirus. En ellos, se considera la posible existencia de diferentes sitios de unión a ligandos y la posible existencia de una actividad de *capping* asociada al complejo transcripcional. Además, en la **Figura 12.62.** se postulan dos modelos alternativos para la iniciación de la replicación en los arnavirus. En ellos, se considera la posible existencia de diferentes sitios de unión a ligandos y un posible mecanismo de reparación de extremos delecionados.

Por otra parte, en la **Figura 12.63.** se presenta un modelo con un posible mecanismo mediante el cual, en algún momento de la evolución, se incorporó un extremo 5' genómico diferencial en el virus Junín *versus* el resto de los arnavirus. Este modelo se basa en la predicción de la existencia (dentro de las secuencias codificantes de los mamíferos) de un segmento de 18 nt. con la siguiente secuencia: 5' CGCACGGCUGCAGUAAG 3', donde los

primeros 8 nt. corresponden al extremo 5' consenso de los arnavirus y los siguientes corresponden al extremo 5' del RNA S de la cepa Candid #1 del virus Junín.

Mediante cálculo (ver **Apéndice 10**) se puede estimar que las probabilidades de existencia de este oligonucleótido en los genomas de mamíferos son:

Nº de veces	Probabilidad
1	$6,5 \times 10^{-3}$
2	$2,1 \times 10^{-5}$
5	$9,2 \times 10^{-14}$
10	$3,4 \times 10^{-29}$

Luego de una búsqueda exhaustiva en el GenBank, se encontraron dos tipos de secuencias que cumplen satisfactoriamente con las condiciones requeridas. Una de ellas presenta una identidad de 83,3 % (15 de 18 nt.) y corresponde a un segmento codificado en el cromosoma 22 humano. Mientras que la segunda secuencia presenta una identidad de 88,9 % (16 de 18 nt.) y corresponde a un segmento codificado en el cromosoma X humano (₅₂₈₉₂GCACAGGCUGCAGUAAG₅₂₉₀₈, número de acceso: AC006121). Es importante destacar que la segunda A es uno de los cambios posibles detectados en LCM-Armstrong (ver **Figura 12.59.B**).

En adición a lo anterior, en la **Figura 12.64.** se presenta un modelo con un posible mecanismo mediante el cual, en algún evento replicativo se puede incorporar un extremo 5' genómico coincidente con el consenso de los arnavirus. Este modelo se basa en la predicción de la existencia (dentro de las secuencias codificantes de los mamíferos) de un segmento de 16 nt. con la siguiente secuencia: 5' CGCACAGGCGCACAGG 3', donde existen dos repeticiones de 8 nt. que corresponden a una de las secuencias encontradas en el extremo 5' del virus LCM-Armstrong.

Mediante cálculo (ver **Apéndice 10**) se puede estimar que las probabilidades de

existencia de este oligonucleótido en los genomas de mamíferos son:

Nº de veces	Probabilidad
1	$7,1 \times 10^{-2}$
2	$2,7 \times 10^{-3}$
5	$2,0 \times 10^{-8}$
10	$1,8 \times 10^{-18}$

Luego de una búsqueda exhaustiva en el GenBank, se encontraron dos tipos de secuencias que cumplen satisfactoriamente con las condiciones requeridas. Una de ellas presenta una identidad de 87,5 % (14 de 16 nt.) y corresponde al clon humano DJ05114A23. Mientras que la segunda secuencia presenta una identidad de 93,8 % (15 de 16 nt.) y corresponde a un segmento codificado en el cromosoma 11 humano (₁₄₇₂₀CGCACAGGCGCACAG₁₄₇₃₄, número de acceso: U73644).

Para la confección del modelo se utilizó como base esta última secuencia. Además, se tienen en cuenta las mismas suposiciones mencionadas anteriormente (posible existencia de secuencias homólogas en porciones de los genomas de mamíferos no disponibles hoy día en el GenBank).

Por último, la persistencia de uno u otro tipo de extremo 5' genómico en los distintos arenavirus puede estar avalada por las diferentes frecuencias de existencia (en los genomas de mamíferos) de oligonucleótidos correspondientes a los primeros 10 ó 12 nt. de los extremos.

Considerando las secuencias de los extremos hasta el conjunto de 4 Gs, la probabilidad de existencia del oligonucleótido 5' UGCAGUAAGGGG 3' (virus Junín) en los genomas de mamíferos son:

Nº de veces	Probabilidad
2	$8,8 \times 10^{-15}$
28	$1,3 \times 10^{-2}$
36	$5,9 \times 10^{-2}$
100	$1,6 \times 10^{-16}$
370	$3,4 \times 10^{-220}$

Mientras que las probabilidades para el oligonucleótido 5' CGCACAGGGG 3' (resto de los arenavirus) en los genomas de mamíferos son:

Nº de veces	Probabilidad
2	$1,4 \times 10^{-156}$
28	$5,4 \times 10^{-119}$
36	$1,6 \times 10^{-110}$
100	$1,4 \times 10^{-62}$
370	$2,1 \times 10^{-2}$

En función de los cálculos anteriores, en los genomas de los arenavirus es esperable encontrar con mucha mayor frecuencia el extremo 5' considerado como consenso (5' CGCACAGGGG 3') que el extremo hallado en las cepas del virus Junín.

En este punto, es interesante destacar que una búsqueda en el GenBank de la posible existencia del oligonucleótido descrito para el virus Junín (5' UGCAGUAAGGGG 3') ha permitido detectar su presencia, al menos 36 veces, en regiones codificantes del genoma humano. Este valor es coincidente con los cálculos probabilísticos.

Por restricciones en el sistema de búsqueda (BLAST), no se ha podido evaluar la existencia del oligonucleótido de 10 residuos típico de la mayor parte de los arenavirus.

Obviamente, todos los modelos anteriormente presentados son explicaciones potenciales que deberían ser confirmadas experimentalmente. Sin embargo, el hecho absolutamente ineludible es que los eventos de crecimiento, purificación de viriones, síntesis de cDNA y secuenciación de los RNA S de las cepas Candid #1 y MC2 del virus Junín fueron absolutamente independientes. No obstante esto, los extremos 5' virales han resultado ser idénticos entre sí y diferentes a los extremos 5' del resto de los arenavirus.

A	LCA	5'	AUAGCGCACCGGGGAUCCUAGGC...	3'	
	LCA	5'	ACUGGCGCACCGGGGAUCCUAGGC...	3'	
	LCA	5'	AGCGGCGCACCGGGGAUCCUAGGC...	3'	mRNAs
	LCA	5'	UAUAGCGCACCGGGGAUCCUAGGC...	3'	
	LCA	5'	AUUGGCGCACCGGGGAUCCUAGGC...	3'	
	LCA	5'	AGGCGGCGCACCGGGGAUCCUAGGC...	3'	
	LCA	3'	CGCGUGGCCCCUAGGAUCCG...	5'	Antigenómico
	LCA	5'	CGCGCACAGUGGAUCCUAGGC...	3'	
	LCA	5'	ACAGCGCACCGGGGAUCCUAGGC...	3'	RNA S (v)
	LCA	5'	GCACACCGGGGAUCCUAGGC...	3'	
	LCA	5'	GCAGCGCGGGGAUCCUAGGC...	3'	
	LCA	5'	GCAGCGCGGGGAUCCUAGGC...	3'	
	B	CD1	5'	CUGCAGUAAGGGGAUCCUAGGC...	3'
MC2		5'	UGCAGUAAGGGGAUCCUAGGC...	3'	RNA S (v)
MC2 1/10		5'	CGCGCGCCGGGGUCCUAGGC...	3'	mRNA?
MC2 2/47		5'	GGCGCACCGGGGAUCCUAGGC...	3'	mRNA?
MC2 2/52		5'	GGGCGCACCGGGGAUCCUAGGC...	3'	mRNA?
MC2 2/59		5'	CA-CGCACCGGGGA-CCUAGGC...	3'	mRNA?
Arenas		5'	CGCACAGUGGAUCCUAGGC...	3'	Antigenómico
LCE		5'	GCGCACCGGGGAUCCUAGGC...	3'	RNA S (v)
LCA		5'	CGCGCACAGUGGAUCCUAGGC...	3'	RNA S (v)
LCA		5'	ACAGCGCACCGGGGAUCCUAGGC...	3'	RNA S (v)
LCA		5'	GCACACCGGGGAUCCUAGGC...	3'	RNA S (v)
LCA		5'	GCGCGCCGGGGAUCCUAGGC...	3'	RNA S (v)
MOP	5'	-----GUGGAUCCUAGGC...	3'	RNA S (v)	
					** *****...

Figura 12.59. Extremos 5' de RNAs genómicos y subgenómicos de arenavirus. En los dos alineamientos se destacan con colores diferentes detalles, en azul se señala un pentanucleótido dentro del cual se encuentran las mayores diferencias nucleotídicas, en rojo se indican las variaciones nucleotídicas y en violeta se señalan los nucleótidos adicionales en el 5'. **A.** Alineamiento múltiple de las secuencias correspondientes a la región no codificante 5' del RNA S (genómico) y a la región no codificante 5' de RNAs virales intracelulares del virus LCM-Armstrong (LCA), tomado de Meyer & Southern (1993, 1994, 1997). Intercalada entre las secuencias de los mRNAs y de los RNAs S virales (v) se muestra el extremo 3' antigenómico (fucsia) para destacar la complementariedad total o parcial de las diferentes secuencias. **B.** En la parte superior se muestran los extremos de las secuencias genómicas de las cepas Candid #1 (CD1) y MC2 del virus Junín y secuencias correspondientes a clones de cDNA obtenidos a partir de RNA total de células BHK infectadas con la cepa MC2 del virus Junín. En la parte inferior se muestra la secuencia del extremo 5' del RNA S viral del virus LCM-WE (LCE) como referencia de lo descrito para la mayoría de los arenavirus. Además, se muestran secuencias de extremos 5' genómicos de otros arenavirus: LCM-Armstrong (LCA) y Mopeia (MOP). En el centro se muestra el extremo 5' antigenómico (fucsia) para destacar el grado de conservación nucleotídica en ambos extremos (indicado con asteriscos).

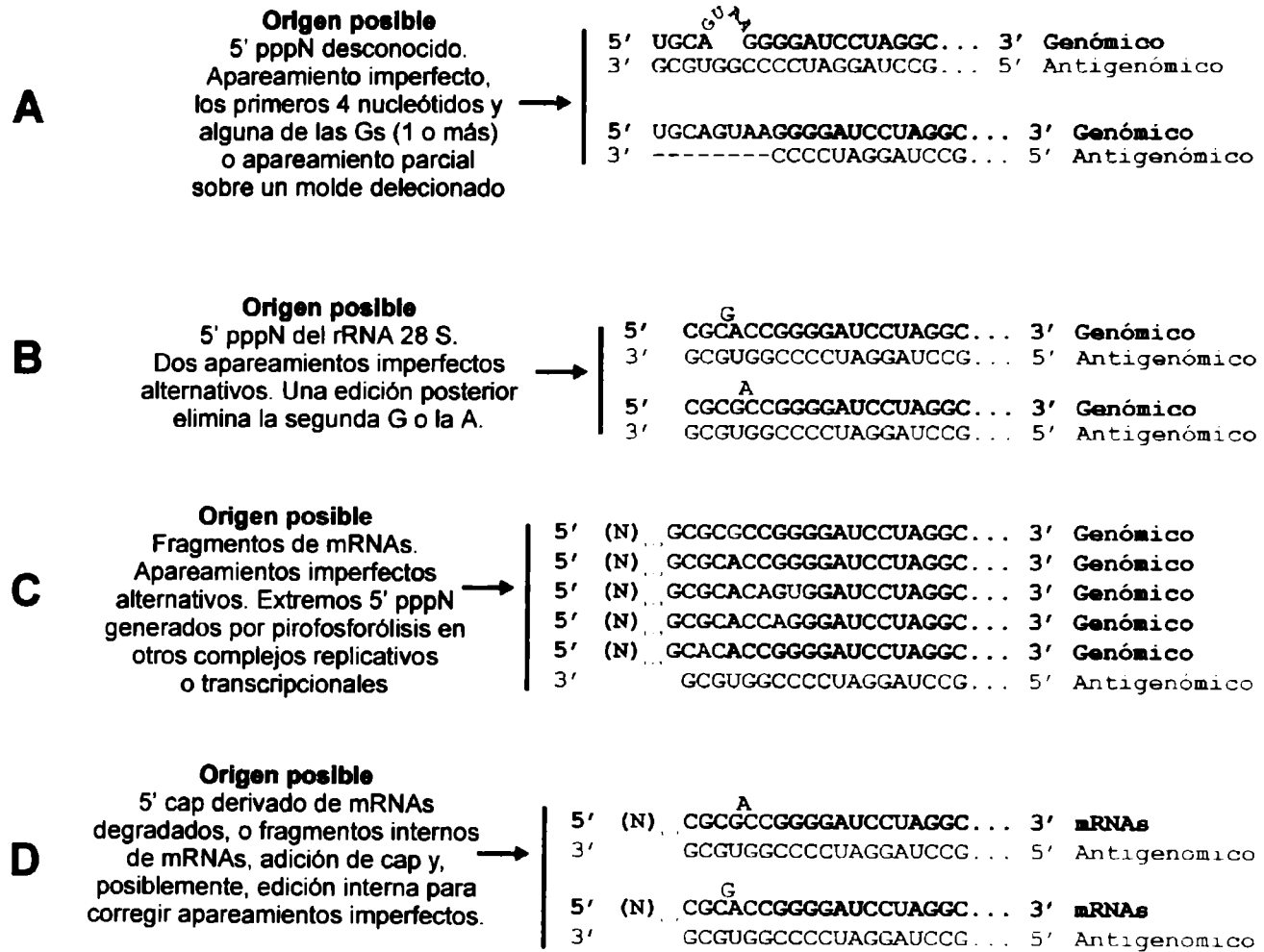


Figura 12.60. Orígenes posibles de los extremos 5' de los RNAs genómicos y subgenómicos de arnavirus. En los bloques se destacan con colores diferentes detalles, en negro se indican los nucleótidos coincidentes con el consenso, en azul se señalan las secuencias nucleotídicas mínimas que pueden hibridar con el molde (extremo 3' antigenómico, indicado en fucsia), en rojo se indican las variaciones nucleotídicas y en violeta se señalan los nucleótidos adicionales en el 5'. **A.** Origen posible de los extremos 5' de los RNAs S de las cepas Candid #1 y MC2 del virus Junín: un segmento de RNA celular con un 5'-ppp. **B.** Origen posible de los extremos 5' de los RNAs S de otros arnavirus: a partir del extremo 5'-trifosfato del rRNA 28 S. En este caso no existen nucleótidos adicionales en el 5'. **C.** Origen posible de los extremos 5' de los RNAs S de otros arnavirus: a partir de extremos 5'-trifosfato generados por pirofosforólisis en otros complejos replicativos o transcripcionales. En este caso pueden existir nucleótidos adicionales en el 5'. **D.** Origen posible origen de los extremos 5' de los mRNAs de arnavirus: a partir de extremos 5'-cap derivados de extremos 5' de mensajeros celulares degradados, o a partir de extremos 5'-trifosfato generados por pirofosforólisis en otros complejos replicativos o transcripcionales y posterior adición del cap. En este caso pueden existir nucleótidos adicionales en el 5'.

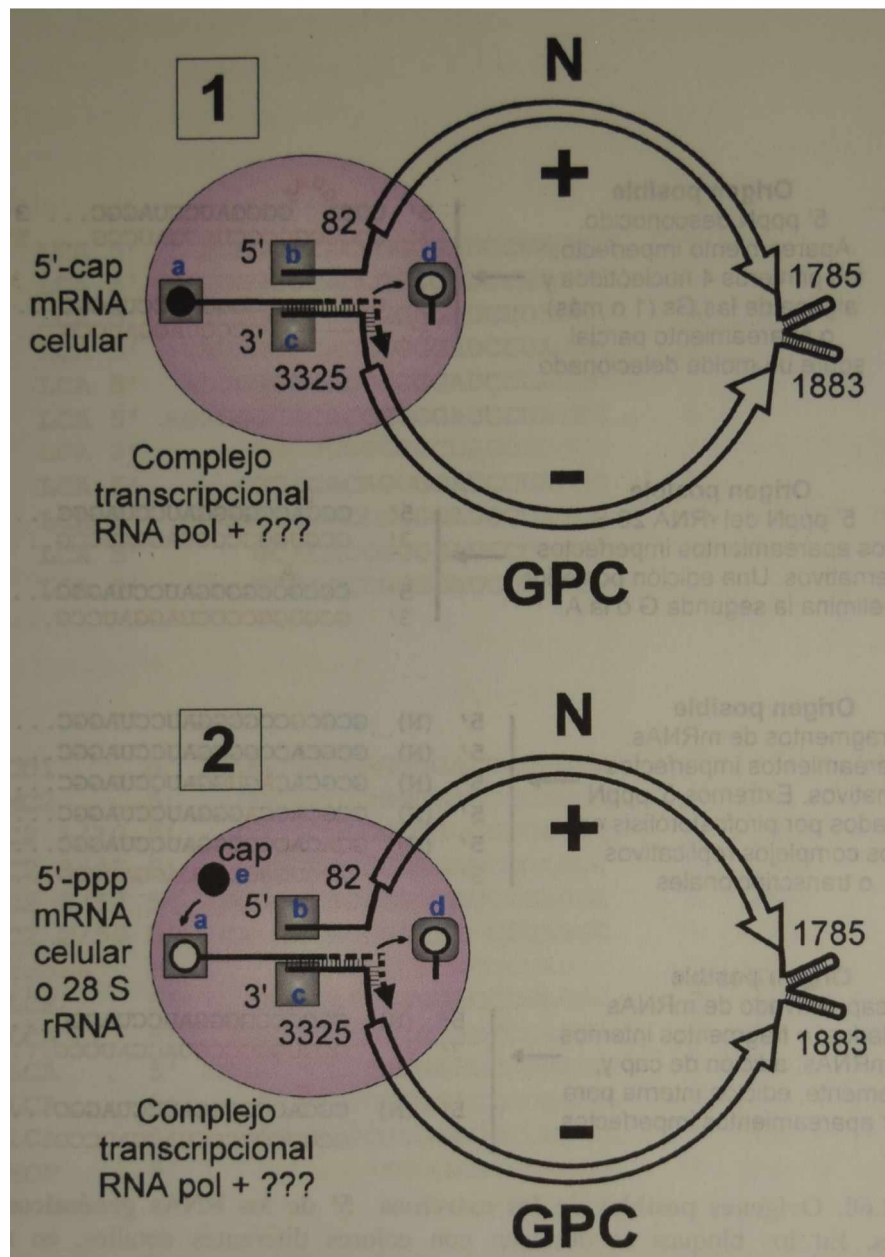


Figura 12.61. Modelos alternativos de transcripción basados en el RNA viral complementario de la cepa Candid #1 del virus Junín. Con un óvalo de color lila se indica el complejo transcripcional, formado por la RNA polimerasa (proteína L). Las diferencias entre el complejo transcripcional y el replicativo pueden deberse a la asociación de la proteína L con alguna otra proteína viral (por ejemplo la proteína Z) y/o una o más proteínas celulares. Con recuadros de color gris se destacan posibles sitios de unión de ligandos.

a. Sitio de unión de extremos 5' m⁷G-ppp-N o 5' ppp-N de RNAs celulares. **b.** Sitio de unión del extremo 5' intacto, viral complementario o viral. **c.** Sitio de unión del extremo 3' intacto, viral complementario o viral. **d.** Sitio de unión de un segmento de RNA con extremo 5' ppp-N generado por pirofosforólisis. **e.** Sitio de unión de m⁷G.

1. El complejo transcripcional, con los extremos 5' y 3' viral complementarios en sus respectivos sitios **b** y **c**, une un extremo 5' m⁷G-ppp-Ns derivado de mRNAs celulares y con cierto grado de complementariedad con el molde (1 o más nucleótidos) en alguna porción cerca del 5'. El complejo transcripcional avanza o queda detenido en función de la complementariedad con el molde de la región 3' del segmento de mRNA unido. El avance de la transcripción se asegura mediante escisión de la porción 3' no complementaria por pirofosforólisis, quedando un extremo 5'-ppp-N en el fragmento liberado. La escisión también podría ocurrir mediante una actividad endonucleolítica de la RNA polimerasa, quedando un extremo 5'-p-N en el fragmento liberado. El fragmento de RNA liberado puede ser retenido en el sitio de unión **d**.

2. El complejo transcripcional, con los extremos 5' y 3' viral complementarios en sus respectivos sitios **b** y **c**, une un extremo 5' ppp-Ns derivado de mRNAs celulares, con cierto grado de complementariedad con el molde (1 o más nucleótidos) en alguna porción cerca del 5', o de los primeros nucleótidos del rRNA 28 S. Una actividad asociada al complejo transcripcional agrega m⁷G en el extremo. El complejo transcripcional avanza o queda detenido en función de la complementariedad con el molde de la región 3' del segmento de mRNA unido. El resto del proceso sería idéntico a lo descrito para el modelo 1.

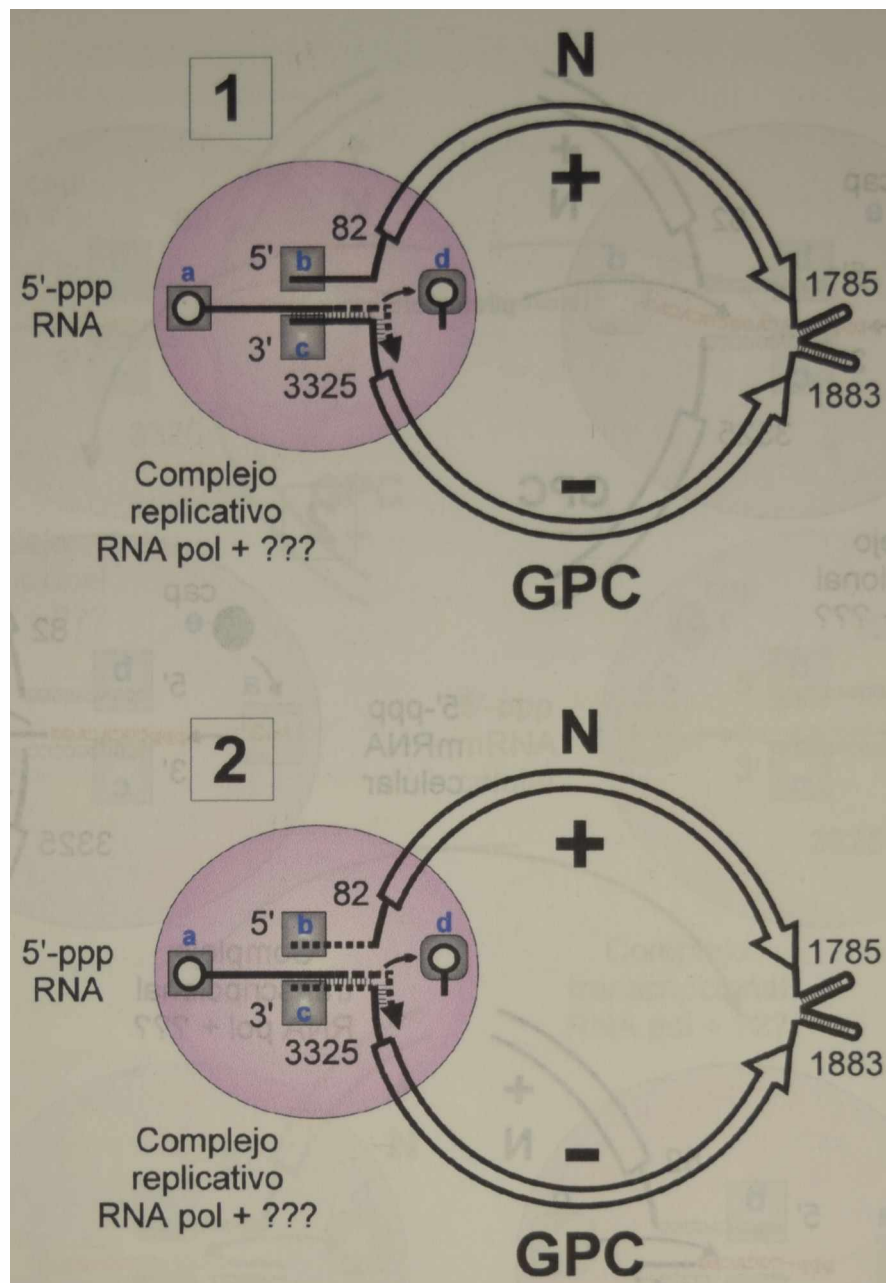


Figura 12.62. Modelos alternativos de replicación basados en el RNA viral complementario de la cepa Candid #1 del virus Junín. Con un óvalo de color lila se indica el complejo replicativo, formado por la RNA polimerasa (proteína L). Las diferencias entre el complejo replicativo y el transcripcional pueden deberse a la asociación de la proteína L con alguna otra proteína viral (por ejemplo la proteína N) y/o una o más proteínas celulares. Con recuadros de color gris se destacan posibles sitios de unión de ligandos. **a.** Sitio de unión de extremos 5' ppp-N o 5' m⁷G-ppp-N de RNAs celulares. **b.** Sitio de unión del extremo 5' intacto, viral complementario o viral. **c.** Sitio de unión del extremo 3' intacto, viral complementario o viral. **d.** Sitio de unión de un segmento de RNA con extremo 5' ppp-N generado por pirofosforólisis.

1. El complejo replicativo, con los extremos 5' y 3' viral complementarios intactos en sus respectivos sitios **b** y **c**, une un extremo 5' ppp-Ns derivado de RNAs celulares (puede ser el fragmento retenido en **d**) y con cierto grado de complementariedad con el molde (1 o más nucleótidos) en alguna porción cerca del 5', o derivado de los primeros nucleótidos del rRNA 28 S. El complejo replicativo avanza o queda detenido en función de la complementariedad con el molde de la región 3' del segmento de RNA unido. El avance de la replicación se asegura mediante escisión de la porción 3' no complementaria por pirofosforólisis, quedando un extremo 5'-ppp-N en el fragmento liberado. La escisión también podría ocurrir mediante una actividad endonucleolítica de la RNA polimerasa, quedando un extremo 5'-p-N en el fragmento liberado. El fragmento de RNA liberado puede ser retenido en el sitio de unión **d**.

2. El complejo replicativo, con los extremos 5' y 3' viral complementarios parcialmente delecionados (uno o los dos extremos) en sus respectivos sitios **b** y **c**, une un extremo 5' ppp-Ns derivado de RNAs celulares (puede ser el fragmento retenido en **d**) y con cierto grado de complementariedad con el molde (1 o más nucleótidos) en alguna porción cerca del 5', o derivado de los primeros nucleótidos del rRNA 28 S. El complejo replicativo avanza o queda detenido en función de la complementariedad con el molde de la región 3' del segmento de RNA unido. El resto del proceso sería idéntico a lo descrito para el modelo 1. En esta segunda alternativa se reparan los extremos delecionados con altas chances de introducir cambios drásticos en los nucleótidos de los extremos.

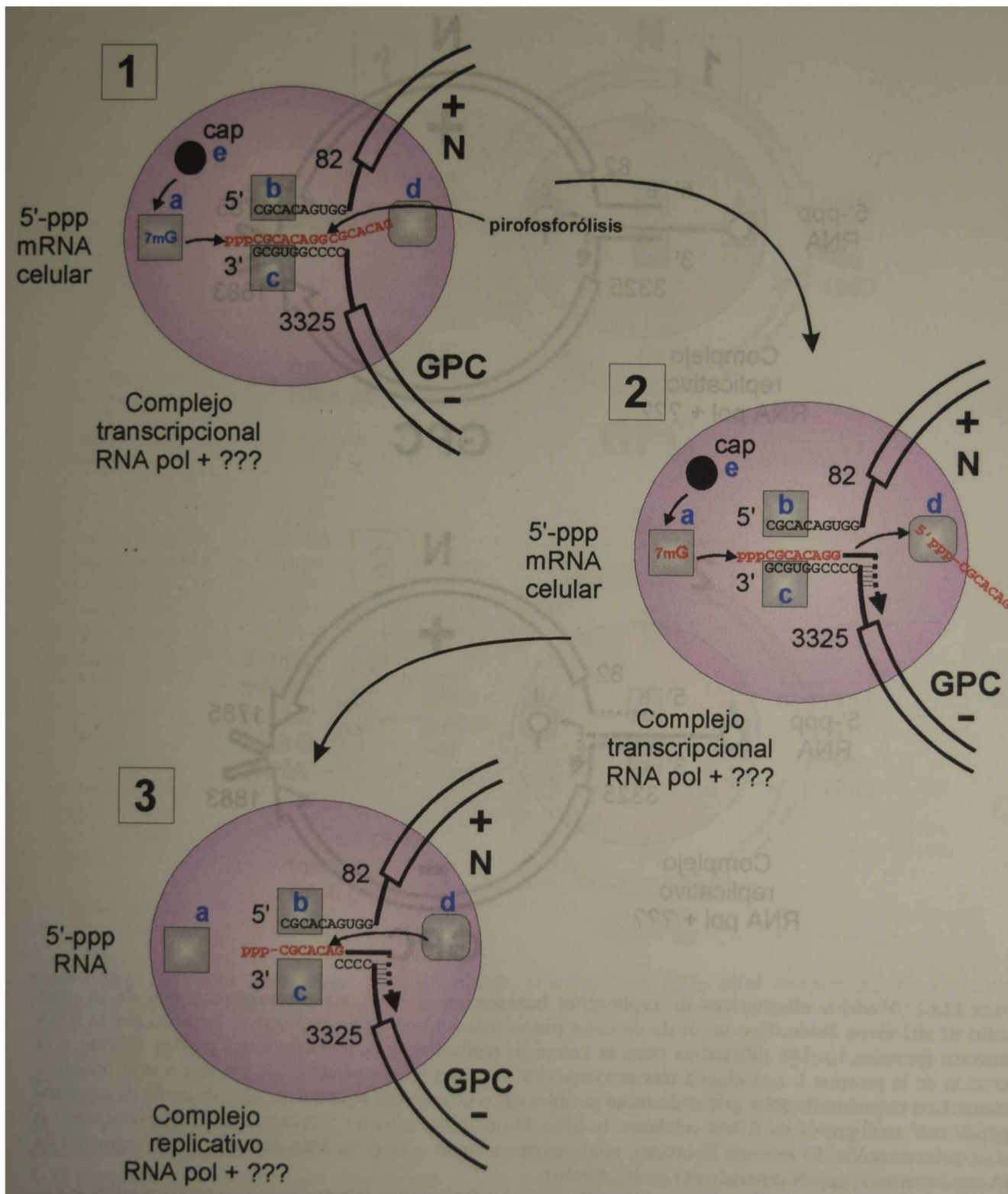


Figura 12.63. Modelo explicativo de la aparición de un extremo 5' genómico consenso en el virus Junín. El modelo se muestra en detalle en base al RNA viral complementario de la cepa Candid #1 del virus Junín. Con un óvalo de color lila se indican los complejos transcripcional y replicativo, formados como se describe en las figuras anteriores.

1. El complejo transcripcional, con los extremos 5' y 3' viral complementarios intactos en sus respectivos sitios b y c, une un extremo 5' ppp-Ns derivado de un RNA celular con cierto grado de complementariedad con el molde (8 nucleótidos). El complejo transcripcional queda detenido en función de la ausencia de complementariedad con el molde de la región 3' del segmento de RNA unido. Una actividad de capping asociada al complejo transcripcional agrega m'G al extremo 5' del fragmento de RNA hibridado con el molde.
2. El avance de la transcripción se asegura mediante escisión de la porción 3' no complementaria por pirofosforólisis, quedando un extremo 5'-ppp-N en el fragmento liberado. Este fragmento de RNA es retenido en el sitio de unión d.
3. El complejo replicativo, asociado a una molécula de RNA viral complementario con el extremo 3' parcialmente deletado unido al sitio c y el extremo 5' intacto unido al sitio b, une el segmento de RNA retenido en d por complementariedad parcial con el molde (1 o más nucleótidos) en alguna porción cerca del 5'. El complejo replicativo avanza o queda detenido en función de la complementariedad con el molde de la región 3' del segmento de RNA unido. Si el complejo queda detenido, se produce una nueva hidrólisis que asegura la disponibilidad de un extremo 3'-OH para continuar con el evento replicativo. Continúa la replicación y el nuevo extremo 5' queda incorporado al genoma.

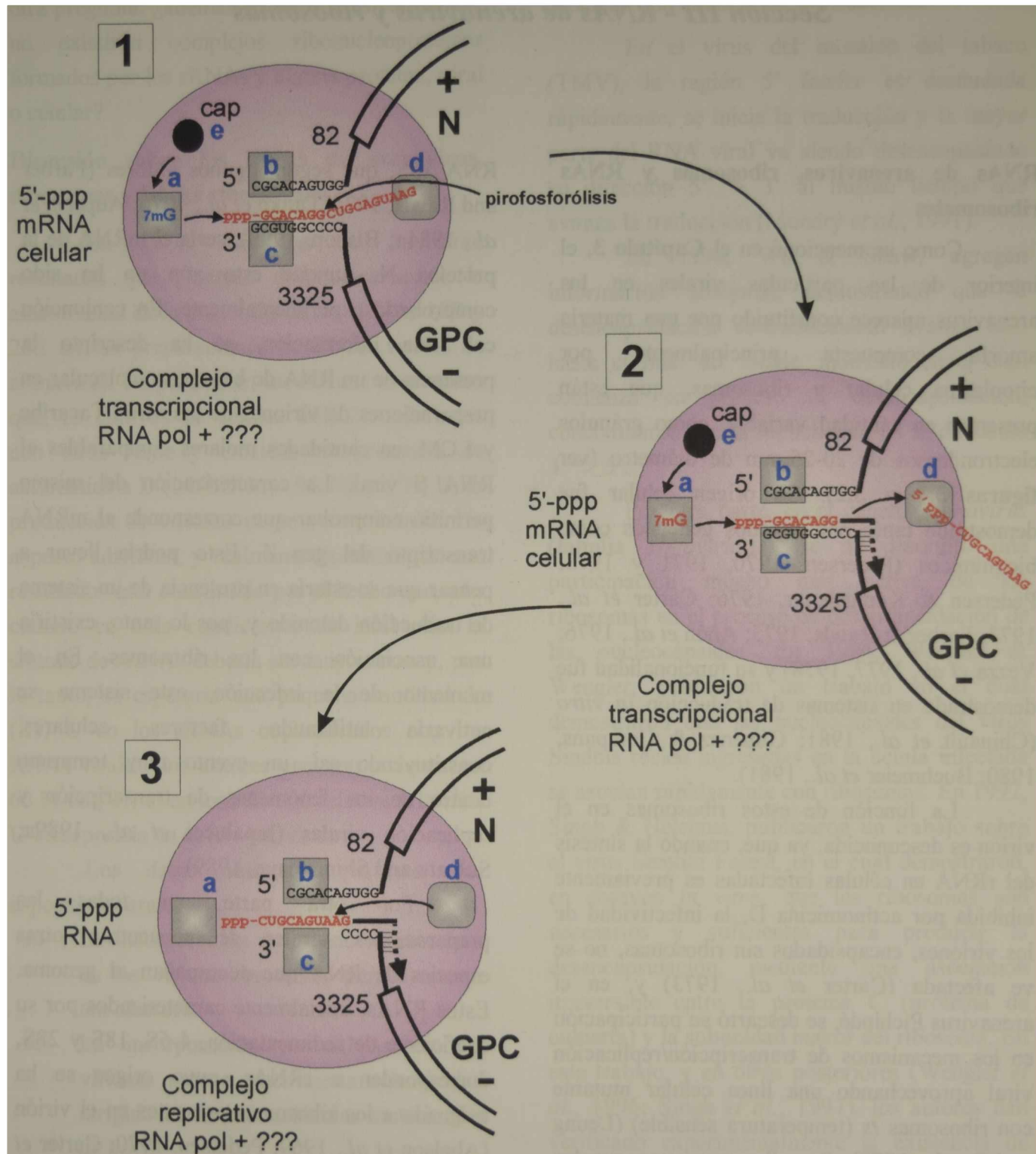


Figura 12.64. Modelo explicativo de la aparición de un extremo 5' genómico diferencial en el virus Junín versus el resto de los arenavirus. El modelo se muestra en detalle en base al RNA viral complementario de la cepa Candid #1 del virus Junín. Con un óvalo de color lila se indican los complejos transcripcional y replicativo, formados como se describe en las figuras anteriores.

1. El complejo transcripcional, con los extremos 5' y 3' viral complementarios intactos en sus respectivos sitios b y c, une un extremo 5' ppp-Ns derivado de un RNA celular con cierto grado de complementariedad con el molde (7 nucleótidos). El complejo transcripcional queda detenido en función de la ausencia de complementariedad con el molde de la región 3' del segmento de RNA unido. Una actividad de capping asociada al complejo transcripcional agrega m⁷G al extremo 5' del fragmento de RNA hibridado con el molde.
2. El avance de la transcripción se asegura mediante escisión de la porción 3' no complementaria por pirofosforólisis, quedando un extremo 5'-ppp-N en el fragmento liberado. Este fragmento de RNA es retenido en el sitio de unión d.
3. El complejo replicativo, asociado a una molécula de RNA viral complementario con el extremo 3' parcialmente delecionado unido al sitio c y el extremo 5' intacto unido al sitio b, une el segmento de RNA retenido en d por complementariedad parcial con el molde (1 o más nucleótidos) en alguna porción cerca del 5'. El complejo replicativo avanza o queda detenido en función de la complementariedad con el molde de la región 3' del segmento de RNA unido. Si el complejo queda detenido, se produce una nueva hidrólisis que asegura la disponibilidad de un extremo 3'-OH para continuar con el evento replicativo. Continúa la replicación y el nuevo extremo 5' queda incorporado al genoma.

Sección III - RNAs de arnavirus y ribosomas

RNAs de arnavirus, ribosomas y RNAs ribosomales

Como se mencionó en el Capítulo 3, el interior de las partículas virales en los arnavirus aparece constituido por una materia amorfa compuesta principalmente por citoplasma celular y ribosomas, que están presentes en cantidad variable, como gránulos electrondensos de 20-25 nm de diámetro (ver figuras 3.1 y 3.2). Su origen celular fue demostrado tanto por métodos genéticos como bioquímicos (Pedersen, 1970, 1971 y 1973; Pedersen & Königshofer, 1976; Carter *et al.*, 1973; Farber & Rawls, 1975; Añón *et al.*, 1976; Vezza *et al.*, 1977, 1978) y su funcionalidad fue demostrada en sistemas de traducción *in vitro* (Chinault *et al.*, 1981; Giménez & Compans, 1980; Buchmeier *et al.*, 1981).

La función de estos ribosomas en el virión es desconocida, ya que, cuando la síntesis del rRNA en células infectadas es previamente inhibida por actinomicina D, la infectividad de los viriones, encapsidados sin ribosomas, no se ve afectada (Carter *et al.*, 1973) y, en el arnavirus Pichindé, se descartó su participación en los mecanismos de transcripción/replicación viral aprovechando una línea celular mutante con ribosomas *ts* (temperatura sensible) (Leung and Rawls, 1977). Por ello, se supone que los ribosomas son incluidos dentro del virión debido a un mecanismo de encapsidación poco específico.

Además, los arnavirus son capaces de encapsidar también, algunas de las especies de RNAs utilizadas como intermediarios en la replicación viral, tales como la forma antigenómica del RNA S, aunque su proporción no supera el 5% de la masa del RNA S genómico. En algunas preparaciones del virus Pichindé, se ha detectado por *Northern blot* un

RNA 15S, que según algunos autores (Farber and Rawls, 1975; Dutko *et al.*, 1976; Auperin *et al.*, 1984a; Bishop, 1990), sería el mRNA de la proteína N, aunque esto aún no ha sido comprobado experimentalmente. En conjunción con esta información, se ha descrito la presencia de un RNA de bajo peso molecular en preparaciones de viriones de los virus Tacaribe y LCM, en cantidades molares comparables al RNA S viral. La caracterización del mismo permitió comprobar que corresponde al mRNA transcrito del gen Z. Esto podría llevar a pensar que se estaría en presencia de un sistema de traducción detenido y, por lo tanto, existiría una asociación con los ribosomas. En el momento de la infección, este sistema se activaría utilizando factores celulares, constituyendo así, un evento muy temprano dentro de los fenómenos de transcripción y replicación virales (Iapalucci *et al.*, 1989a; Salvato and Shimomaye, 1989).

Por otra parte, en todas las preparaciones virales se encuentran otras especies de RNA que acompañan al genoma. Estos RNAs, inicialmente caracterizados por su coeficiente de sedimentación, 4-6S, 18S y 28S, corresponden a rRNAs, cuyo origen se ha asignado a los ribosomas presentes en el virión (Abelson *et al.*, 1969; Pedersen, 1970; Carter *et al.*, 1973; Añón *et al.*, 1976; Vezza *et al.*, 1977; Ramsingh *et al.*, 1980; Dutko *et al.*, 1981). Como ya se indicó, los ribosomas no parecen tener ninguna función en el ciclo replicativo viral, y su presencia dentro de los virus, podría deberse a un sistema de encapsidación y brotación poco específico.

Sin embargo, ninguna de las publicaciones anteriores resulta concluyente, y se mantiene abierta la discusión sobre el papel de los ribosomas en las partículas virales de los

arnavirus. También se mantiene subyacente otra pregunta: ¿además de ribosomas completos, no existirán complejos ribonucleoproteicos formados por los rRNAs y alguna proteína, viral o celular?

Discusión sobre los RNAs de arnavirus, ribosomas y RNAs ribosomales

En 1976, Añón *et al.* muestran resultados que plantean dudas acerca de la aleatoriedad de la presencia de los rRNAs 18S y 28S en las preparaciones de RNAs virales de arnavirus. En particular, los autores observaron que, en cultivos de células BHK21 infectadas con virus Junín y crecidas en presencia de actinomicina D (inhibición ~ 94 %), y (³H) U, la producción de viriones no se veía afectada. Un aspecto adicional, y realmente llamativo, de sus resultados está constituido por el hecho que, aún cuando en esas condiciones de inhibición la síntesis de rRNAs debería ser muy escasa y, por lo tanto, se esperaría una pequeña cantidad de (³H) U en los rRNAs copurificados con los RNAs virales, los autores encuentran hasta un 60 % de (³H) U en los rRNAs, relativa a la (³H) U incorporada en el RNA L viral.

Los datos anteriores plantean dos aspectos centrales:

- ⇒ La incorporación de ribonucleoproteínas en los viriones brotantes parece no ser tan aleatoria.
- ⇒ La incorporación de rRNAs en los viriones parece tener cierta selectividad dirigida hacia las moléculas sintetizadas *de novo*.

Por otra parte, tampoco se puede descartar la posible participación de los ribosomas, o alguno de sus componentes (proteínas y/o ácidos nucleicos), en el proceso de desencapsidación (*uncoating*) de las nucleocápsides.

Desde este punto de vista, en varios virus con genoma de RNA positivo se ha verificado la necesidad de una interacción entre los ribosomas celulares y las nucleocápsides para asegurar la desencapsidación de las

mismas. Este evento, en algunos casos, ha sido asociado con el inicio de la traducción.

En el virus del mosaico del tabaco (TMV), la región 5' *leader* es desnudada rápidamente, se inicia la traducción y la mayor parte del RNA viral va siendo desencapsidado en dirección 5' → 3' al mismo tiempo que avanza la traducción (Mundry *et al.*, 1991).

En 1996, Wu & Shaw, agregan información adicional, demostrando que la desencapsidación co-traducciona avanza casi hasta el final del molde, momento en el cual comienza un proceso de desencapsidación, concomitante con la replicación, en la dirección 3' → 5'.

Por otra parte, en el género *Alphavirus* (familia *Togaviridae*), se ha descrito una participación mucho más activa de los ribosomas en el proceso de desencapsidación de las nucleocápsides. En 1984, Wengler & Wengler, publicaron un trabajo en el cual demostraron que las nucleocápsides del virus Sindbis recién ingresadas en la célula infectada se asocian rápidamente con ribosomas. En 1992, Singh & Helenius, publicaron un trabajo sobre el virus Semliki Forest, en el cual demostraron, en ensayos *in vitro*, que los ribosomas son necesarios y suficientes para producir la desencapsidación, mediante una asociación irreversible entre la proteína C (proteína de cubierta) y la subunidad mayor del ribosoma. En este trabajo, y en otros posteriores (Wengler *et al.*, 1996; Singh *et al.*, 1997), los autores han verificado experimentalmente la existencia de complejos proteína C-ribosomas dentro de las células infectadas.

La responsabilidad de la formación de complejos proteína C-subunidad ribosomal 60S, ha sido atribuida a la interacción entre la proteína C y una región del rRNA 28S que presenta un 54% de homología con el RNA de los alfavirus.

En función de todas las consideraciones anteriores, y buscando conjugar observaciones previas sobre arnavirus (Abelson *et al.*, 1969;

Pedersen, 1970; Carter *et al.*, 1973; Añón *et al.*, 1976; Vezza *et al.*, 1977; Ramsingh *et al.*, 1980; Dutko *et al.*, 1981; Lukashevich *et al.*, 1984; Rivera Pomar *et al.*, 1991) con nuevas interpretaciones, se realizó una búsqueda de similitud entre los RNA S de los arenavirus y los rRNAs ribosomales 28S y 18S. Para ello, se utilizó un esquema de análisis de similitud basado en la técnica de *dot plot*, utilizando el programa DotPlot v1.0 (Nakisa, 1993).

Como se puede observar en la **Figura 12.65.(A - K)**, en todos los arenavirus, la región intergénica presenta una alta frecuencia de repeticiones imperfectas distribuidas a lo largo de todo el rRNA 28S humano. Los histogramas respectivos muestran la frecuencia de repeticiones imperfectas con una identidad superior al 50% (ver pie de figura). En la **Figura 12.65.L** se muestra un gráfico de curvas superpuestas correspondientes a distintas comparaciones entre la región intergénica del RNA S de la cepa Candid #1 del virus Junín y otros RNAs (rRNA 28S humano, rRNA 18S humano y el propio RNA viral de Candid #1); como se puede observar, la hiperabundancia de repeticiones imperfectas de la región intergénica de CDI en el rRNA 28S humano es muy significativa. (Ver también **Apéndice 11**).

Para confirmar la información obtenida, se realizó un análisis complementario buscando la existencia de regiones de homología de 7 residuos o más, con homología mayor al 90%, entre el RNA S de la cepa Candid #1 del virus Junín (polaridad viral y viral complementario) y los rRNAs 28S y 18S. Para ello, se utilizó una implementación para PC del programa BLAST2SEQ (Altschul *et al.*, 1990). En las **Figuras 12.66 y 12.67** (ver también **Apéndice 11**) se pueden observar los resultados de estos análisis. Nuevamente, se encontró una significativa hiperabundancia de repeticiones correspondientes específicamente a la región de las estructuras tipo *hairpin loop* intergénicas de Candid #1. La mayor abundancia de repeticiones en el par RNA S de Candid

#1 /rRNA 28S humano vs. RNA S de Candid #1/rRNA 18S humano, puede deberse, no sólo a la diferencia de longitud (rRNA 28S: 5035 nt. y rRNA 18S: 1871 nt.), sino también a la diferencia en contenido de GC (rRNA 28S: 69,2%; rRNA 18S: 56,1%; RNA S Candid #1: 44,6%). Los resultados obtenidos con los análisis anteriores, permiten elaborar hipótesis adicionales sobre los mecanismos de incorporación de complejos ribonucleoproteicos (ribosomas y nucleocápsides) en los viriones brotantes de los arenavirus.

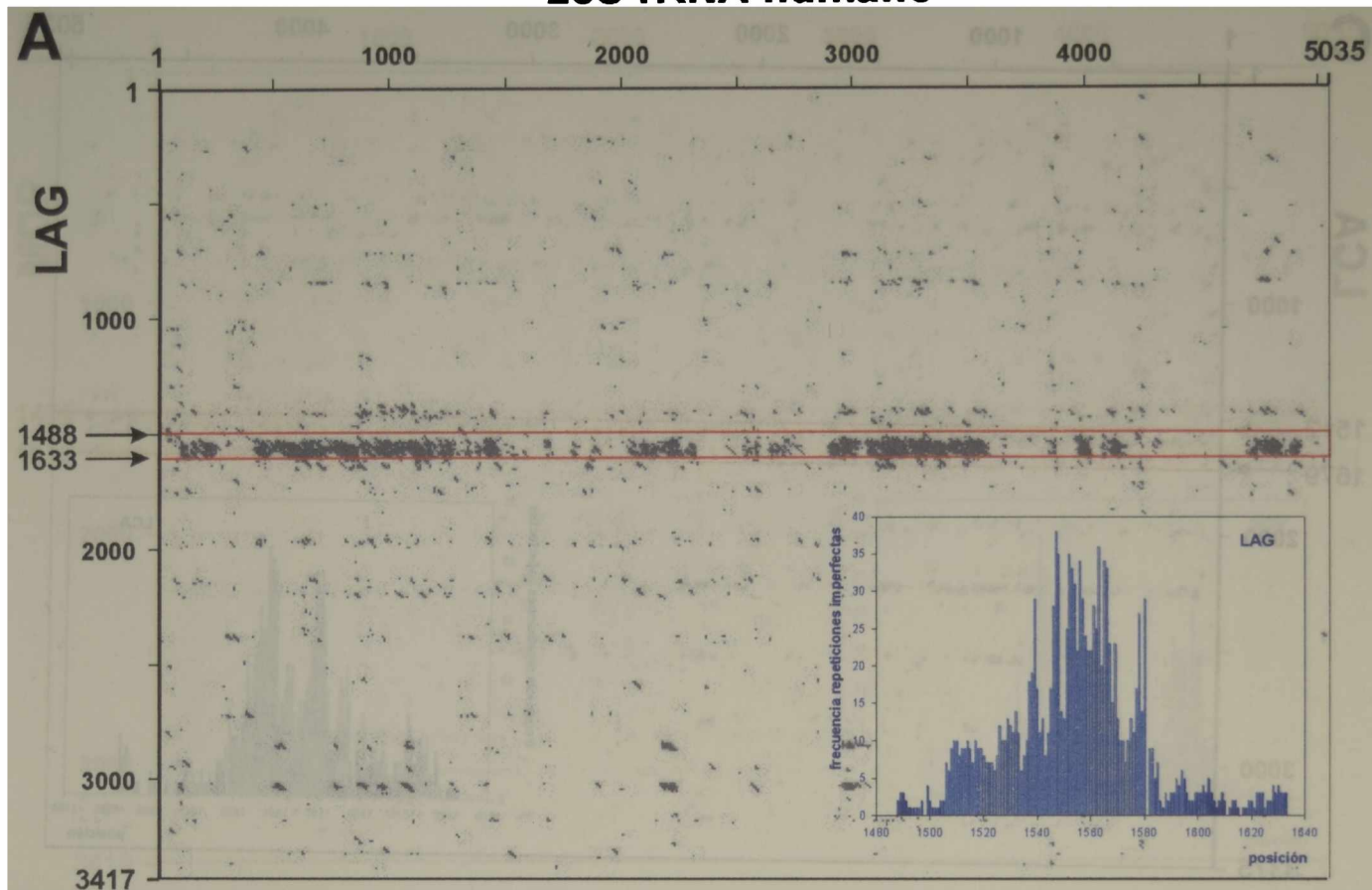
Teniendo en cuenta la disposición perinuclear (mayor concentración local) de la proteína N en células CV1 infectadas con la cepa MC2 del virus Junín (**Figura 12.68.A y B**), es probable que la misma pueda interactuar con distintas regiones del rRNA 28S (y/o el rRNA 18S) accesibles en los ribosomas inmaduros (ensamblados o desensamblados), de reciente formación, que emergen por los poros nucleares; o accesibles en complejos ribosoma-mensajero viral (por ejemplo el mRNA codificante para Z).

Estos complejos proteína N-ribosomas (o subunidades ribosomales) podrían persistir en una cierta proporción durante el resto del ciclo viral. Además, la existencia de estos complejos permitiría una relativa selectividad en la incorporación de ribosomas, con rRNAs sintetizados *de novo*, en los viriones brotantes, de la misma manera que se incorporan con cierta selectividad las nucleocápsides (**Figura 12.68.C**).

Por otra parte, la posible capacidad de interacción proteína N-ribosomas facilitaría la desencapsidación de los RNAs virales una vez ingresados en la célula. Una vez liberadas las nucleocápsides en el citoplasma, los ribosomas de la célula infectada pueden secuestrar la proteína N asociada con los RNAs virales.

Obviamente, el modelo anteriormente presentado es una explicación potencial que debería ser confirmada experimentalmente.

28S rRNA humano



28S rRNA humano

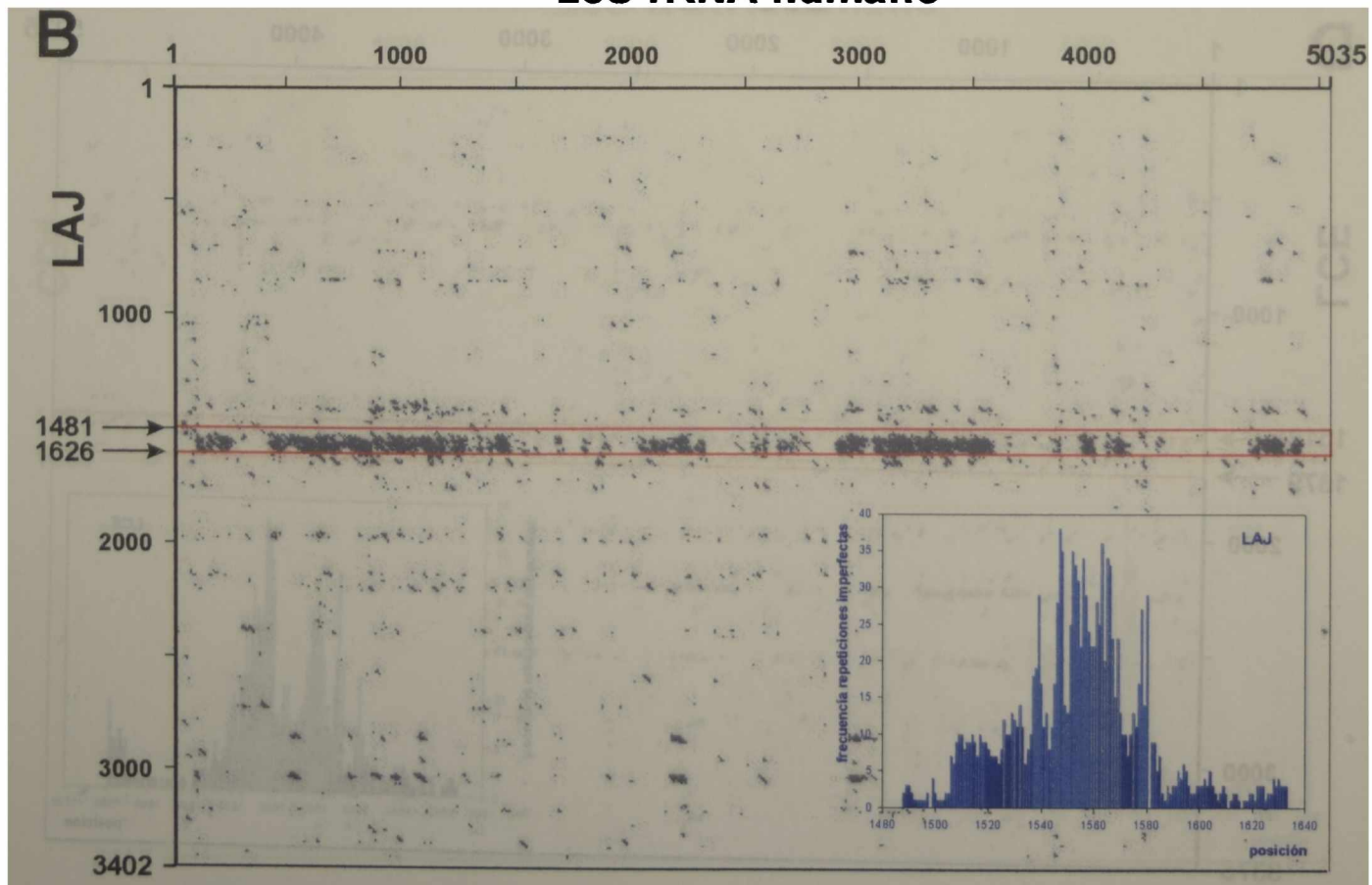
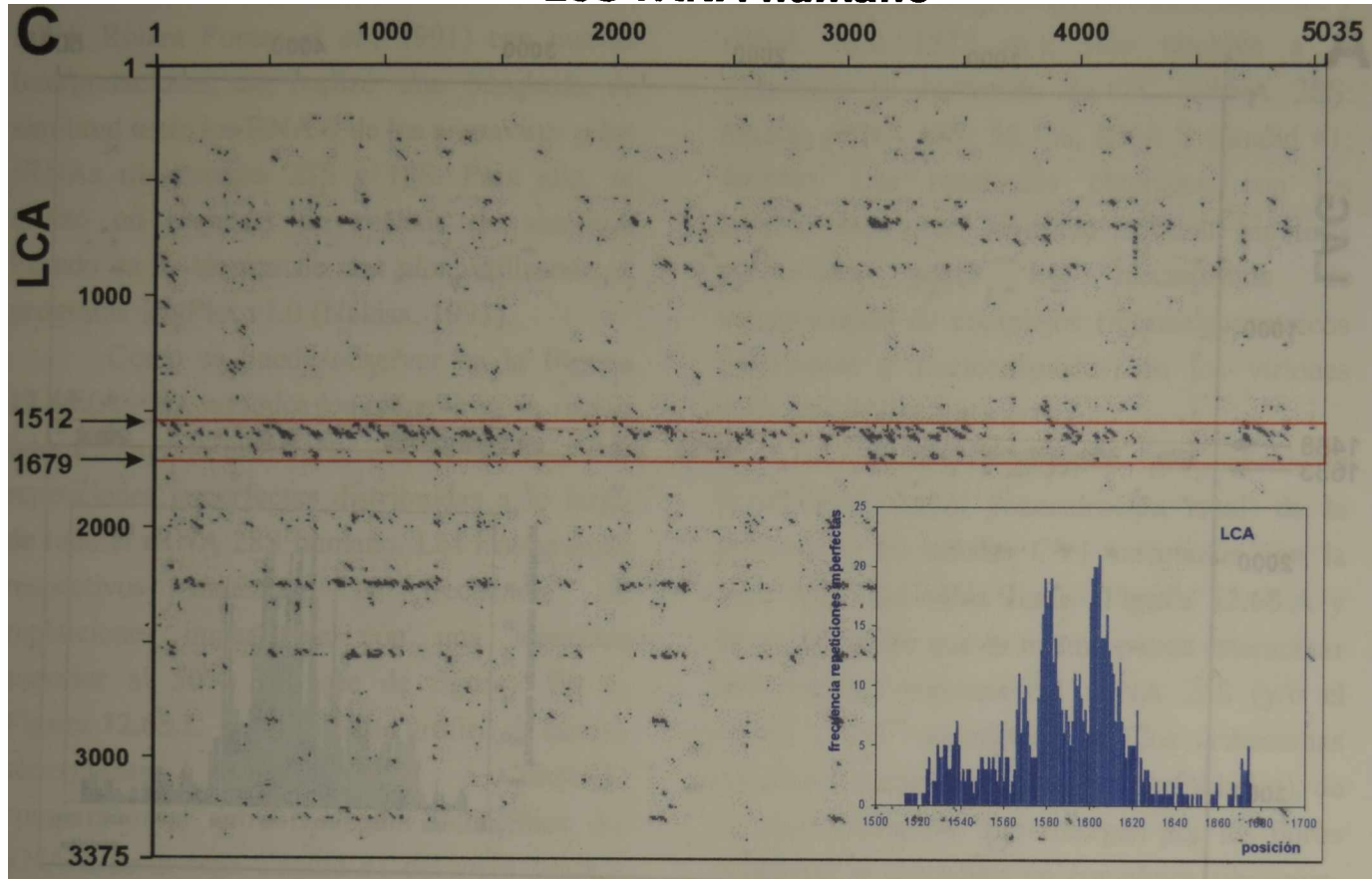


Figura 12.65. (Ver pie de figura en página 196)

28S rRNA humano



28S rRNA humano

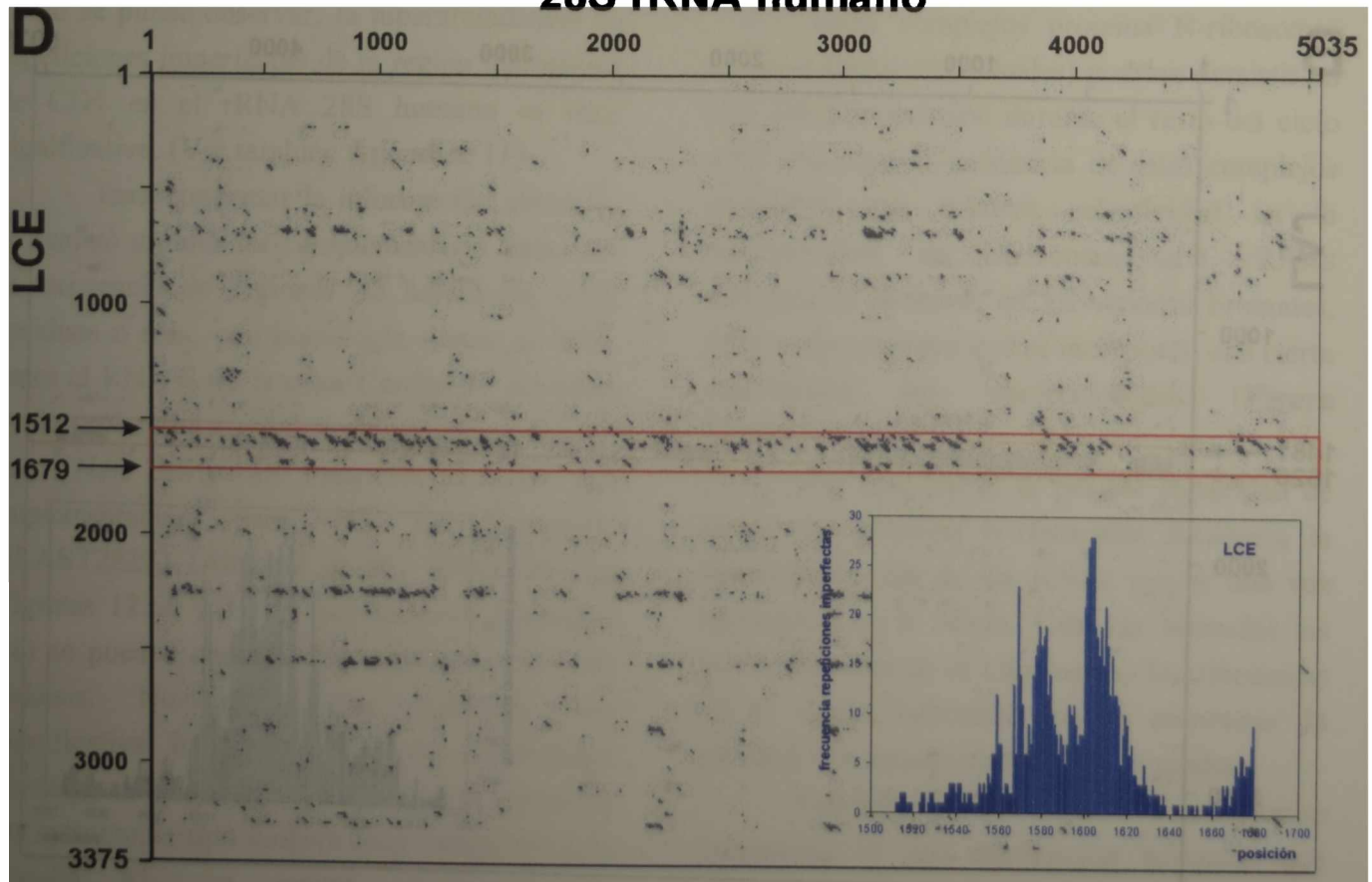
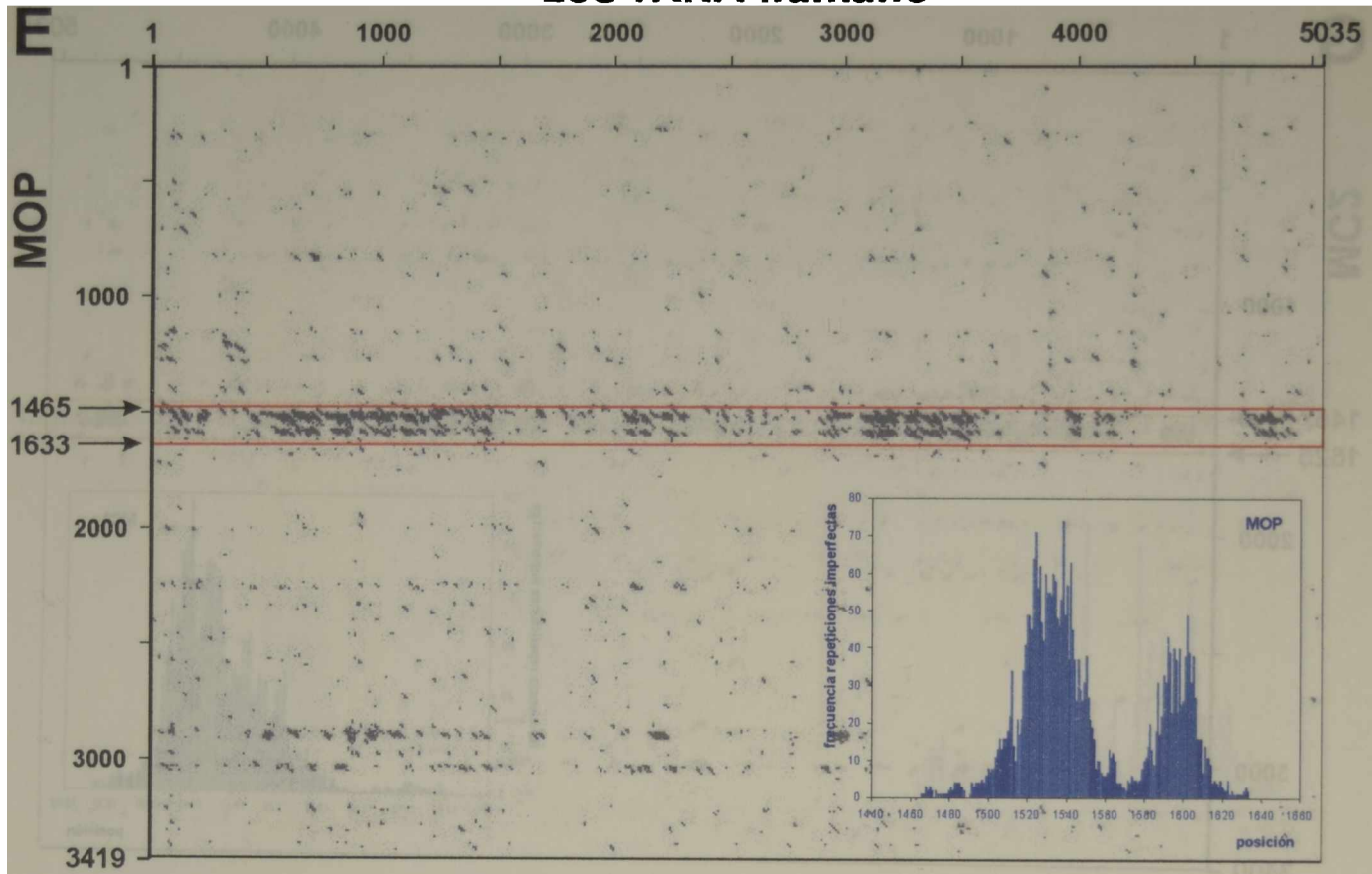
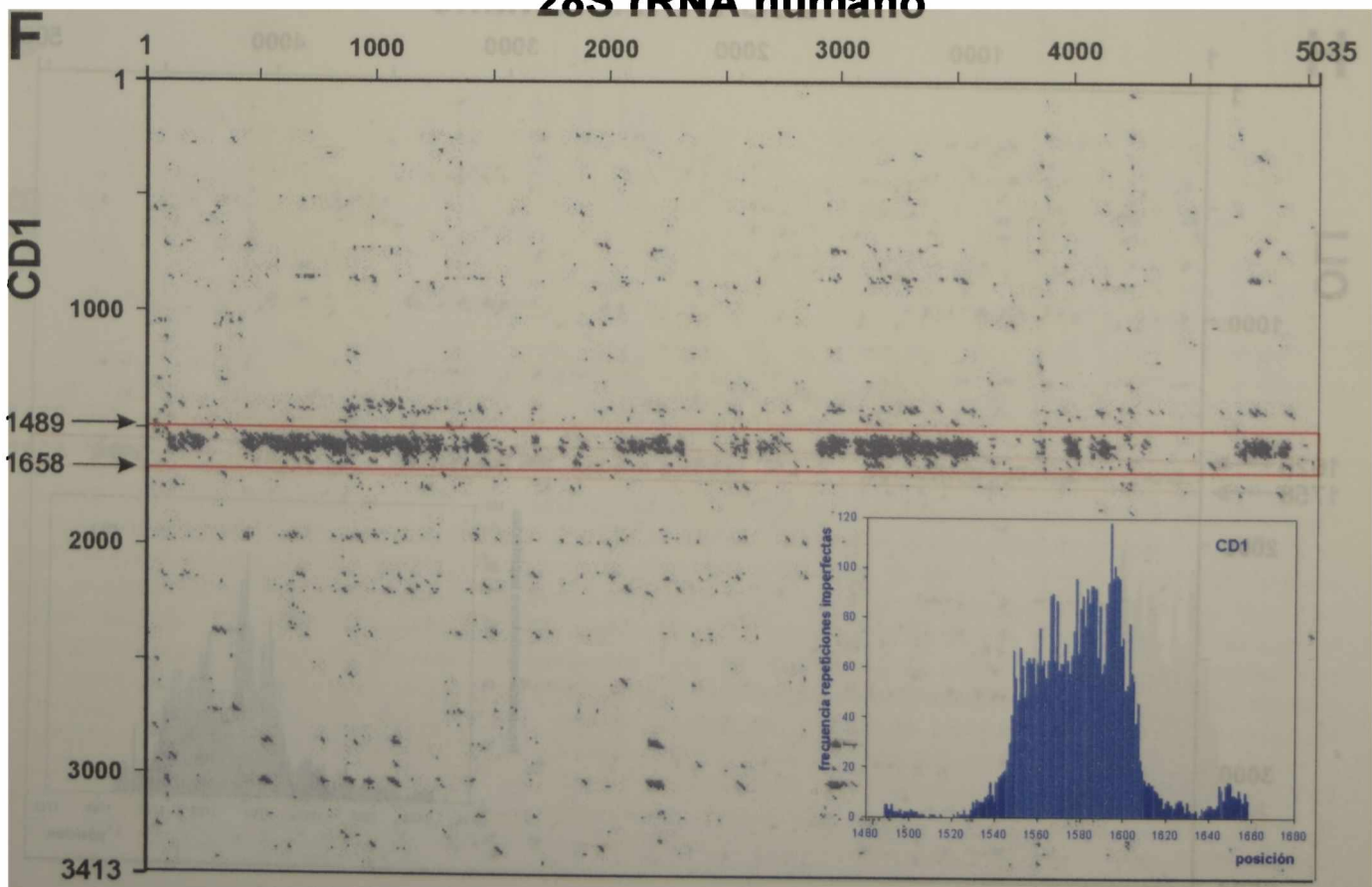


Figura 12.65. (Ver pie de figura en página 196)

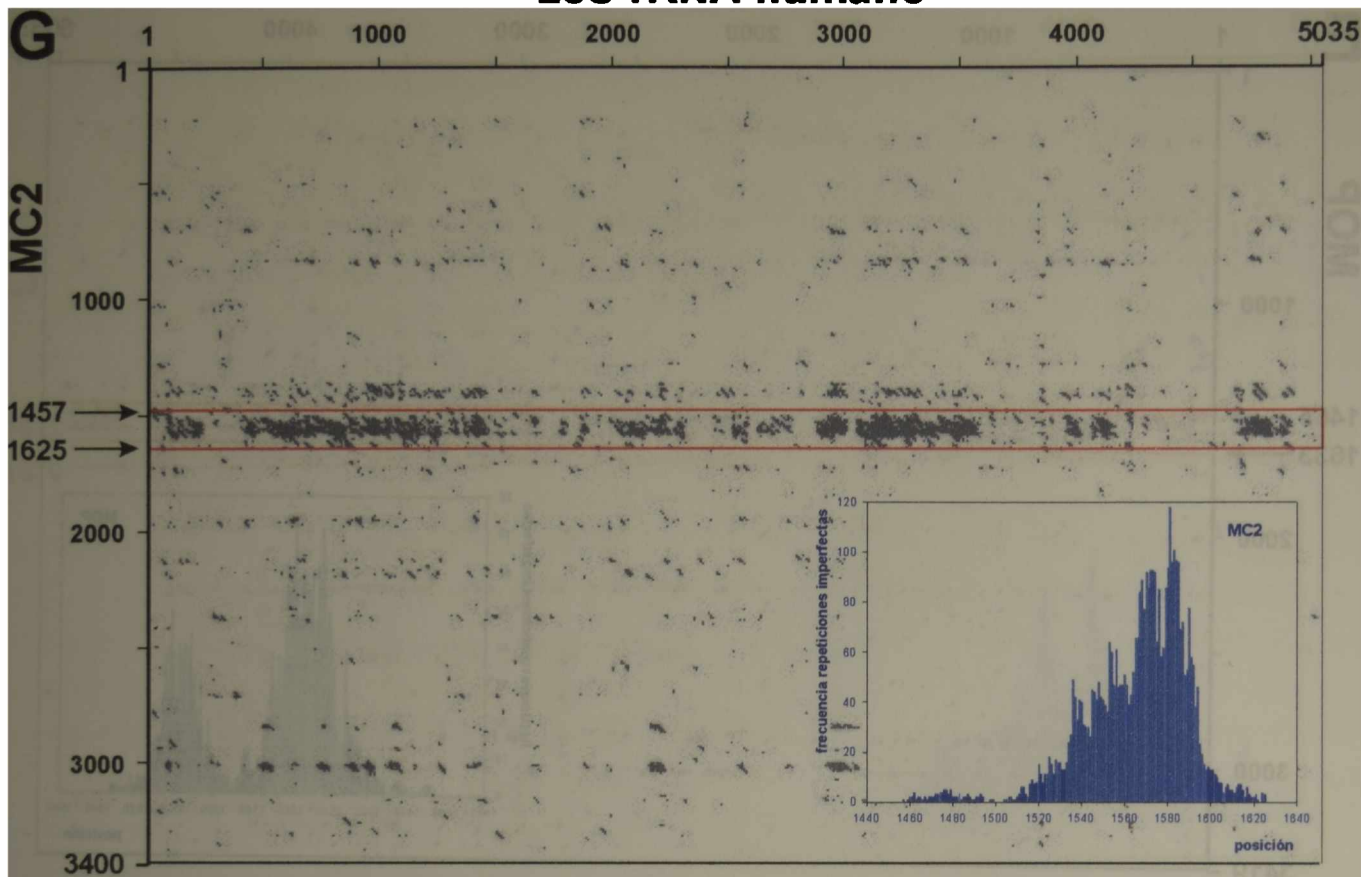
28S rRNA humano



28S rRNA humano



28S rRNA humano



28S rRNA humano

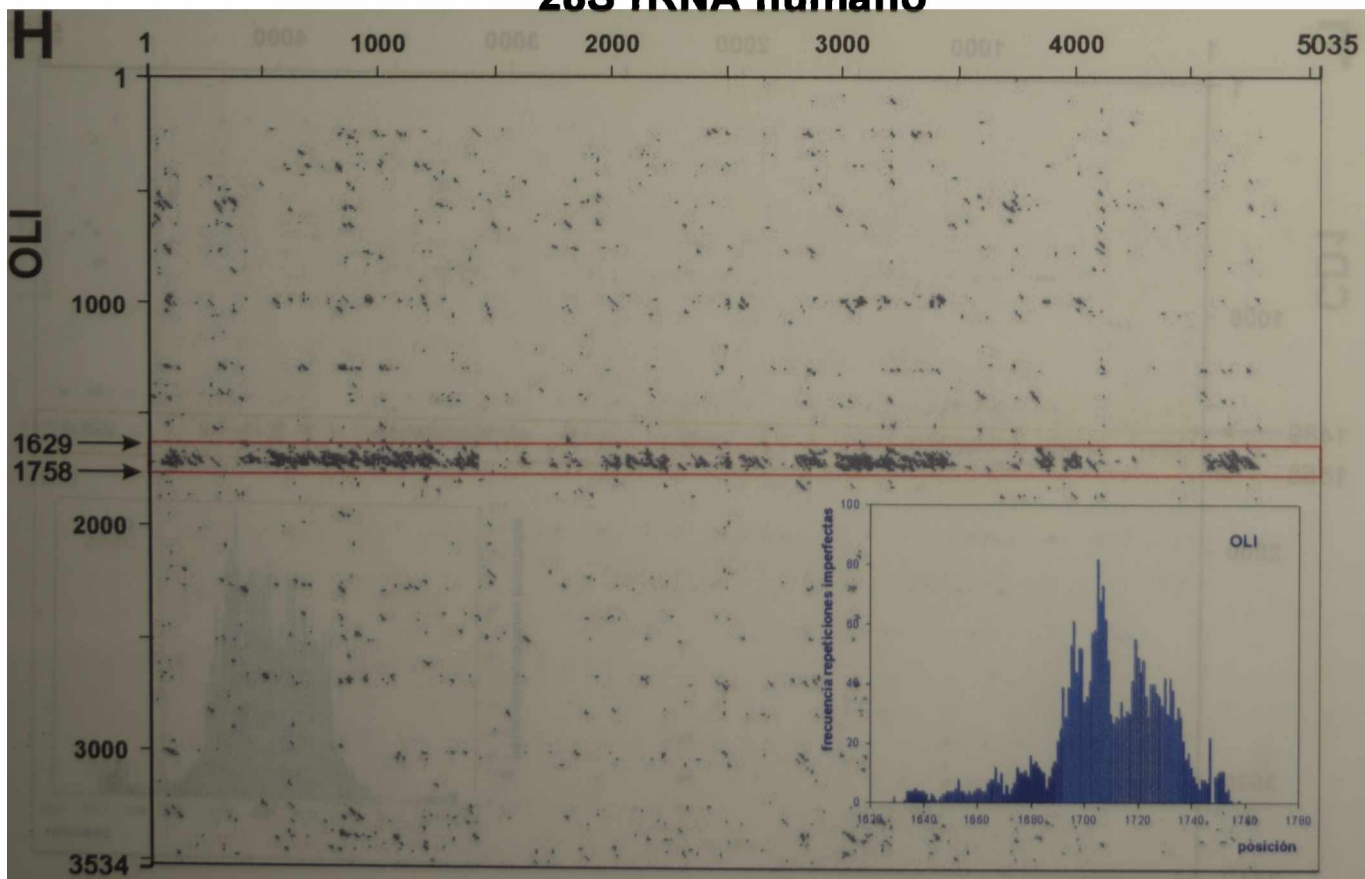
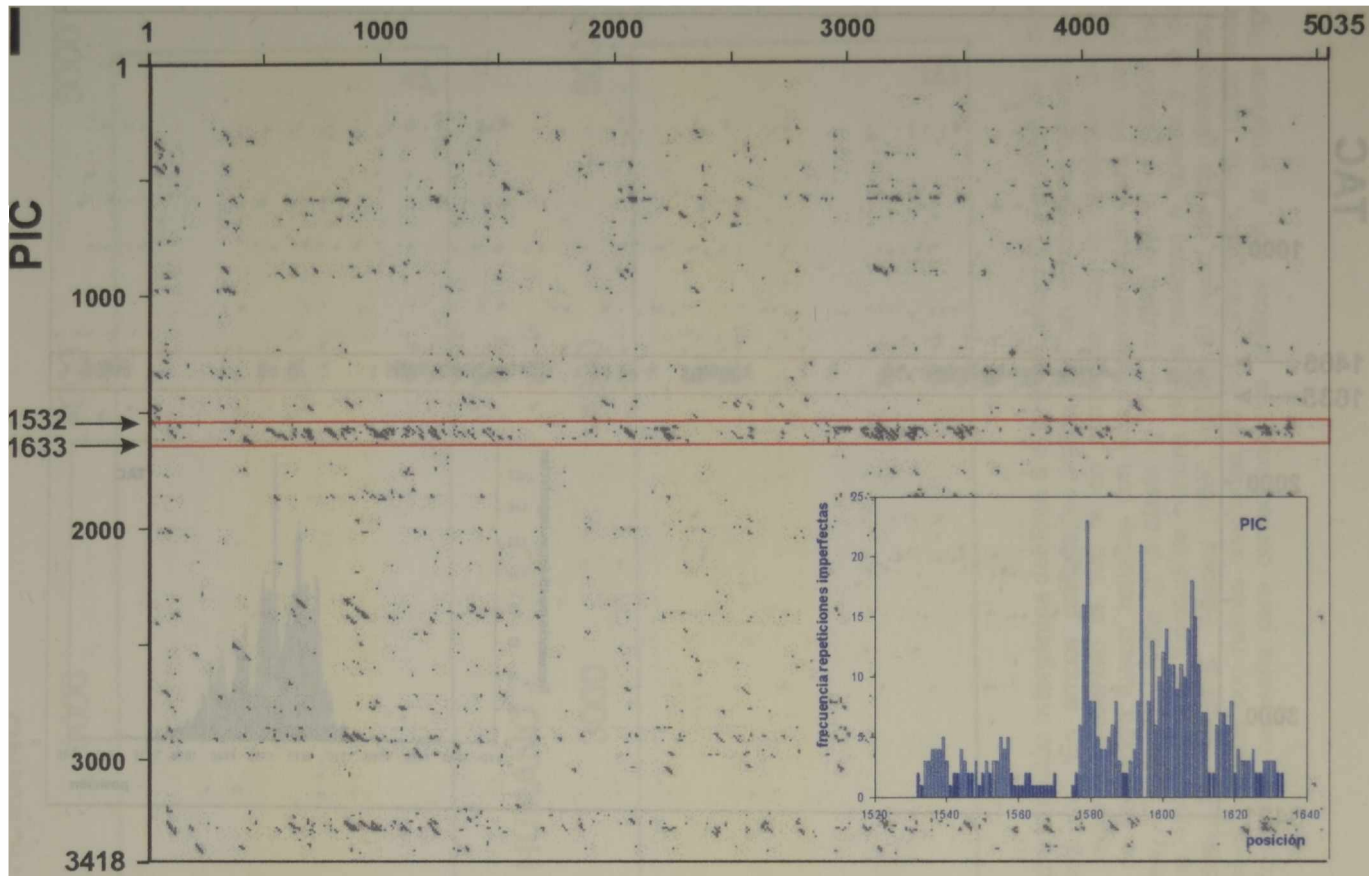


Figura 12.65. (Ver pie de figura en página 196)

28S rRNA humano



28S rRNA humano

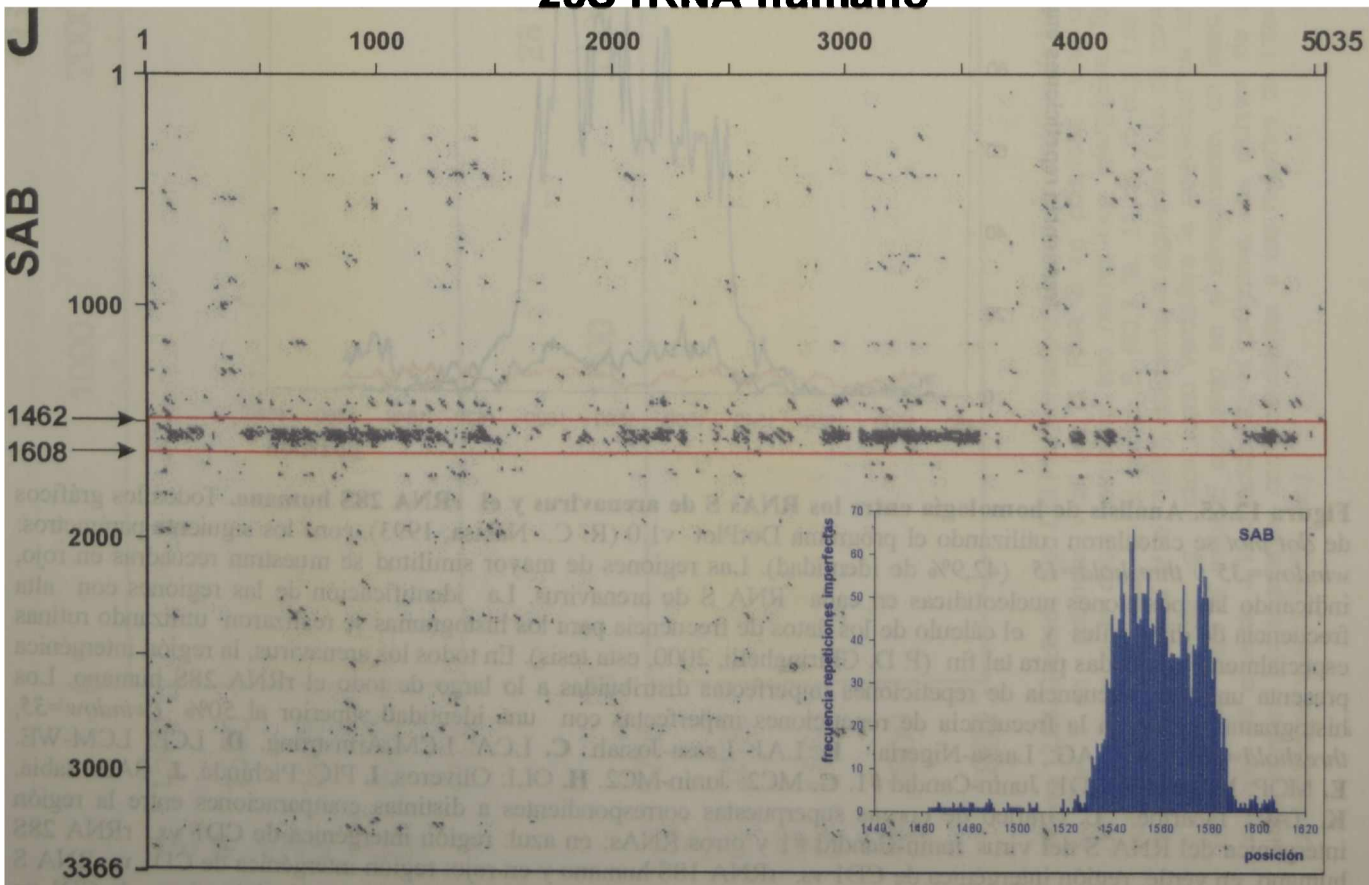


Figura 12.65. (Ver pie de figura en página 196)

28S rRNA humano

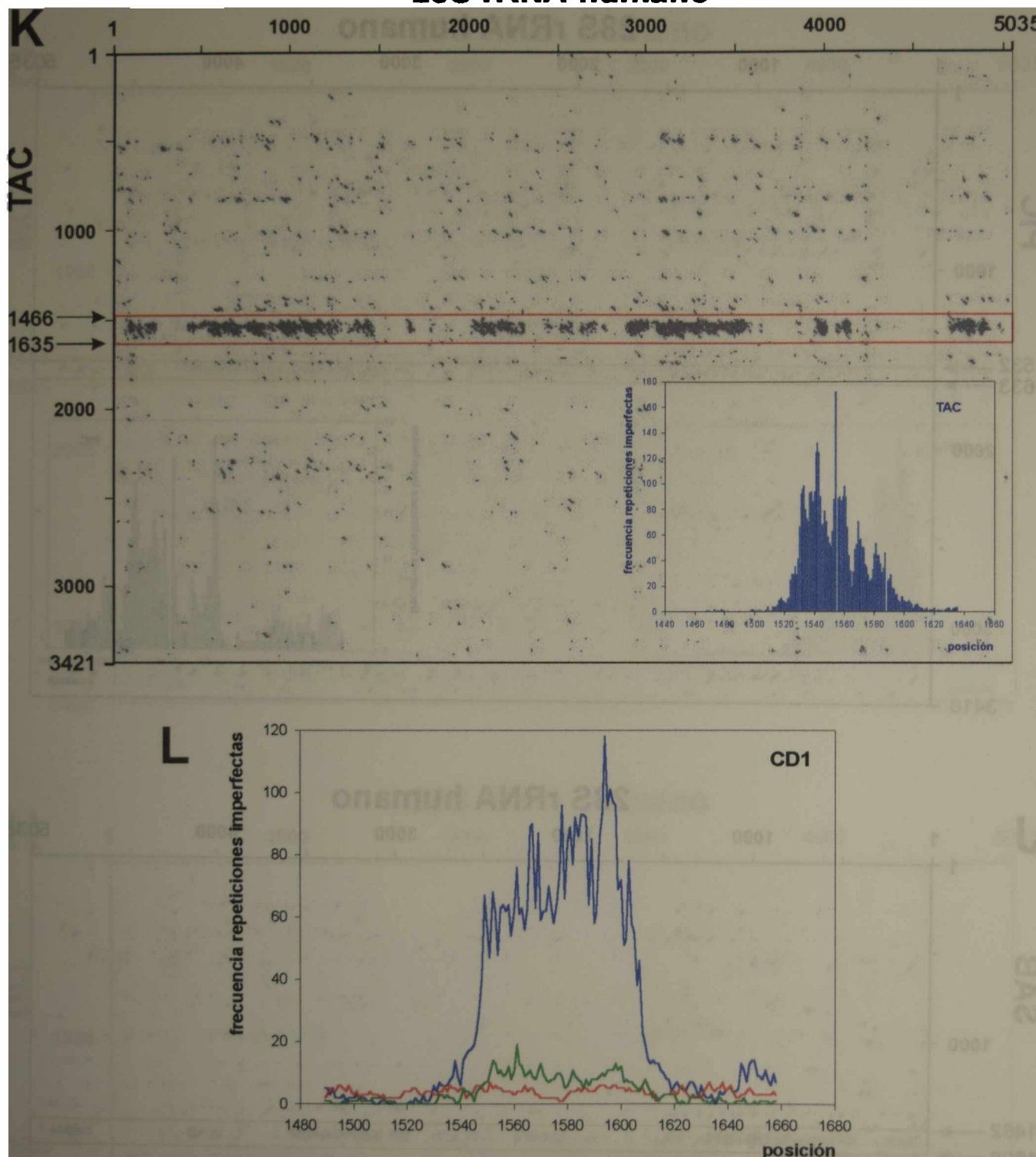


Figura 12.65. Análisis de homología entre los RNAs S de arnavirus y el rRNA 28S humano. Todos los gráficos de dot plot se calcularon utilizando el programa DotPlot v1.0 (R. C. Nakisa, 1993), con los siguiente parámetros: *window*=35 y *threshold*=15 (42,9% de identidad). Las regiones de mayor similitud se muestran recuadradas en rojo, indicando las posiciones nucleotídicas en cada RNA S de arnavirus. La identificación de las regiones con alta frecuencia de diagonales y el cálculo de los datos de frecuencia para los histogramas se realizaron utilizando rutinas especialmente diseñadas para tal fin (P. D. Ghiringhelli, 2000, esta tesis). En todos los arnavirus, la región intergénica presenta una alta frecuencia de repeticiones imperfectas distribuidas a lo largo de todo el rRNA 28S humano. Los histogramas muestran la frecuencia de repeticiones imperfectas con una identidad superior al 50% (*window*=35, *threshold*=18). **A.** LAG: Lassa-Nigeria. **B.** LAJ: Lassa-Josiah. **C.** LCA: LCM-Armstrong. **D.** LCE: LCM-WE. **E.** MOP: Mopeia. **F.** CD1: Junín-Candid #1. **G.** MC2: Junín-MC2. **H.** OLI: Oliveros. **I.** PIC: Pichindé. **J.** SAB: Sabiá. **K.** TAC: Tacaribe. **L.** Gráfico de curvas superpuestas correspondientes a distintas comparaciones entre la región intergénica del RNA S del virus Junín-Candid #1 y otros RNAs; en azul: región intergénica de CD1 vs. rRNA 28S humano, en verde: región intergénica de CD1 vs. rRNA 18S humano y en rojo: región intergénica de CD1 vs. RNA S de CD1. Como se puede observar la hiperabundancia de repeticiones imperfectas de la región intergénica de CD1 en el rRNA 28S humano es significativa.

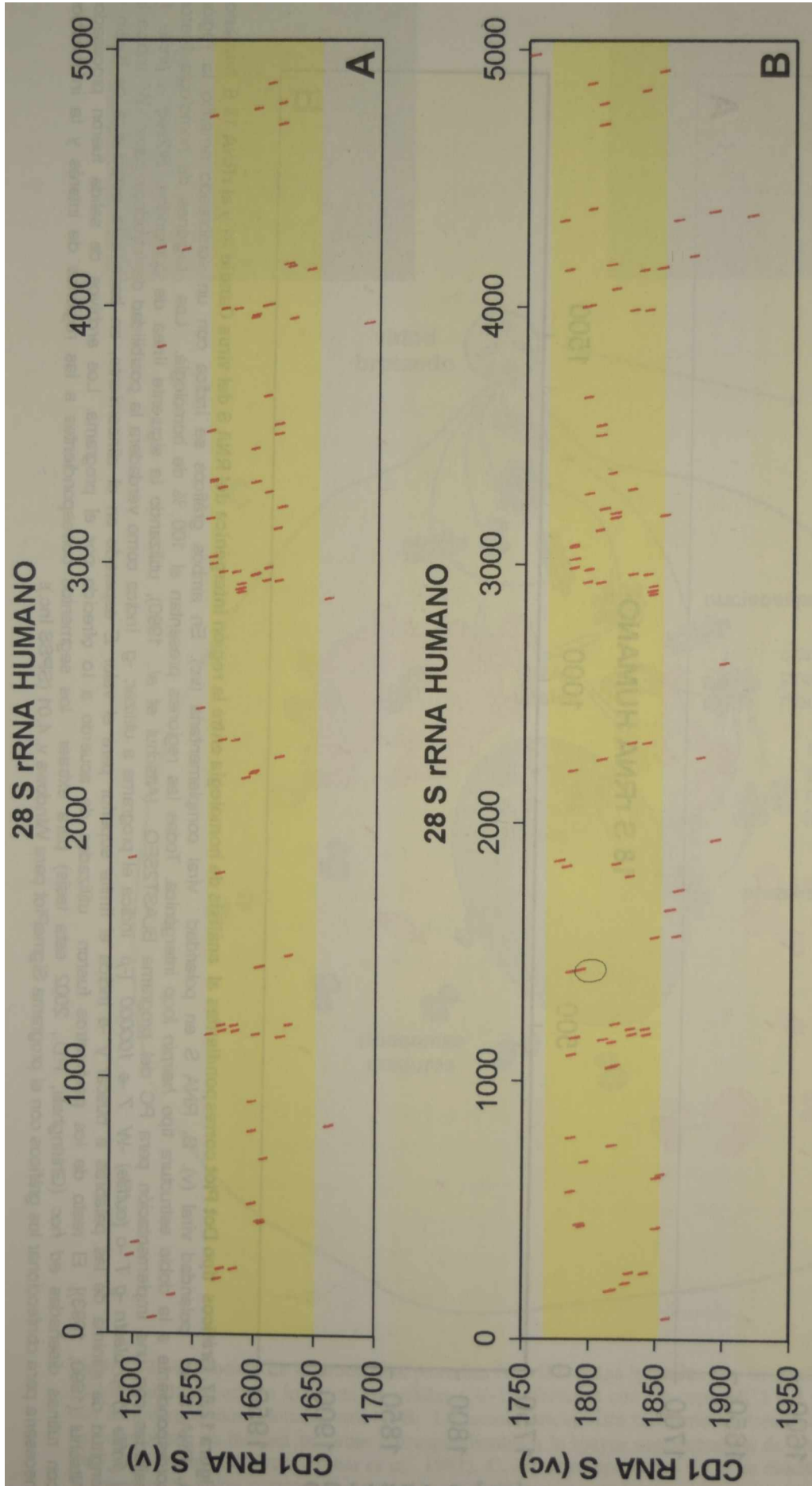


Figura 12.66. Gráficos tipo Dot Plot correspondientes al análisis de homología entre la región Intergénica del RNA S del virus Candid #1 y el rRNA 28 S humano.
A. RNA S en polaridad viral (v). **B.** RNA S en polaridad viral complementaria (vc). En ambos gráficos se indica con un sombreado amarillo la región correspondiente a la doble estructura tipo *hairpin loop* intergénica. Todas las regiones de homología presentan el 100 % de homología, con la excepción de la región señalada con un círculo en el gráfico B (93,3 %, 14 de 15 nt.). Las regiones de homología fueron halladas con una implementación para PC del programa BLAST2SEQ (Atschul *et al.*, 1990), utilizando la siguiente línea de comandos: *Bl2seq -i (infile 1) -j (infile 2) -p blastn -g T -o (outfile) -W 7 -e 100000* [-p indica el programa a utilizar; -g indica como verdadera la posibilidad de introducir *gaps*; -W indica la longitud de mínima de las palabras a buscar y -e indica el límite superior para el valor E esperado en el alineamiento, en base a la estadística de Karlin & Altschul (1990, 1993)]. El resto de los parámetros fueron utilizados de acuerdo a lo ofrecido por el programa. Los archivos de salida fueron procesados con rutinas diseñadas *ad hoc* (Ghinghelli, P.D., 2002 esta tesis) para extraer los segmentos correspondientes a las regiones de interés y la información necesaria para confeccionar los gráficos con el programa SigmaPlot para Windows v. 4.01 (SPSS Inc.).

18 S rRNA HUMANO

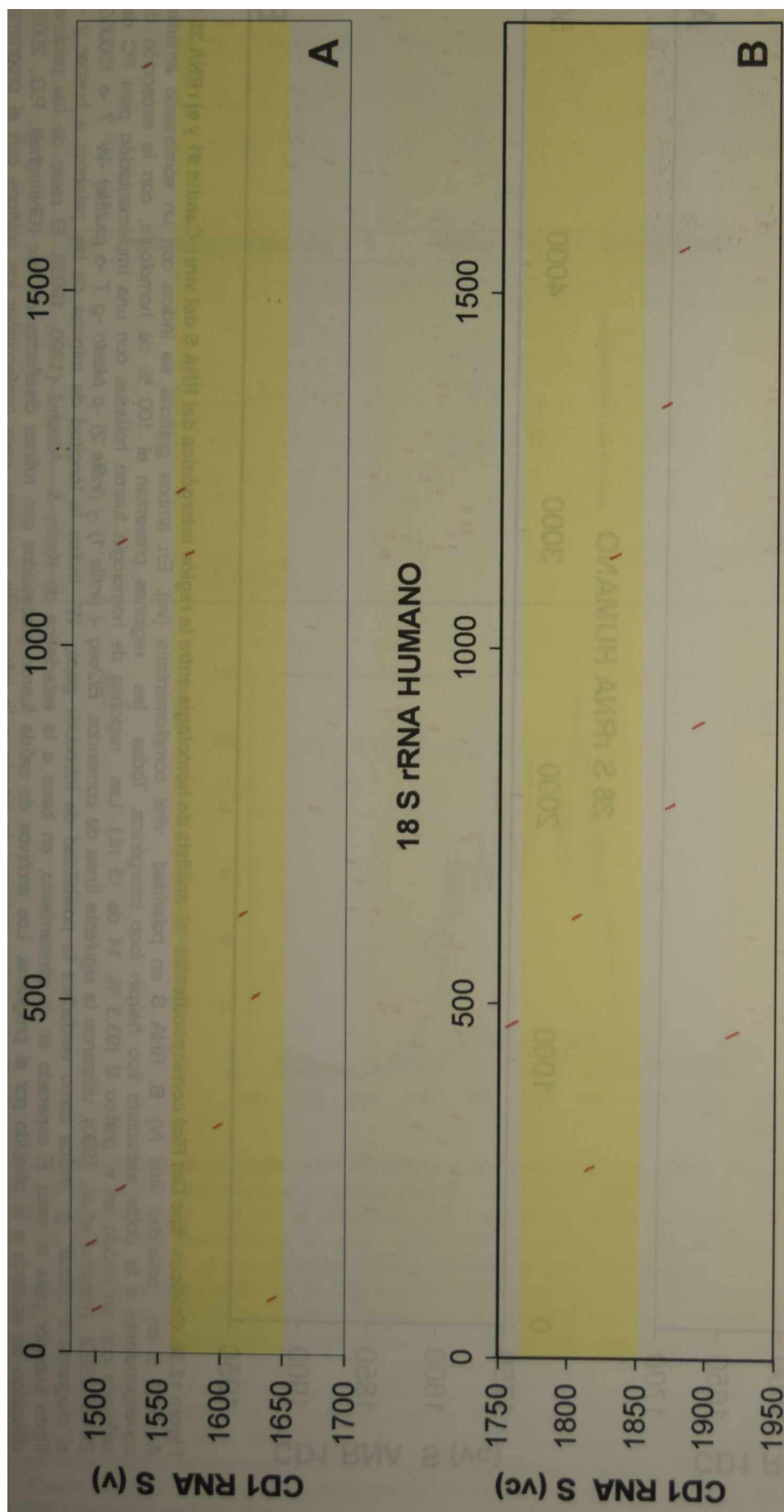


Figura 12.67. Gráficos tipo Dot Plot correspondientes al análisis de homología entre la región intergénica del RNA S del virus Candid #1 y el rRNA 18 S humano. A. RNA S en polaridad viral (v). B. RNA S en polaridad viral complementaria (vc). En ambos gráficos se indica con un sombreado amarillo la región correspondiente a la doble estructura tipo *hairpin loop* intergénica. Todas las regiones presentan el 100 % de homología. Las regiones de homología fueron halladas con una implementación para PC del programa BLAST2SEQ (Altschul et al., 1990), utilizando la siguiente línea de comandos: *Bl2seq -i (infile 1) -j (infile 2) -p blastn -g T -o (outfile) -W 7 -e 100000 [-p indica el programa a utilizar, -g indica como verdadera la posibilidad de introducir gaps; -W indica la longitud de mínima de las palabras a buscar y -e indica el límite superior para el valor E esperado en el alineamiento, en base a la estadística de Karlin y Altschul (1990, 1993)]. El resto de los parámetros fueron utilizados de acuerdo a lo ofrecido por el programa. Los archivos de salida fueron procesados con rutinas diseñadas *ad hoc* (Ghiringhelli, P.D., 2002 esta tesis) para extraer los segmentos correspondientes a las regiones de interés y la información necesaria para confeccionar los gráficos con el programa Sigma Plot para Windows v. 4.01 (SPSS Inc.).*

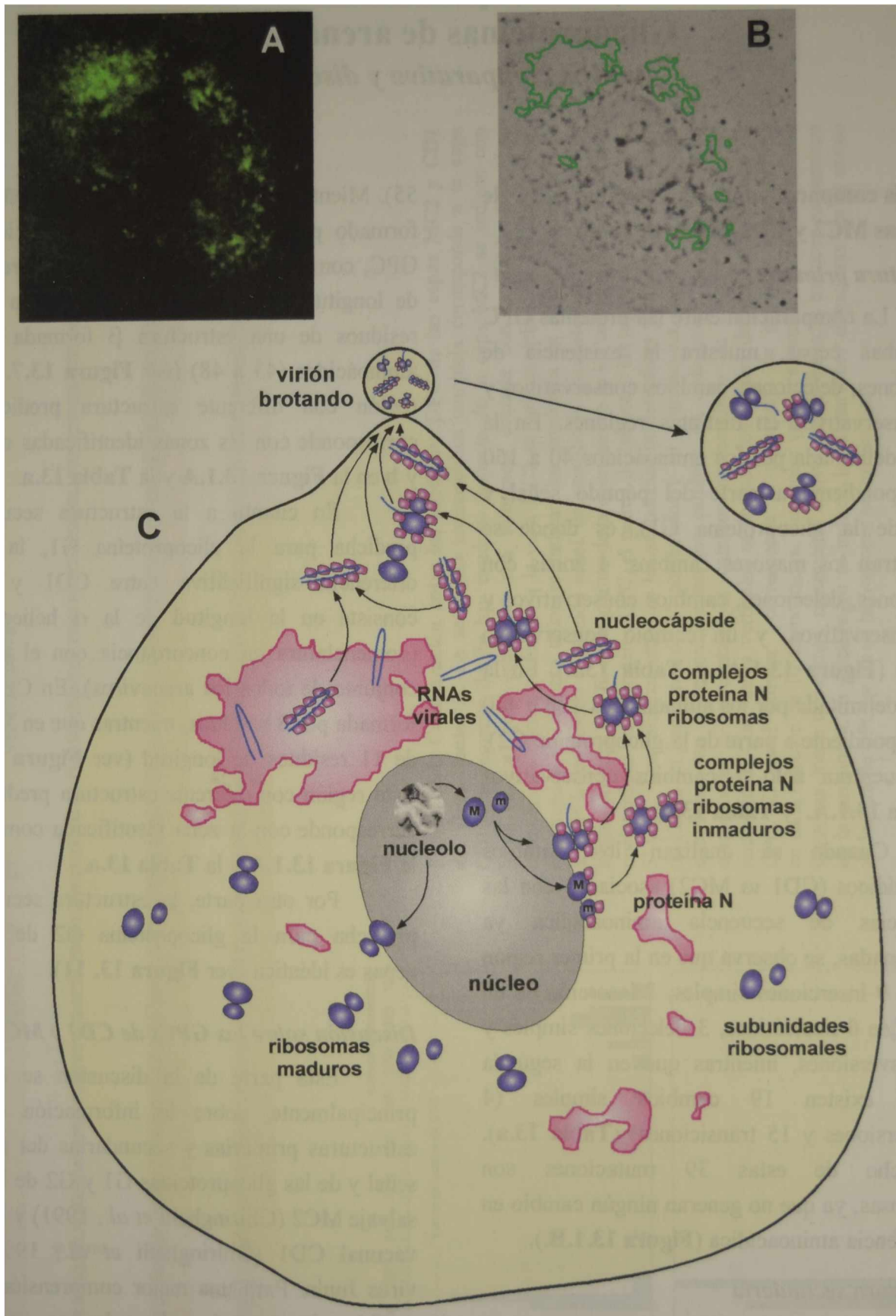


Figura 12.68. Modelo de la asociación proteína N - ribosomas inmaduros y brotación en los arenavirus. **A.** Inmunofluorescencia indirecta de células CV-1 infectadas con la cepa MC2 del virus Junin, reveladas con anticuerpos monoclonales anti-N. **B.** La misma célula vista mediante contraste de fase. Con contornos de color verde se indican las áreas correspondientes a la mayor concentración de proteína N en la foto A (**A** y **B** tomadas de Rivera Pomar *et al.*, 1991). **C.** Modelo hipotético. En él se muestra la asociación entre la proteína N y los ribosomas inmaduros, probablemente, mediante contactos proteína N-rRNA 28S y/o proteína N-rRNA 18S. Las áreas de color rosa corresponden a las zonas de mayor concentración de proteína N en la foto A. Los corpúsculos color rosa corresponden a monómeros de proteína N. Con líneas azules se indican los RNAs genómicos virales. En violeta se muestran los ribosomas, maduros e inmaduros.

Capítulo 13

Glicoproteínas de arnavirus

Análisis comparativo y discusión

Análisis comparativo de las proteínas GPC de las cepas MC2 y CD1 del virus Junín

Estructura primaria

La comparación entre las proteínas GPC de ambas cepas, muestra la existencia de inserciones, deleciones, cambios conservativos y no conservativos en distintas regiones. En la región delimitada por los aminoácidos 40 a 150 (correspondiente a parte del péptido señal y parte de la glicoproteína G1), es donde se encuentran los mayores cambios: 4 zonas con inserciones, deleciones, cambios conservativos y no conservativos, y un cambio conservativo aislado (**Figura 13.1.A.** y **Tabla 13.a**). En la región delimitada por los aminoácidos 420 a 450 (correspondiente a parte de la glicoproteína G2), se encuentran sólo 2 cambios conservativos (**Figura 13.1.A.** y **Tabla 13.a**).

Cuando se analizan los cambios nucleotídicos (CD1 vs MC2) asociados con las diferencias de secuencia aminoacídica ya mencionadas, se observa que en la primer región existen 9 inserciones simples, 1 inserción de un bloque de 6 nucleótidos, 3 deleciones simples y 2 transversiones, mientras que en la segunda región existen 19 cambios simples (4 transversiones y 15 transiciones) (**Tabla 13.a**). Dieciocho de estas 39 mutaciones son silenciosas, ya que no generan ningún cambio en la secuencia aminoacídica (**Figura 13.1.B.**).

Estructura secundaria

Desde este punto de vista, la única diferencia significativa observada corresponde a la estructura secundaria predicha para el péptido señal. En CD1 el péptido señal está formado por los primeros 58 aminoácidos de GPC, con una α hélice de 53 residuos de longitud (residuos 3 a

55). Mientras que en MC2 el péptido señal está formado por los primeros 54 aminoácidos de GPC, con una α hélice predicha de 38 residuos de longitud (residuos 3 a 40) separada por 2 residuos de una estructura β formada por 6 aminoácidos (43 a 48) (ver **Figura 13.7.**). Esta región con diferente estructura predicha se corresponde con las zonas identificadas como **a** y **b** en la **Figura 13.1.A** y la **Tabla 13.a**.

En cuanto a la estructura secundaria predicha para la glicoproteína G1, la única diferencia significativa entre CD1 y MC2 consiste en la longitud de la α hélice $\alpha 0-1$ (nomenclatura en concordancia con el análisis conjunto de todos los arnavirus). En CD1 está formada por 8 residuos, mientras que en MC2 es de 11 residuos de longitud (ver **Figura 13.9.**). Esta región con diferente estructura predicha se corresponde con la zona identificada como **d** en la **Figura 13.1.A** y la **Tabla 13.a**.

Por otra parte, la estructura secundaria predicha para la glicoproteína G2 de ambas cepas es idéntica (ver **Figura 13. 11**).

Discusión sobre las GPCs de CD1 y MC2

Esta parte de la discusión se enfoca, principalmente, sobre la información de las estructuras primarias y secundarias del péptido señal y de las glicoproteínas G1 y G2 de la cepa salvaje MC2 (Ghiringhelli *et al.*, 1991) y la cepa vacunal CD1 (Ghiringhelli *et al.*, 1997) del virus Junín. Para una mejor comprensión de la posible importancia de algunos cambios observados, en algunos puntos se agrega información de los antecesores filogenéticos de la cepa vacunal CD1, las cepas XJ-Parodi (PARO) y XJ#44 (XJ44) del virus Junín (Albariño, 1997; Albariño *et al.*, 1997).

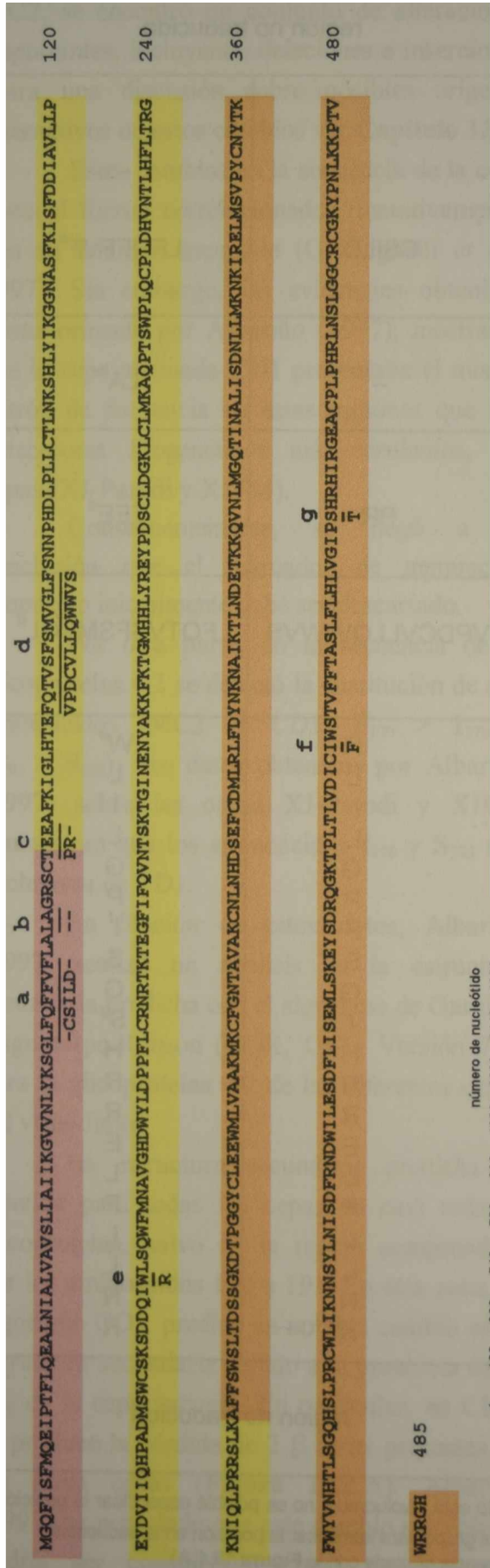
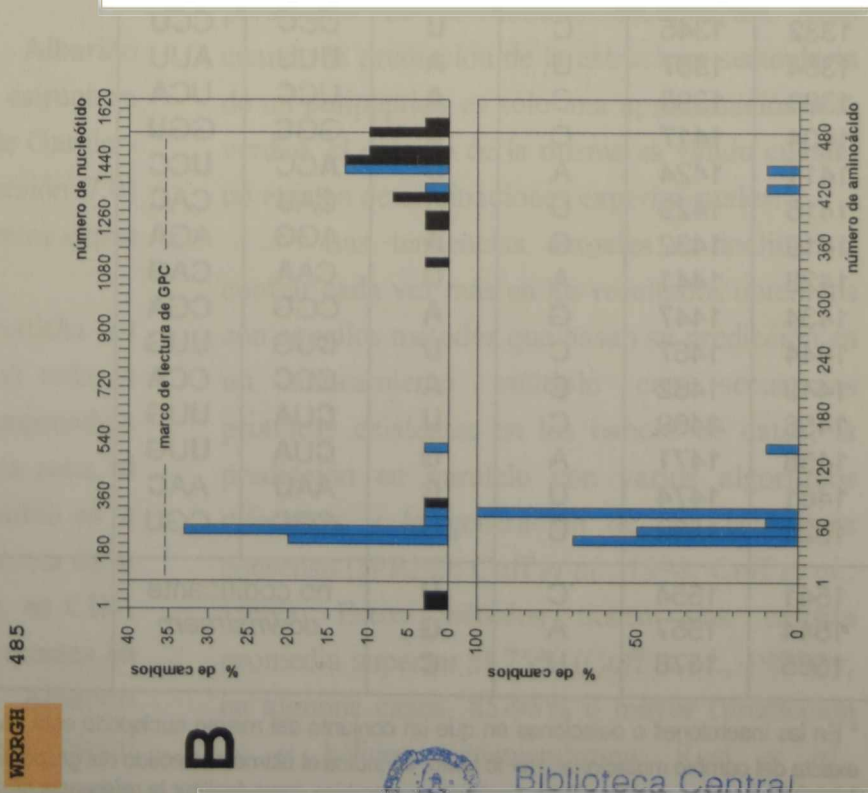


Figura 13.1. Comparación entre las proteínas GPC de las cepas MC2 y CD1 del virus Junín. A. La secuencia aminoacídica completa corresponde a la cepa CD1. Las regiones donde hay cambios respecto de la cepa MC2 se indican con letras (a,...,g) para establecer una referencia cruzada con la **Tabla 13.a**. Los aminoácidos correspondientes a la cepa MC2 se indican debajo de las líneas. Con sombreados de distinto color se indican los productos de clivaje proteolítico de las GPCs (en rosado el péptido señal, en amarillo la glicoproteína G1 y en naranja la glicoproteína G2. B. Histogramas de la distribución de los cambios aminoacídicos y nucleotídicos en la proteína GPC, el marco de lectura y las regiones no codificantes 5' y 3'. Los alineamientos de las secuencias nucleotídicas y aminoacídicas fueron divididos en fracciones de 30 nucleótidos y 10 aminoácidos, respectivamente, para establecer una correlación entre los cambios en cada nivel. El panel superior corresponde al porcentaje de cambios nucleotídicos en cada fracción de 30 nucleótidos. Con líneas verticales se indican los límites del marco de lectura de GPC. El panel inferior corresponde al porcentaje de cambios aminoacídicos en cada fracción de 10 aminoácidos. Con color celeste en ambos paneles se indican las zonas donde uno o más mutaciones a nivel nucleotídico se traducen en cambios en la proteína. Con color negro se indican las mutaciones silenciosas (en el marco de lectura de GPC) y las de las regiones no traducidas.



Biblioteca Central
 Fac. Cs. Exactas
 U.N.L.P.
 120
 986

Tabla 13.a. Cambios en las secuencias de las proteínas GPC de las cepas MC2 y CD1 del virus Junín

Posición		Nucleótido		Codón		Aminoácido	
MC2	CD1	MC2	CD1	MC2	CD1	MC2	CD1
— 37	1 38	- A	C G	no codificante <i>upstream</i>		región no traducida	
214 218 225 226 228 229	*216 *221 *229 231 *234 *236	ins. ins. ins. ins. ins. ins.	U C U U U C	cambio de marco de lectura		CSILD	LFQFFVF ^a
234 a 235	242 a 247	inserción de 6 nt. en bloque CUCGCA		— —	CUC GCA	--	LA ^b
247 248 250 252	260 262 265 268	C ins. ins. ins.	A A A A	cambio de marco de lectura (1 codón extra)		PR	TEE ^c
280 300 319	*295 *314 *332	del. del. del.	G U U	cambio de marco de lectura (1 codón menos)		VPDCVLLQWWVS	FQTVSFSMVGL ^d
502 1123 1239 1264 1323 1332 1354 1383 1404 1411 1416 1419 1428 1434 1444 1449 1456 1458 1461 1521	515 1136 1252 1277 1336 1345 1367 1396 1417 1424 1429 1432 1441 1447 1457 1462 1469 1471 1474 1534	A U U C C C U C C U A C C G C C C A U G A A G G C C C C C C U U C C	U C C U U U A A U C A G A U C A G C U A G C U C U	AGG UUG AUU CUA GGC CCC UUU UCC GGC ACC CAU AGG CAA CGG CUG CCC CUA CUA AAU CGC	UGG CUG AUC UUA GGU CCU AUU UCA GGU UCC CAC AGA CAG CGA UUG CCA UUG UUG AAC CGU	R L L L L G P F S G T H R R E L P L L N R	W ^e L L L G P I S G S ^g H R R E L P L L N R
1541 1544 1565	1554 1557 1578	C A U	U G C	no codificante <i>downstream</i>		región no traducida	

* En las inserciones o deleciones en que un conjunto del mismo nucleótido está involucrado, no es posible especificar la posición exacta del cambio mutacional. Por lo tanto, se indica el último nucleótido del grupo para identificar la posición en la secuencia.
^{a, b, c, d, e, f, g} letras que identifican regiones de cambios (para facilitar la referencia cruzada con la Figura 13.1.A).
 Los guiones representan gaps.

Cuando se compararon las secuencias aminoacídicas de la región carboxilo del péptido señal y del extremo amino de la glicoproteína G1, de la cepa vacunal CD1 y la cepa salvaje MC2, se encontró un conjunto de alteraciones importantes, incluyendo deleciones e inserciones (para una discusión sobre posibles orígenes alternativos de estos cambios ver **Capítulo 12**).

Estos cambios en la secuencia de la cepa vacunal fueron correlacionados, tentativamente, con su fenotipo atenuado (Ghiringhelli *et al.*, 1997). Sin embargo, las evidencias obtenidas posteriormente por Albariño (1997), mostraron que la cepa atenuada CD1 presentaba el mismo patrón de secuencia en estas regiones que sus antecesores filogenéticos más virulentos, las cepas (XJ-Parodi y XJ#44).

Consecuentemente, se llegó a la conclusión que el marcador de atenuación propuesto inicialmente debe ser descartado.

Por otra parte, en la secuencia de la glicoproteína G2 se detectó la sustitución de dos aminoácidos (MC2 vs CD1, $F_{176} > I_{176}$ y $T_{193} > S_{193}$). Los datos obtenidos por Albariño (1997) sobre las cepas XJ-Parodi y XJ#44 demuestran que los aminoácidos I_{176} y S_{193} son exclusivos de CD1.

En función de estos datos, Albariño (1997) realizó un análisis de la estructura secundaria predicha con el algoritmo de Garnier-Osguthorpe-Robson (GOR, GCG Versión 7.0) para la glicoproteína G2 de las diferentes cepas del virus Junín.

La estructura secundaria predicha es idéntica para todas las cepas en casi toda la glicoproteína, salvo en la región comprendida por los aminoácidos 170 a 191. En esta zona, el algoritmo GOR predice un notable cambio en la estructura secundaria debido a la presencia de la I_{176} en la cepa vacunal. En particular, en CD1, se produce la pérdida de 2 β turns presentes en las otras cepas (**Figura 13.2.A**). Albariño (1997) ha postulado que este cambio estructural podría ser considerado como un marcador

genético de la atenuación de la cepa CD1.

De acuerdo a lo postulado por Burns & Buchmeier (1991), las glicoproteínas G2 se asocian como tetrámeros y se insertan en la envoltura lipídica. Por fuera se encuentran tetrámeros de la glicoproteína G1, asociados con G2 mediante interacciones no covalentes. Estas estructuras constituyen las espículas observadas en la envoltura viral.

En base a estas consideraciones, resulta sumamente atractiva la observación de diferencias en la predicción de estructura secundaria de la G2 de CD1 *versus* las G2 de las cepas más virulentas.

Sin embargo, teniendo en cuenta que, cuando se comparan las 4 cepas del virus Junín, la identidad aminoacídica local (residuos 170-191) es 95,4% y la homología 100%, es lógico preguntarse si la predicción de estructura secundaria derivada del algoritmo GOR es correcta. Los datos obtenidos con otros algoritmos de predicción son diferentes (**Figura 13.2.B**) y son más coherentes con un análisis del perfil de hidropatía local (**Figura 13.2.C**).

En este punto es útil hacer algunas consideraciones acerca de los métodos de predicción de estructuras secundarias. Aún cuando la predicción de la estructura secundaria de un polipéptido es sólo una aproximación a la verdad, el cálculo de la misma es válido cuando no existen determinaciones experimentales.

Las tendencias actuales se inclinan a confiar cada vez más en los resultados obtenidos con aquellos métodos que basan su predicción en un alineamiento múltiple con secuencias proteicas existentes en los bancos de datos, la predicción en paralelo con varios algoritmos diferentes y la generación de una estructura consenso (JPRED: Cuff *et al.*, 1998; Cuff *et al.*, 1999). Estos métodos tienen una certeza promedio superior al 75% (Cuff *et al.*, 1998) y, en algunos casos, 85-95% o mayor (predicción de α hélices transmembrana: Rost *et al.*, 1996d; von Heijne *et al.*, 1994,1996).

Mientras que las predicciones que utilizan exclusivamente la secuencia de la proteína en cuestión y tablas de tendencias estructurales para cada aminoácido (en función de sus vecinos), tales como el GOR IV (Garnier *et al.*, 1996), apenas llegan a una certeza promedio de 60-65%. (Ver Apéndice 8).

En base a las consideraciones anteriores y a la información mostrada en la Figura 13.2, el marcador genético postulado debe ser descartado, al menos en lo que respecta a la existencia de cambios drásticos en la estructura secundaria local y la posible relación de los mismos con la atenuación de la virulencia.

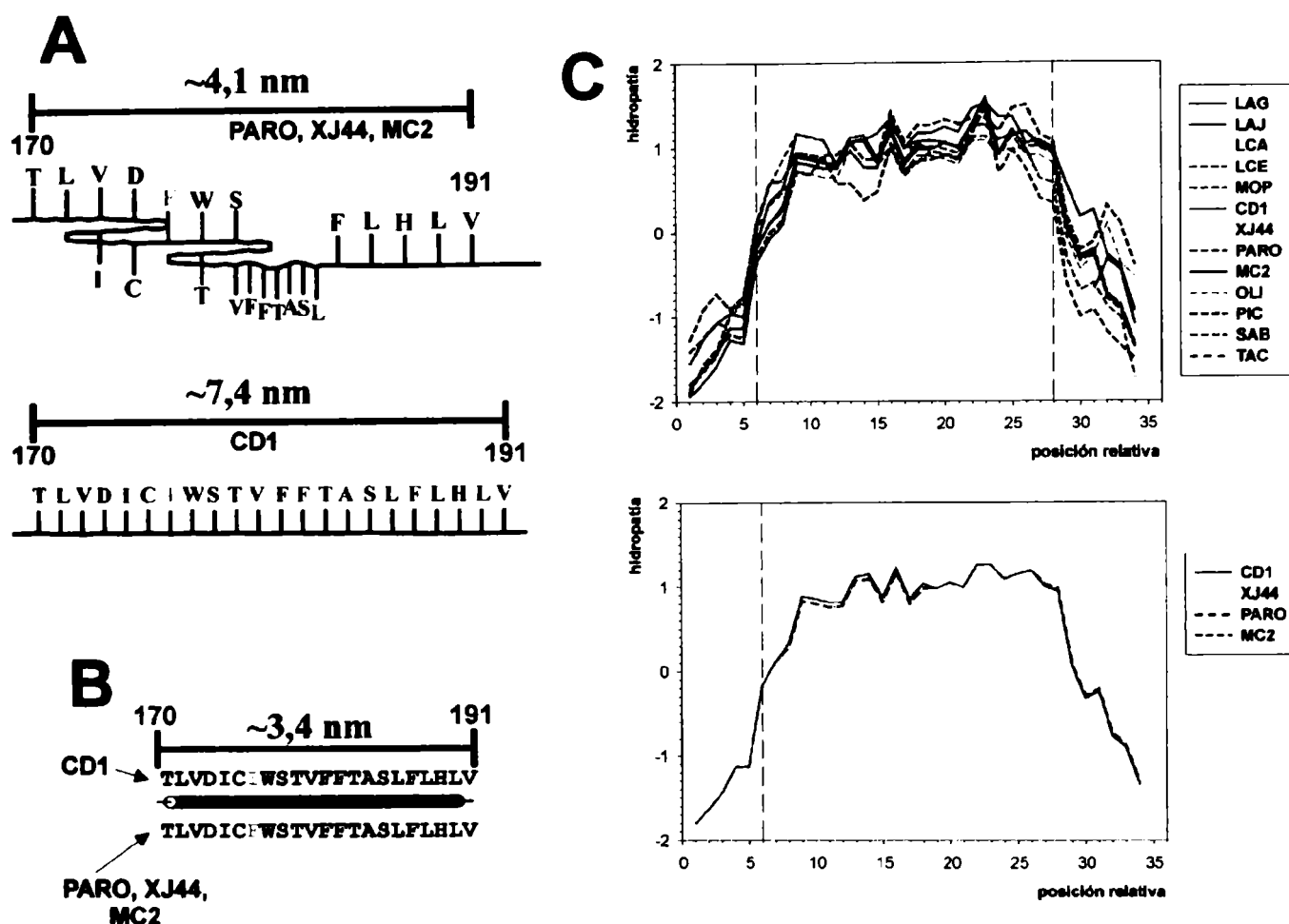


Figura 13.2. Análisis estructural de la potencial región transmembrana de la glicoproteína G2 de cepas del virus Junín con diferente grado de virulencia. A. Predicción de estructura secundaria realizada con el algoritmo GOR (GCG, versión 7) correspondiente a la región transmembrana de G2 de las cepas MC2, XJ#44 (XJ44) y XJ-Parodi (PARO), y la cepa CD1 del virus Junín. Por razón de espacio, las escalas no son idénticas. Se observan los 2 β turns predichos entre los aminoácidos 172 a 180 para las cepas más virulentas, y su ausencia en CD1. En naranja se indican los aminoácidos diferenciales entre las cepas. Además, se indican las longitudes aproximadas de los segmentos proteicos respectivos, calculadas en función de las estructuras locales predichas. Debemos tener en cuenta que el centro apolar de una membrana plasmática tiene ~ 3 nm de espesor (Engelman *et al.*, 1986). B. Consenso de predicción de estructura secundaria obtenido con el servidor JPRED (<http://circinus.ac.uk/jpred>). Las predicciones para cada cepa se realizaron en forma independiente. Se observa que para todas las cepas la predicción es la misma y corresponde a una α hélice de 22 residuos de longitud. En naranja se indica el aminoácido diferencial entre CD1 y el resto de las cepas. Además, se indica la longitud aproximada del segmento proteico, calculada en función de la estructura predicha. C. Perfiles de hidropatía correspondientes al segmento proteico transmembrana. En el panel superior se han superpuesto los perfiles correspondientes a 13 arnavirus y en el panel inferior se han superpuesto los perfiles correspondientes a las cuatro cepas del virus Junín. Dado que las posiciones aminoacídicas en cada proteína son diferentes, el cálculo del perfil se realizó en forma independiente para cada segmento (desde -11 a +11 de las posiciones originales) y los perfiles fueron alineados para el gráfico utilizando una escala posicional relativa en el eje X que abarca desde el aminoácido -6 a +6 del segmento proteico transmembrana predicho para cada arnavirus. Las líneas de guiones verticales delimitan específicamente la región transmembrana. Como se puede observar, en el panel superior, los perfiles de hidropatía muestran poca dispersión (la identidad aminoacídica global en esta zona es 22,7% y la homología 59,1%), mientras que en el panel inferior se observa que los perfiles correspondientes a las distintas cepas del virus Junín son prácticamente idénticos (la identidad aminoacídica en esta zona es 95,4% y la homología 100%). LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia; CD1: Junín-Candid #1; XJ44: Junín-XJ#44; PARO: Junín-XJ-Parodi; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe.

En 1990, Matloubian *et al.*, presentaron las primeras evidencias sólidas que establecen con claridad una asociación entre cambios en una región genómica particular con la patogenicidad de los arenavirus. En particular, estos trabajos fueron realizados mediante la generación de reasociantes genómicos entre la cepa *wild type* Armstrong 53b del virus LCM y variantes esplénicas de LCM: los clones 13 y 28b. Estos resultados fueron confirmados en varios trabajos posteriores de los mismos autores y de otro grupo de investigación (Ahmed *et al.*, 1991; Salvato *et al.*, 1991; Matloubian *et al.*, 1993). En síntesis, en estos trabajos se ha establecido la importancia de cambios aminoacídicos simples, en una o más proteínas virales, asociados con:

1. un cambio en el tropismo predominante: prevalencia en el sistema nervioso central a prevalencia en tejidos linfoides del bazo (en especial en macrófagos)
2. un cambio en el fenotipo CTL: de fenotipo CTL⁺ a CTL⁻ relacionado con la persistencia viral y el establecimiento de una infección crónica.
3. un incremento significativo en la capacidad de replicación en macrófagos.

Según Matloubian *et al.* (1993), los cambios fenotípicos 1 y 3 están relacionados con una mutación en la RNA polimerasa viral (proteína L): (K₁₀₇₉ > Q₁₀₇₉) en el clon 13 ó (K₁₀₇₉ > N₁₀₇₉) en el clon 28b. Mientras que los cambios fenotípicos 1 y 2 están asociados con la mutación ya mencionada en la proteína L y una mutación adicional de la glicoproteína G1 (F₂₆₀ > L₂₆₀). En ninguna de las variantes encontraron mutaciones en las regiones no codificantes.

Según el trabajo de Salvato *et al.* (1991), la mutación de la glicoproteína G1 (F₂₆₀ > L₂₆₀) es necesaria pero no suficiente para producir el cambio de fenotipo CTL (cambio fenotípico 2).

Por otra parte, según Zhang *et al.* (1999) y Salvato (2000, comunicación

personal), en estudios realizados con el arenavirus Pichindé mediante la generación de reasociantes genómicos han permitido establecer una relación entre los cambios en la virulencia y el RNA S. Todavía no se conoce con exactitud cuáles son esos cambios, ni su ubicación en el RNA S (regiones codificantes y/o no codificantes).

En última instancia, en las cepas del virus Junín clonadas y secuenciadas hasta ahora, la I₁₇₆ y la S₁₉₃ son peculiaridades de la glicoproteína G2 exclusivas de la cepa vacunal Candid #1, y constituyen los únicos cambios aminoacídicos no comunes con ninguna de las otras cepas. La S₁₉₃ está ubicada en la porción citoplásmica de G2, mientras que la I₁₇₆ está ubicada en la α hélice transmembrana. Dado que el residuo isoleucina es más hidrofóbico que el residuo fenilalanina, los grupos R de ambos aminoácidos poseen diferente geometría (**Figura 13.3.A**) y, además, la I₁₇₆ en CD1 (o F₁₇₆ en MC2, XJ#44 y XJ-Parodi) está ubicada en el tercio más externo de la región apolar de la membrana plasmática (**Figura 13.3.B**), sería razonable suponer que el cambio de aminoácido está asociado con la atenuación de la virulencia.

La influencia sobre la atenuación de la virulencia puede explicarse, al menos, de 3 maneras: un cambio en la geometría del segmento transmembrana (segmento de anclaje) que se traslada a la porción externa de G2, alterando la geometría de la espícula de glicoproteínas y, por lo tanto, dificultando el ingreso del virus en la célula y/o cambiando el tropismo viral; o un cambio en la estabilidad de la interacción entre las hemihélices más hidrofóbicas y los lípidos de la membrana plasmática (**Figura 13.4.**), que dificulta la actividad fusogénica de G2 y la liberación de las nucleocápsides en el citoplasma; o la combinación de ambas explicaciones.

De todas maneras, el cambio existe y es llamativo que sólo se presente en Junín-Candid #1.

Estas consideraciones no excluyen la presencia de otros cambios en regiones genómicas no analizadas en esta parte del

estudio (proteína N) o desconocidas hasta el momento (RNA L y/o sus productos génicos).

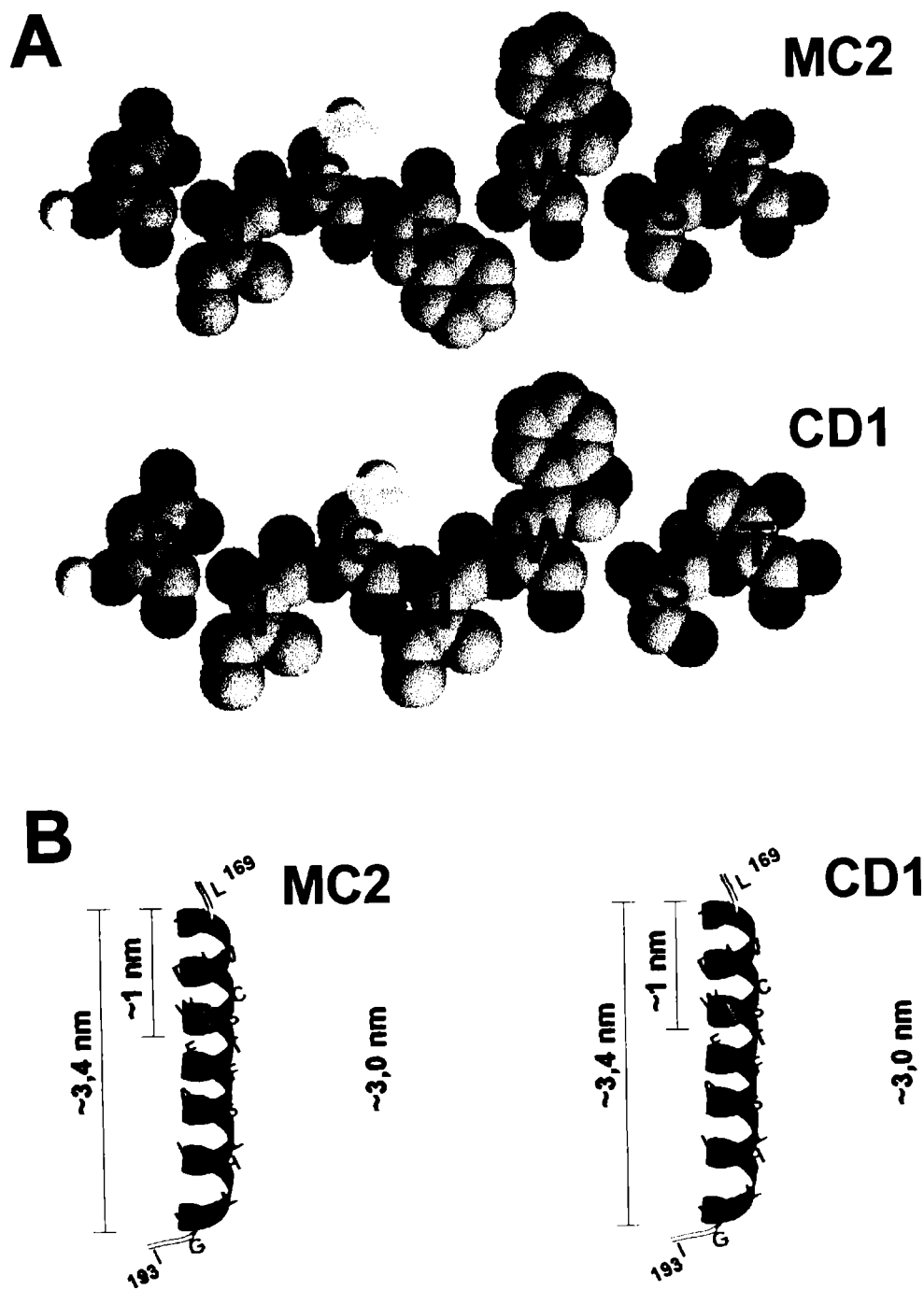


Figura 13.3. Representaciones esquemáticas de las α hélices transmembrana de las cepas MC2 y CD1 del virus Junín. A. Representación lineal, con modelos de esferas, de los aminoácidos 173 a 179. En azul se indican los nitrógenos, en rojo los oxígenos, en naranja el azufre, en gris los carbonos y en blanco sombreado los hidrógenos (sólo los del N-terminal y del grupo SH). Con símbolos de una letra, superpuestos sobre la cadena carbonada, se indican los aminoácidos respectivos. Como se puede observar la geometría y el volumen de los grupos R de la isoleucina y de la fenilalanina son muy diferentes. Los espacios entre aminoácidos son sólo artefactos del dibujo. B. Representación esquemática de la α hélice transmembrana inserta en la región netamente apolar de la membrana plasmática (recuadro de color amarillento). Con símbolos de una letra se indican los aminoácidos respectivos, en rojo los ubicados en el plano anterior y en verde los ubicados en el plano posterior (para facilitar la visualización algunos aminoácidos están ligeramente desplazados de la posición original). El aminoácido diferencial entre las cepas se muestra en celeste. Se indican en nanómetros la longitud aproximada de la α hélice, el espesor de la región apolar de la membrana y la posición del residuo diferencial (medida en distancia desde el extremo extracelular de la α hélice).

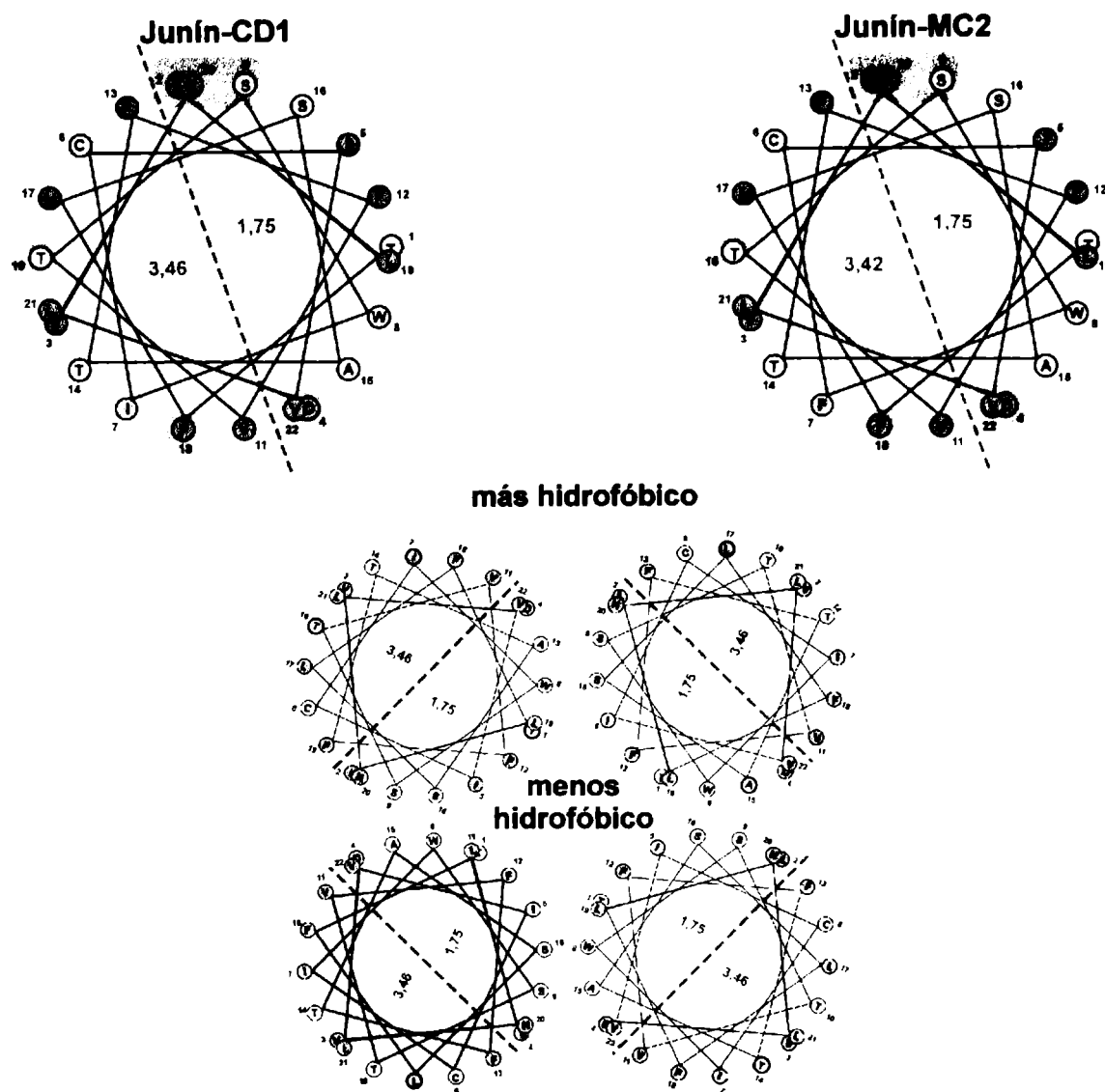


Figura 13.4. Representaciones tipo *helical wheel* de las α hélices transmembrana de las cepas MC2 y CD1 del virus Junín. En el panel superior se muestran los diagramas individuales de cada cepa. Los aminoácidos fueron numerados secuencialmente desde el extremo N-terminal de la estructura. Basado en Karlin *et al.* (1990), se hace referencia al diferente grado de hidrofobicidad del residuo con distinto tipo de fondo (fondo celeste: residuos hidrofílicos; fondo grisado: residuos hidrofóbicos; fondo blanco: los residuos ambivalentes). Con borde rojo se indica el residuo diferencial entre las cepas. La línea de guiones roja corresponde a un plano hipotético que divide la α hélice en dos hemihélices. Los valores indicados en rojo corresponden a la hidrofobicidad media de la respectiva hemihélice. Como se puede observar, 20 de los 22 residuos que forman las hélices son hidrofóbicos o ambivalentes y, por lo tanto, las hélices son netamente hidrofóbicas. Sin embargo, la distribución de los residuos hidrofóbicos no es exactamente uniforme, y el plano de división en hemihélices permite identificar dos regiones de diferente hidrofobicidad media. Además, las dos hemihélices presentan una marcada asimetría en la conservación de residuos en todos los arnavirus: la hemihélice menos hidrofóbica presenta 6 residuos altamente conservados en tipo y posición, mientras que la hemihélice más hidrofóbica presenta sólo 2 residuos conservados. El grado de conservación se indica con fondos sombreados de distinto color (fondo rosa: 100%, fondo amarillo pálido: ~90%, fondo verde claro: ~80%). Por otra parte, algunos de los residuos tienen escasa probabilidad de estar presentes en la hemihélice que interactúa exclusivamente con los lípidos de membrana, como el aspartato (D). En función de estas consideraciones, en el panel inferior se representa un modelo hipotético de la disposición de las cuatro α hélices que soportan el anclaje en la membrana del tetramero de G2. Las hemihélices de menor hidrofobicidad media podrían estabilizar el conjunto mediante diferentes tipos de interacciones (hidrofóbicas, puentes de hidrógeno, etc.) entre ellas, mientras que las hemihélices de mayor hidrofobicidad media podrían estabilizar el conjunto, principalmente, mediante interacciones hidrofóbicas con los lípidos de la membrana. Con fondo de color amarillento se representa la región netamente apolar de la membrana. Con la exclusiva finalidad de diferenciar las regiones, la posible zona de influencia de las hemihélices menos hidrofóbicas se ha representado con fondo blanco.

Coincidentemente con lo publicado para otros sistemas virales, los resultados de este análisis sugieren la participación de las glicoproteínas de superficie en la atenuación de la virulencia.

En este sentido, se han obtenido evidencias experimentales sobre la relación entre la alteración de las proteínas de superficie y la atenuación de la virulencia para las cepas vacunales del virus de la polio, Sabin tipo 1 y 3 (Tatem *et al.*, 1992), y del virus de la fiebre amarilla (17D; Hahn *et al.*, 1987). Del mismo modo, se encuentran numerosas publicaciones en la literatura, sobre los efectos profundos en la virulencia y el tropismo tisular provocados por sustituciones de aminoácidos en las proteínas de la envoltura de reovirus, rhabdovirus, togavirus, bunyavirus, picornavirus, birnavirus y flavivirus (Spriggs *et al.*, 1983; Vrati *et al.*, 1986; Prehaud *et al.*, 1988; Takehara *et al.*, 1989; Polo *et al.*, 1991; Jensen *et al.*, 1993; Vialat *et al.*, 1997; Brandt *et al.*, 2001; Monath *et al.*, 2002).

Por otra parte, cualquier cambio que altere el clivaje proteolítico del precursor de las glicoproteínas de superficie, también está relacionado con cambios en el tropismo y la virulencia en paramyxovirus y togavirus (Tashiro *et al.*, 1990; Davis *et al.*, 1995). Respecto de esto, en ortomyxovirus y togavirus se ha detectado que alteraciones en el patrón de glicosilación de las proteínas de superficie está relacionado con el fenómeno de atenuación (Deshpande *et al.*, 1987; Zheng *et al.*, 1989).

En este sentido, en el arnavirus LCM, Wright *et al.* (1990) han publicado que la inhibición de la glicosilación de GPC con tunicamicina dificulta severamente el procesamiento proteolítico y, por lo tanto, la generación de progenie.

Respecto de las cepas del virus Junín, no se han observado cambios de aminoácidos cercanos al sitio de clivaje del precursor de las glicoproteínas. Además, las cepas MC2, XJ-

Parodi, XJ#44 y Candid #1 del virus Junín presentan conservados en secuencia y posición todos los sitios de N-glicosilación, tanto en G1 como en G2.

Por lo tanto, se puede concluir que el clivaje del precursor GPC y el patrón de glicosilación no constituyen una base para los cambios en la virulencia de las cepas estudiadas.

Análisis comparativo de las proteínas GPC de los arnavirus

Estructura primaria

La comparación entre las proteínas GPC de los arnavirus, permite observar la existencia de diferencias significativas en cuanto al grado de conservación de la secuencia aminoacídica en las tres regiones peptídicas (**Figura 13.5**).

Un análisis más detallado, basado en alineamientos múltiples (ver **Apéndice 6**), de la similitud aminoacídica en cada región, muestra que:

- En el **peptido señal** se observa que la identidad varía entre 33,3 % (MC2-LCA) y 56,9 % (PIC-LAG o PIC-LAJ) y la homología entre 65,0 % (MC2-LCA o MC2-LCE) y 84,5 % (PIC-LCA o PIC-LCE), cuando se tienen en cuenta los arnavirus del Nuevo Mundo vs. los del Viejo Mundo. Por otra parte, cuando se consideran por separado los dos grandes grupos de arnavirus, en los del Nuevo Mundo la identidad varía entre 46,6 % (MC2-PIC) y 82,8 % (CD1-TAC) y la homología entre 72,4 % (MC2-PIC) y 93,1 % (OLI-SAB); mientras que en los del Viejo Mundo la identidad varía entre 55,2 % (LCA-MOP) y 75,9 % (LAG-MOP o LAJ-MOP) y la homología entre 86,2 % (LAG-LCA, LAJ-LCA, LCA-MOP o LCE-MOP) y 96,6 % (LAG-MOP o LAJ-MOP) (**Tablas 13.b.1 y 13.b.2**).

- En la **glicoproteína G1** se observa que la identidad varía entre 14,1 % (MC2-LCA) y 30,0 % (PIC-LAJ) y la homología entre 38,0 % (MC2-LCA) y 56,6 % (PIC-MOP), cuando se tienen en cuenta los arnavirus del Nuevo

Mundo vs. los del Viejo Mundo. Por otra parte, cuando se consideran por separado los dos grandes grupos de arenavirus, en los del Nuevo Mundo la identidad varía entre 14,0 % (MC2-PIC) y 44,5 % (CD1-TAC) y la homología entre 43,0 % (PIC-TAC) y 74,9 % (CD1-TAC); mientras que en los del Viejo Mundo la identidad varía entre 43,7 % (LCE-MOP) y 73,0 % (LAG-MOP) y la homología entre 66,2 % (LCA-MOP) y 89,0 % (LAG-MOP o LAJ-MOP) (Tablas 13.c.1 y 13.c.2).

- En la **glicoproteína G2** se observa que la identidad varía entre 48,3 % (SAB-LCE) y 57,8 % (OLI-LCA) y la homología entre 75,2 % (PIC-LCE) y 78,9 % (OLI-LCA), cuando se tienen en cuenta los arenavirus del Nuevo Mundo vs. los del Viejo Mundo. Por otra parte, cuando se consideran por separado los dos grandes grupos de arenavirus, en los del Nuevo Mundo la identidad varía entre 51,7 % (PIC-TAC) y 82,2 % (MC2-TAC) y la homología entre 78,4 % (PIC-TAC) y 94,9 % (CD1-TAC o MC2-TAC); mientras que en los del Viejo Mundo la identidad varía entre 66,2 % (LCE-MOP) y 80,3 % (LAG-MOP) y la homología entre 87,6 % (LAG-LCE o LAJ-LCE) y 92,7 % (LAG-MOP o LAJ-MOP) (Tablas 13.d.1 y 13.d.2).

En síntesis, el orden de conservación de secuencias es G2, péptido señal y G1; lo cual, probablemente, sea un reflejo de las respectivas funciones en los arenavirus.

Estudiando las secuencias desde el punto de vista de las frecuencias aminoacídicas en cada región peptídica (Figura 13.6.A, 13.6.B y 13.6.C), el análisis muestra que el péptido señal es la región con mayor diversidad y dispersión, mientras que en G1 y G2 las frecuencias son mucho más conservadas. Por otra parte, cuando se agrupan los aminoácidos en hidrofóbicos, hidrofílicos y ambivalentes (Karlín *et al.*, 1990), las dispersiones disminuyen significativamente en todas las regiones. En particular, en G1 y G2, se observa un detalle muy interesante:

independientemente de las diferencias en conservación de secuencias (identidad y homología) para cada una, el patrón de frecuencias es prácticamente el mismo.

Estructura secundaria

La predicción de estructuras secundarias y el análisis de las mismas, muestran que existe un grado de conservación estructural muy alto. Esto es especialmente cierto tanto para el péptido señal como para G2, mientras que en G1 existe mayor variabilidad.

En 10 de los 11 arenavirus analizados, el péptido señal está formado por los primeros 58 aminoácidos de GPC, con una α hélice de 53 residuos de longitud. La única excepción es la cepa MC2 del virus Junín, donde el péptido señal es 4 residuos más corto (54 aminoácidos), y la estructura predicha está formada por una α hélice de 38 residuos, separada por 3 aminoácidos de una estructura β de 6 residuos de longitud (Figura 13.7).

En las glicoproteínas 1 (G1) de todos los arenavirus, como se puede observar en las Figuras 13.8 y 13.9.A, existe mayor variación. Sin embargo, cuando se divide la comparación de predicciones en regiones definidas en función de las estructuras secundarias predominantes, es posible distinguir un cierto patrón estructural formado por 4 regiones predominantemente β y 2 regiones predominantemente α . En las regiones β , el número de estructuras individuales es variable y, en algunos casos, la predicción es diferente. Así, en las cepas del virus Junín (CD1 y MC2), la región β 1 no existe y, en su lugar, la predicción indica la presencia de una α hélice (identificada como α 0-1); mientras que en los virus Oliveros, Pichindé y Tacaribe la estructura β 4-2 no existe y, en su lugar, la predicción indica la presencia de una α hélice (identificada como α 3-1) (Figura 13.9.A).

En las glicoproteínas 2 (G2) de todos los arenavirus, como se puede observar en las Figuras 13.10 y 13.11.A, existe una gran

conservación estructural. En base a esto es posible definir 4 regiones de estructura tipo β (β_1, \dots, β_4), 4 regiones de α hélice ($\alpha_1, \dots, \alpha_4$) y 8 regiones definidas como *loops*.

Discusión sobre las GPCs de arenavirus

Péptido señal

Los péptidos señal de cada arenavirus han sido analizados en forma individual, definiendo el límite C-terminal en base a la bibliografía con datos experimentales (virus LCM-Armstrong, Pichindé y Tacaribe, Buchmeier *et al.*, 1993; virus Junín, Padula *et al.*, 1996), o por homología y análisis estructural en los restantes arenavirus.

La información obtenida es concordante entre sí y con las características generales, asociadas con la función, de los péptidos señal de proteínas heterólogas (Lüthy & Eisenberg, 1991). La hidrofobicidad media de las α hélices predichas varía entre 1,9 (Junín-MC2) y 2,6 (Tacaribe), coincidentemente con la alta proporción de aminoácidos hidrofóbicos (~50 %) y ambivalentes (~30 %).

En función de lo anterior, no se deberían esperar valores de anfipaticidad (AI) superiores a 2, cuando se realiza el análisis de acuerdo a Cornette *et al.* (1987, 1995). Sin embargo, los valores obtenidos son superiores a 2, para casi todas las α hélices del péptido señal, con la excepción de Sabiá y Tacaribe.

Un análisis detallado de los períodos dominantes en las α hélices predichas en el péptido señal de los distintos arenavirus (ver **Apéndice 7**), permite detectar la existencia de un período dominante de ~3,5. Esto explica el valor de AI superior a 2, ya que el área comprendido entre 85° y 110° se hace significativa *versus* el área total (ver **Apéndice 5**). Por otra parte, se observa la presencia de otros períodos de importancia. Probablemente, esto último está en relación con la longitud de la estructura (40 o 50 residuos) y su carácter predominante apolar, lo cual permite detectar

repeticiones periódicas de diferente magnitud. En conclusión, para estructuras superiores a 15 o 20 residuos, el valor de AI que permite definir anfipaticidad debería ser muy superior a 2.

Por último, desde el punto de vista de la función del péptido señal, es interesante destacar que la búsqueda en la base de datos PROSITE predice la existencia de sólo dos tipos de sitios potenciales: de fosforilación y de miristoilación.

Respecto de los sitios potenciales de fosforilación, por caseínquinasa, proteínquinasa C o quinasas dependientes de cAMP (datos no mostrados), hasta el presente no existe información experimental que demuestre que la función biológica de un péptido señal requiera de este tipo de modificación postraducciona, por lo cual deberían ser descartados como señales biológicamente utilizadas. Una razón adicional que sustenta esta consideración es que en 5 de los 11 arenavirus estudiados estos sitios potenciales no existen.

Respecto del sitio potencial de miristoilación (**Figura 13.7**; ver también **Apéndice 6, Figura Ap6.1**), aún cuando no existen pruebas experimentales directas que demuestren que esta modificación proteica ocurre, es interesante destacar que el sitio potencial está totalmente conservado en todos los arenavirus estudiados y, en particular, está ubicado en una posición “ideal” (GQIVTF, residuos 2 a 7) para ser utilizado *in vivo*.

Acerca de esta hipótesis, es necesario realizar algunas consideraciones adicionales.

La adición de grupos miristato a ciertas proteínas es una modificación cotraducciona catalizada por una enzima citosólica, la *Myristoyl CoA:protein N-myristoyltransferase* (NMT), muy frecuente en células eucariotas (Grand, 1989; Magee *et al.*, 1989; Raju *et al.*, 1997; Boutin, 1997). Numerosas publicaciones indican que el rol principal de esta modificación es una correcta localización subcelular y, en particular, el direccionamiento a membranas

endosomales, del retículo endoplásmico y plasmática (Busconi and Michel, 1993; D'Souza-Schorey and Stahl, 1995; Zlatkine *et al.*, 1997; McCabe and Berthiaume, 1999). Además, en muchos sistemas virales se ha demostrado que la acilación de las glicoproteínas juega un rol importante en el direccionamiento subcelular correcto. Tal vez uno de los sistemas más estudiados es el de HIV, donde se ha verificado la necesidad de la miristoilación para un transporte eficiente de la proteína Env a la superficie celular (Shiraishi *et al.*, 2001).

En síntesis, la similitud en la posición del residuo miristato (siempre una glicina N-terminal, normalmente en posición 1 ó 2) y la similitud en la localización subcelular en varias proteínas aciladas con este tipo de residuos, sugiere que las propiedades biofísicas de las mismas y las de las membranas biológicas son los principales determinantes en la selección apropiada del tipo de membrana. (Johnson *et al.*, 1990; McCabe and Berthiaume, 1999).

En función de las consideraciones anteriores, y de las peculiaridades del sitio de miristoilación existentes en el péptido señal de los arnavirus (conservación en todos y la G ubicada en +2), es altamente probable que esta modificación cotraduccional ocurra durante la síntesis de GPC en la célula infectada y asegure el direccionamiento del precursor de las glicoproteínas virales al retículo endoplásmico.

Por otra parte, en 1999, Cordo *et al.* publican información sobre la inhibición de la replicación del virus Junín utilizando análogos del ácido mirístico, lo cual demuestra que alguna de las proteínas del virus Junín se miristoila durante el ciclo viral.

Glicoproteína G1

Los extremos N-terminales de las G1 han sido definidos experimentalmente en varios arnavirus (LCM-Armstrong, Pichindé y Tacaribe, Buchmeier *et al.*, 1993; Junín, Padula *et al.*, 1996) y por homología en los restantes. Mientras que el C-terminal ha sido definido en

todos considerando el sitio de reconocimiento para la proteasa de origen celular, caracterizado por Buchmeier *et al.* (1987) en LCM y por Lenz *et al.* (2000) en Lassa.

La longitud en aminoácidos de G1 varía desde 191 residuos en Tacaribe hasta 225 residuos en Oliveros, lo cual es coincidente con la escasa conservación de secuencias ya mencionada (ver **Tabla 13.c**).

Por otra parte, la búsqueda en la base de datos PROSITE predice la existencia de cantidades variables de sitios potenciales de glicosilación, fosforilación (datos no mostrados) y miristoilación (ver **Apéndice 6, Figura Ap6.2**). Hasta el momento, no existen datos experimentales que demuestren la fosforilación o miristoilación en G1, por lo cual se desconoce si estos sitios son biológicamente significativos.

Dado que la bibliografía actual indica que las acilaciones por adición de miristoilos ocurren en la porción N-terminal, la utilización *in vivo* de estos sitios debería descartarse. Sin embargo, se ha publicado una asociación entre la miristoilación de proteínas en sitios internos y la fosforilación por proteínquinasa C (Utsumi *et al.*, 1994). En última instancia, la confirmación (o no) del empleo *in vivo* de algunos (o todos) los sitios de fosforilación y/o miristoilación, debería realizarse experimentalmente.

Respecto de los sitios de N-glicosilación, la cantidad varía desde 4 sitios (cepas de Junín y Tacaribe) hasta 11 sitios (Pichindé) (ver **Apéndice 6, Figura Ap.6.2**). En la mayor parte de los arnavirus no se sabe cuántos ni cuáles de los sitios potenciales están glicosilados *in vivo*. El único informe que muestra en forma experimental cuántos de los sitios potenciales son utilizados *in vivo* corresponde al virus LCM, cepa Armstrong (Wright *et al.*, 1990), donde se muestra que de los 6 sitios existentes sólo se utilizan 5. Es posible que la variabilidad en la cantidad de sitios de N-glicosilación y los sitios realmente utilizados *in vivo* (en número y posición) en los distintos arnavirus tengan una

íntima relación con aspectos diferenciales de la patogenia viral. Un estudio mutacional detallado, con ensayos de actividad biológica en modelos animales, permitiría resolver este punto.

Desde un punto de vista estructural, las glicoproteínas G1 de los diferentes arnavirus tienen, en general, una distribución aproximadamente regular de aminoácidos hidrofóbicos y no hidrofóbicos, lo cual redundaría en una mayor probabilidad promedio de exposición al solvente y sugiere que, en todos los casos, podría ser una proteína soluble (análisis general no mostrado). Esta información es coincidente con la predicción computacional realizada para las G1 de las cepas del virus Junín (ver **Capítulos 7 y 11**), que sugiere que las G1 son proteínas periféricas y con una conformación de tipo globular. Además, también es coincidente con el modelo de Buchmeier *et al.* (1987, 1993), quien postula que la porción externa de las espículas virales está formada por tetrámeros de proteínas globulares (G1).

La asociación tetramérica está estabilizada covalentemente mediante uniones disulfuro. En este punto es interesante destacar que la cantidad de cisteínas en las G1 varía en el rango de 5 a 8 en los distintos arnavirus, 3 de las mismas están totalmente conservadas en posición en un alineamiento múltiple (ver **Apéndice 6, Figura Ap.6.2**). Es probable que la conservación de estas cisteínas esté relacionada con el mantenimiento de un patrón de interacciones, común a todos los arnavirus, que asegure la correcta formación de la porción más externa de las espículas virales.

Por otra parte, desde el punto de vista de la estructura secundaria, un estudio de los perfiles de hidropatía de las glicoproteínas G1 (**Figura 13.9.B**), permite observar que no existe un perfil común a todos los arnavirus. La similitud de perfiles de hidropatía varía entre 48,2 % (OLI-MOP) y 75,2 % (TAC-MOP), cuando se tienen en cuenta los arnavirus del Nuevo Mundo vs. los del Viejo Mundo. Por otra

parte, cuando se consideran por separado los dos grandes grupos de arnavirus, en los del Nuevo Mundo la similitud varía entre 56,0 % (MC2-OLI) y 85,2 % (CD1-TAC); mientras que en los del Viejo Mundo la similitud varía entre 78,5 % (LAG-LCE y LAJ-LCE) y 86,4 % (LAJ-MOP) (**Tabla 13.e.1, Figuras 13.9.C y 13.9.D**).

Estas características son coincidentes con la ausencia de un patrón común observada en las predicciones de estructura secundaria, tanto respecto del tipo de estructura predicha como respecto de algunas de las características fisicoquímicas de las mismas (AI: índice de anfipaticidad, hidrofobicidad media local, etc.) (ver **Figura 13.9.A**).

En síntesis, en cualquier tipo de comparación que se realice, las diferencias entre las glicoproteínas G1 de todos los arnavirus son mayores que las similitudes.

Como se ha mencionado anteriormente, no ha sido posible encontrar un patrón común más restringido que la simple definición de 4 regiones con abundancia de estructuras tipo β , 2 regiones α hélice y múltiples *loops*, que actuarían como conectores (es probable que en estas regiones existan verdaderos *loops* y β -*turns*). Sin embargo, es interesante destacar que la alta predominancia de estructuras tipo β y *loops* es coincidente con lo determinado mediante difracción de rayos X en proteínas de función análoga de otros sistemas virales, tales como la HA₁ de influenza (orthomyxovirus) (Lamb and Krug, 1995).

Todas las características mencionadas anteriormente y las consideraciones conexas, probablemente, están en directa relación con una de las principales funciones biológicas de la glicoproteína G1 en los arnavirus.

Aunque resulte bastante obvio, la infección de una célula por un virus comienza con la unión (adsorción) del virus a la superficie celular mediante algún polipéptido viral llamado VAP (*viral attachment protein*) (Lentz, 1990). Como ha sido demostrado por Borrow &

Oldstone (1992) para el virus LCM (cepa Armstrong 53b), la glicoproteína G1 es la VAP de los arnavirus. En este trabajo, los autores muestran la unión específica de la glicoproteína G1 a una proteína de membrana de ~120-140 kDa presente en fibroblastos y células similares (murinas y humanas) y, aparentemente, ausente (o muy escasa) en linfocitos T. Posteriormente, la proteína de membrana fue identificada como el α -dístroglicano (Cao *et al.*, 1998; Smelt *et al.*, 2001; Kunz *et al.*, 2001), una proteína periférica con un alto grado de glicosilación, aunque la presencia de los residuos glucosídicos no es imprescindible para la unión del virus LCM. Dada la aparente ausencia del receptor en los linfocitos, todavía no está claro como ingresa el virus LCM en los mismos cuando se establecen infecciones persistentes; los autores suponen que una de las posibilidades es la existencia de una subpoblación de linfocitos (no detectada) que posee el receptor, pero en menor proporción que el resto de las células.

Con mayor o menor detalle, en la bibliografía se encuentran descripciones de otros virus envueltos en los cuales se han identificado los VAPs y, en algunos casos, los receptores celulares. En todos los casos las VAPs son proteínas virales presentes en la envoltura, mientras que los receptores son, generalmente, constituyentes normales de la membrana plasmática (Lentz, 1990). La unión de las partículas virales a los receptores hace que se establezca una íntima interacción VAP/receptor, con el concomitante ingreso del virus a la célula mediante fusión directa de membranas o mediada por endocitosis (Marsh and Helenius, 1989). En el virus LCM se ha publicado que el mecanismo de ingreso es mediado por endocitosis (Buchmeier, 1993; Borrow and Oldstone, 1994).

Desde el punto de vista de la interacción VAP/receptor y, considerando que la naturaleza, número y distribución de los receptores celulares es uno de los principales determinantes del rango

de hospedador y del tropismo viral (Lentz, 1990), es altamente probable que las diferencias existentes entre las glicoproteínas G1 de los arnavirus estén íntimamente relacionadas con las diferencias de patogenia entre los distintos arnavirus.

Centrando las hipótesis exclusivamente, en la interacción VAP/receptor (en esta parte de la discusión), una explicación acerca de las bases responsables de algunas de las diferencias de patogenia debería incorporar, no sólo las diferencias de secuencia (estructura primaria), sino también los diferentes patrones de plegamiento potencialmente existentes (en función de la estructura secundaria y los perfiles de hidropatía). Estos patrones podrían permitir una conformación espacial equivalente, proteínas globulares asociadas en tetrámeros (estructuras terciaria y cuaternaria), que favorecerían la interacción de las VAPs de los diferentes arnavirus con receptores celulares alternativos al descrito en LCM o totalmente distintos.

Glicoproteína G2

Los extremos N-terminales de las G2 han sido definidos experimentalmente en varios arnavirus (LCM-Armstrong, Pichindé y Tacaribe, Buchmeier *et al.*, 1993; Junín, Padula *et al.*, 1996; Lassa, Lenz *et al.*, 2000) y por homología en los restantes. Mientras que el C-terminal es el original del precursor de las glicoproteínas (GPC). Actualmente se conoce que la proteólisis para separar G1 y G2 es realizada por una convertasa de proproteínas celular, la SKI-1/S1P (*subtilisin kexin isozyme-1/site-1 protease*; Lenz *et al.*, 2001).

La longitud en aminoácidos de G2 varía desde 230 residuos en Pichindé hasta 235 residuos en Oliveros, lo cual es coincidente con la alta conservación de secuencias ya mencionada (ver **Tabla 13.d**).

Por otra parte, la búsqueda en la base de datos PROSITE predice la existencia de

cantidades variables de sitios potenciales de glicosilación, fosforilación (datos no mostrados) y miristoilación (ver **Apéndice 6, Figura Ap6.3**). Hasta el momento, no existen datos experimentales que demuestren la fosforilación o miristoilación en G2, por lo cual se desconoce si estos sitios son biológicamente significativos. En este punto caben las mismas consideraciones que ya se hicieron sobre los sitios equivalentes en G1. En última instancia, la confirmación (o no) del empleo *in vivo* de algunos (o todos) estos sitios, debería realizarse experimentalmente.

Respecto de los sitios de N-glicosilación, la cantidad varía desde 3 sitios (cepas de LCM) hasta 5 sitios (Pichindé) (ver **Apéndice 6, Figura Ap.6.3**). Es interesante destacar que 3 de los sitios están conservados en posición en todos y un cuarto sitio está conservado en 9 de los 11 arnavirus estudiados (no existe en las cepas de LCM). En la mayor parte de los arnavirus no se sabe cuantos ni cuales de los sitios potenciales están glicosilados *in vivo*. El único informe que muestra experimentalmente cuántos de los sitios potenciales son utilizados *in vivo* corresponde a la cepa Armstrong del virus LCM (Wright *et al.*, 1990), donde se verifica que de los 3 sitios existentes sólo se utilizan 2. Es probable que la gran constancia en la cantidad de sitios de N-glicosilación y los sitios realmente utilizados *in vivo* (en número y posición) en los distintos arnavirus tengan una íntima relación con restricciones evolutivas a tolerar cambios asociados con la función biológica de la glicoproteína G2. Un estudio mutacional, con ensayos de actividad biológica en modelos animales, permitiría resolver este asunto.

Desde un punto de vista estructural, las glicoproteínas G2 de los diferentes arnavirus tienen, en general, una distribución irregular de aminoácidos hidrofóbicos y no hidrofóbicos. Sin embargo, como se puede observar en la **Figura 13.11.B**, los perfiles de hidropatía son prácticamente idénticos. La similitud de perfiles de hidropatía varía entre 81,2 % (CD1-MOP) y

88,1 % (OLI-LCE), cuando se tienen en cuenta los arnavirus del Nuevo Mundo vs. los del Viejo Mundo. Por otra parte, cuando se consideran por separado los dos grandes grupos de arnavirus, en los del Nuevo Mundo la similitud varía entre 84,8 % (CD1-PIC) y 91,2 % (MC2-SAB); mientras que en los del Viejo Mundo la similitud varía entre 85,1 % (LCE-MOP) y 87,8 % (LAG-MOP) (**Tabla 13.f.1, Figuras 13.11.C y 13.11.D**).

Además, es necesario destacar, en particular, la existencia de una región delimitada por los residuos 168 a 198, rica en aminoácidos hidrofóbicos (**Figura 13.11.B**) y con muy escasa probabilidad de exposición al solvente (análisis general no mostrado), las cuales son características típicas de una región transmembrana. Esta información es coincidente con la predicción computacional realizada para las G2 de las cepas del virus Junín (ver **Capítulos 7 y 11**), que sugiere que las G2 son proteínas asociadas a membrana. Además, también es coincidente con el modelo de Buchmeier *et al.* (1987, 1993), quien postula que la base de las espículas virales está formada por tetrámeros de proteínas ancladas a membrana (G2). Dada la ausencia de cisteínas conservadas en todos los arnavirus en la porción transmembrana (ver **Apéndice 6, Figura Ap6.3**), la asociación tetramérica estaría estabilizada por interacciones no covalentes.

Por otra parte, desde el punto de vista de la estructura secundaria, como se puede observar en las **Figuras 13.10 y 13.11.A**, las predicciones son prácticamente idénticas, tanto en el tipo de estructura predicha como en algunas de las características fisicoquímicas (AI: índice de anfipaticidad, hidrofobicidad media local, etc.; ver **Figura 13.11.A**).

En síntesis, en cualquier tipo de comparación que se realice, las similitudes entre las glicoproteínas G2 de todos los arnavirus son mayores que las diferencias.

El patrón estructural uniforme para

todas las G2, probablemente, está en directa relación con dos de las funciones biológicas descritas para la misma:

1. Constituir el soporte físico de anclaje de las espículas virales en la membrana de la envoltura viral.
2. Ser la responsable de la liberación de las nucleocápsides desde las vesículas endocíticas al citoplasma de la célula infectada.

Respecto del primer punto, las únicas pruebas experimentales corresponden al virus prototipo (LCM). Trabajando sobre células infectadas con LCM, se ha determinado, mediante secuenciamiento, el N-terminal de la fracción de G2 protegida del ataque proteolítico por la membrana.

Los resultados obtenidos muestran la protección de un fragmento de 18 residuos (${}_{174}\text{LLMFSTAYLVSIFLHLV}_{191}$) (Burns and Buchmeier, 1993). Aplicando las mismas metodologías sobre las G2 de envolturas virales, los mismos autores encuentran un fragmento protegido más largo (${}_{165}\text{GSTPLALMDLLMFS TSAYLVSIFLHLV}_{191}$) (Burns and Buchmeier, 1993). Además, la predicción mediante computadora sugiere la presencia de una α hélice transmembrana de 22 residuos de longitud (${}_{170}\text{ALMDLLMFSTAYLVSIFLHLV}_{191}$). En todos los casos, el extremo C-terminal es igual.

El análisis computacional de la G2 de todos los arnavirus permite detectar una homología estructural muy alta en esta región (**Figura 13.11 A y B**). El conjunto de bases experimentales y predictivas es consistente con la hipótesis de que la región transmembrana de las G2 de todos los arnavirus está formada, al menos, por una α hélice de 22 residuos de longitud ubicada con el extremo N-terminal hacia la cara extracelular de la membrana. El límite N-terminal real puede (o no) comenzar 5 residuos antes, en una zona definida con estructura tipo *loop* (**Figura 13.11.A**).

Por otra parte, aplicando a las α hélices transmembrana de los diferentes arnavirus el mismo tipo de análisis ya mencionado para las cepas del virus Junín, encontramos que, a pesar de la escasa similitud aminoacídica local (22,7 % de identidad y 59,1 % de homología), la similitud estructural en esta región es prácticamente del 100 % (**Figuras 13.11 A y B, 13.12 y 13.13**). En base a esto, se puede hipotetizar que las 4 hélices α_4 que soportan el anclaje transmembrana están dispuestas de acuerdo a un patrón común para todos los arnavirus analizados (ver **Figura 13.4**).

Además, en esta zona de las G2 también existe un motivo tipo *heptad repeat* (un aminoácido hidrofóbico cada 7 residuos) conservado en todos los arnavirus (**Figura 13.11.A**). En el contexto de la bicapa lipídica y, considerando exclusivamente la hélice α_4 predicha, los residuos 7, 14 y 21 están ubicados en la hemihélice de mayor $\langle H_i \rangle$ (**Figuras 13.12 y 13.13**) en todas las G2. Esta disposición peculiar de los residuos podría colaborar con la estabilización del anclaje transmembrana.

Por otra parte, un argumento adicional que sustenta la hipótesis del patrón estructural común, puede hallarse en la estructura α_2 . Esta estructura tiene ~50 residuos de longitud (en OLI, PIC y TAC son 2 α hélices, de 12 y 38 residuos de longitud respectivamente, separadas por 1 solo residuo). Burns & Buchmeier (1993) para LCM y Chambers *et al.* (1990) para Lassa han postulado que esta región podría constituir una zona con estructura de tipo *coiled-coil* que estabilizaría íntimamente la porción extracelular del tetrámero de G2. Las predicciones presentadas en este trabajo no contemplan el análisis de potenciales estructuras tipo *coiled-coil*. Sin embargo, el análisis con la base de datos PROSITE predice la existencia de un nonapéptido que actuaría como dominio de dimerización en Oliveros (${}_{73}\text{SRLNLLTNT}_{81}$) y en Tacaribe (${}_{73}\text{TRVNVLSHT}_{81}$).

Dada la escasa conservación de secuencia entre los dos motivos hallados en la

G2 (ver también **Capítulo 14**, sitios hallados en las proteínas N), se definió un consenso ampliado que comprende a todos los nonapéptidos hallados con el PROSITE en las proteínas de arnavirus: [CIKMNQSTV]-[ADKQRST]-[AIGLSTV]-[AHIDGNQST]-[ILV]-[FILKMQTV]-[GKNRSTY]-[DHKNQRT]-[ANSTV].

En este punto conviene realizar algunas consideraciones estadísticas. La población total de nonapéptidos posibles es de $5,1 \times 10^{11}$ miembros, mientras que la fracción de esa población que coincide con el consenso ampliado es de $2,3 \times 10^7$ y la probabilidad de encontrar uno de ellos es $4,5 \times 10^{-5}$ (frecuencias iguales para los distintos aminoácidos). Si esta probabilidad se corrige con las frecuencias de los distintos aminoácidos, existentes en G1 y G2, la probabilidad aumenta a $1,8 \times 10^{-4}$. Esto último implica que, para una proteína de ~220 aminoácidos (promedio de los alineamientos múltiples de G1 y G2), la frecuencia esperada de hallazgo de nonapéptidos sería de 0,04 sitios/proteína.

Teniendo en cuenta las consideraciones anteriores, se realizó una búsqueda exhaustiva en las dos glicoproteínas de los arnavirus, con el consenso ampliado.

En las G1, coincidentemente con lo esperado, no se encontró ningún nonapéptido coincidente con el consenso. Mientras que en las G2, se encontró 1 único sitio en todos los arnavirus, en la misma posición (**Figura 13.14**; ver también **Apéndice 6, Figura Ap6.3**).

Por otra parte, con la finalidad de evaluar la confiabilidad de esta predicción, se realizaron dos controles: se buscó la existencia de nonapéptidos coincidentes con el consenso en un conjunto de 20 secuencias aminoacídicas no relacionadas y se generó un conjunto de 20 secuencias con un orden aleatorio de los aminoácidos, basadas en las secuencias de G2 de los arnavirus y manteniendo las frecuencias promedio características. En ninguno de los dos conjuntos se encontró ningún sitio coincidente con el consenso propuesto.

En función de lo anterior, se postula que la zona equivalente a los nonapéptidos hallados con el PROSITE en Oliveros y Tacaribe (residuos 73 a 81), de los otros arnavirus podría actuar también como dominio de dimerización (**Figura 13.11.A**), y sería, en parte, responsable de estabilizar el tetrámero extracelular de G2, ya sea bajo la forma de 4 α hélices estabilizadas como conjunto por interacciones mutuas o de una estructura tipo *coiled-coil*. En cualquiera de los dos casos, en esta región de las G2 de todos los arnavirus, existe un motivo tipo *heptad repeat* (**Figura 13.11.A**). Como se ha descrito para la proteína F de paramyxovirus (Lamb & Kolakofsky, 1995), los *heptad repeats* pueden formar un *triple-stranded coil* formado por 3 α hélices. En el caso de los arnavirus, esta supraestructura debería ser un *tetra-stranded coil* formado por 4 α hélices.

Respecto del segundo punto, la bibliografía describe que en glicoproteínas de otros sistemas virales existen regiones de alta hidrofobicidad, llamadas péptidos de fusión (*fusion peptides*), que son cruciales para la actividad fusogénica de los virus; todos estos péptidos de fusión están presentes en proteínas ancladas en membrana (White, 1992).

En 1989, Glushakova & Lukashevich, describen la susceptibilidad de las etapas tempranas de la infección por el virus Lassa al tratamiento con agentes lisosomotrópicos. En 1990, Glushakova *et al.*, postulan la presencia de 6 potenciales péptidos de fusión en la GPC el virus Lassa, tres de ellos deberían descartarse por estar presentes en G1.

Uno o más de estos péptidos serían los responsables de la fusión de la membrana de la envoltura viral y la membrana endocítica. El mecanismo de fusión se activa por un descenso del pH de la vesícula endocítica, lo cual lleva a la disociación del tetrámero de G1 de las espículas y cambios conformacionales en las G2, con la concomitante exposición del péptido

de fusión que puede interactuar mediante interacciones electrostáticas o hidrofóbicas con la membrana endocítica y permitir la liberación de las nucleocápsides al citoplasma. Estas etapas del mecanismo de entrada se han estudiado para el virus LCM, cepa 4 (Di Simone *et al.*, 1994) y cepas 53b y clon 13 (Borrow y Oldstone, 1994).

En función de las predicciones de Glushakova *et al.* (1990) y de los estudios de Di Simone *et al.* (1994), se han ubicado los correspondientes péptidos de fusión potenciales en el resto de los arnavirus (ver **Figura 13.11.A y B**). Hasta el presente no se conoce si el péptido activo es uno sólo, o todos tienen algún tipo de participación.

El potencial péptido de fusión I, postulado como el candidato ideal por Di Simone *et al.* (1994), está ubicado exactamente en la posición N-terminal de las G2 y, en 7 de los 11 arnavirus analizados, los tres primeros aminoácidos se corresponden con un motivo GXF. Según Di Simone *et al.* (1994), este motivo se encuentra en el extremo N-terminal de péptidos de fusión de otros virus. Sin embargo, existen varias objeciones a esta postulación. En primera instancia, aunque el motivo existe en todas las proteínas HA₂ de influenza (orthomyxovirus) (Lamb & Krug, 1995), en el caso de los arnavirus el motivo no existe en todos; en CD1, MC2 y TAC, los tres primeros aminoácidos presentan un motivo AXF y en OLI es AXL. En segunda instancia, en los virus Sendai y Measles (paramyxovirus), el consenso de los tres primeros aminoácidos es FXG (Yeagle, 1993; Lamb and Kolakofsky, 1995). Por otra parte, el péptido no sería lo suficientemente hidrofóbico como para cumplir la función, en 8 de los 11 arnavirus la $\langle H_i \rangle$ varía en el rango de $-0,03$ a $0,17$ (en OLI es $0,26$; en TAC es $0,28$ y en SAB es $0,33$). Además, la similitud aminoacídica local es: $25,0\%$ de identidad y $68,8\%$ de homología. Esto último discrepa significativamente con lo

descrito para péptidos de fusión de otros sistemas virales; en los cuales la conservación de la identidad aminoacídica es extremadamente alta. En los péptidos de fusión de paramyxovirus la conservación de la identidad es de $\sim 90\%$ (Lamb and Kolakofsky, 1995).

El potencial péptido de fusión II, está adyacente al I y, en todos los casos, los cinco primeros aminoácidos están conservados en los 11 arnavirus (₁₇GGYCL₂₁). Por otra parte, este péptido es el de mayor conservación en cuanto al perfil local de hidrofobicidad, la $\langle H_i \rangle$ varía en el rango de $0,20$ a $0,42$. Además, la similitud aminoacídica local es: $44,0\%$ de identidad y $75,0\%$ de homología.

El potencial péptido de fusión III, está solapado con el dominio de dimerización. Por otra parte, el péptido no sería lo suficientemente hidrofóbico como para cumplir la función, en 7 de los 11 arnavirus la $\langle H_i \rangle$ varía en el rango de $-0,02$ a $0,17$ (en las cepas de LAS es $0,27$ y en las cepas de LCM es $0,20$). Además, la similitud aminoacídica local es: $16,7\%$ de identidad y $72,2\%$ de homología.

Analizando la hélice $\alpha 1$, presente en el potencial péptido de fusión II, mediante gráficos tipo *helical-wheel* y calculando las $\langle H_i \rangle$ de cada hemihélice (divididas mediante un plano arbitrario), se puede observar una marcada asimetría en la $\langle H_i \rangle$, la cual varía en un rango de $\delta \langle H_i \rangle = 0,37$ (OLI) a $\delta \langle H_i \rangle = 0,92$ (SAB) (**Figuras 13.15 y 13.16**). Por otra parte, en alphavirus se ha demostrado experimentalmente la importancia de un doblete GG para la actividad fusogénica y que el péptido de fusión puede estar lejos del N-terminal (75 residuos) (Levy-Mintz and Kielian, 1991). En función de estas consideraciones, se postula que el mejor candidato para todos los arnavirus, potencialmente responsable de la liberación de nucleocápsides al citoplasma, sería el péptido de fusión II. Esta conclusión, es coincidente con lo postulado por Glushakova *et al.* (1990), para 5 de los arnavirus estudiados en esta Tesis.

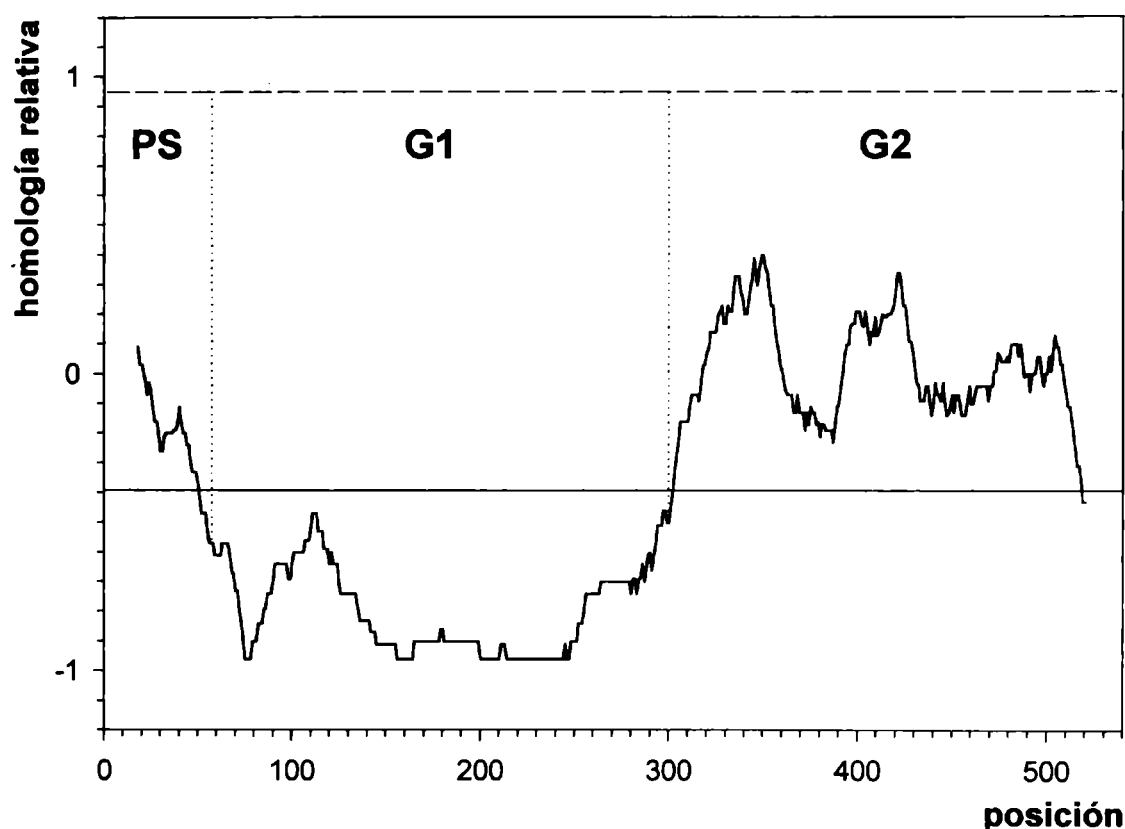


Figura 13.5. Regiones conservadas y regiones variables en las proteínas GPC de los arnavirus Junín-Candid #1, Junín-MC2, Oliveros, Pichindé, Sabiá, Tacaribe, Lassa-Josiah, Lassa-Nigeria, LCM-Armstrong, LCM-WE y Mopeia. El gráfico fué calculado con la rutina HOMGRAPH (Ghiringhelli, P.D., 2002, esta tesis), utilizando una ventana de 35 residuos y una tabla de equivalencias de +1,0 para las identidades, +0,5 y +0,25 para los diferentes grados de cambios conservativos (Schwartz and Dayhoff, 1979), y -1,0 para la ausencia de identidad. La línea llena horizontal indica la media de los valores de homología relativa. Los picos por encima de la media corresponden a las regiones conservadas y los picos por debajo corresponden a las regiones variables. En el gráfico se notan grandes contrastes, se observa un región conservada correspondiente al péptido señal (PS), seguida de una gran región no conservada correspondiente a la glicoproteína G1 (G1) y una región con alternancia de zonas con mayor o menor grado de conservación correspondiente a la glicoproteína G2 (G2). (Ver también Figuras Ap6.1, Ap6.2 y Ap6.3).

Tabla 13.b.1. Identidad de secuencias aminoacídicas. - Péptido señal

	NUEVO MUNDO						VIEJO MUNDO				
	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
CD1	100 %	84,5 %	63,8 %	53,5 %	63,8 %	82,8 %	46,6 %	46,6 %	37,9 %	36,2 %	44,8 %
MC2		100 %	56,9 %	46,6 %	56,9 %	74,1 %	39,3 %	39,3 %	33,3 %	35,0 %	41,7 %
OLI			100 %	50,0 %	69,0 %	60,3 %	48,3 %	48,3 %	43,1 %	43,1 %	51,7 %
PIC				100 %	46,6 %	62, %	56,9 %	56,9 %	43,1 %	41,4 %	53,5 %
SAB					100 %	62,1 %	53,5 %	53,5 %	48,3 %	44,8 %	51,7 %
TAC						100 %	48,3 %	48,3 %	37,9 %	36,2 %	48,3 %
LAG							100 %	100 %	60,3 %	58,6 %	75,9 %
LAJ								100 %	60,3 %	58,6 %	75,9 %
LCA									100 %	91,4 %	55,2 %
LCE										100 %	56,9 %
MOP											100 %

Tabla 13.b.2. Homología de secuencias aminoacídicas. - Péptido señal

	NUEVO MUNDO						VIEJO MUNDO				
	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
CD1	100 %	86,2 %	87,9 %	77,6 %	91,4 %	91,4 %	74,1 %	74,1 %	70,7 %	69,0 %	79,3 %
MC2		100 %	77,6 %	72,4 %	81,0 %	82,8 %	70,5 %	70,5 %	65,0 %	65,0 %	73,3 %
OLI			100 %	84,5 %	93,1 %	84,5 %	77,6 %	77,6 %	77,6 %	75,9 %	79,3 %
PIC				100 %	79,3 %	86,2 %	79,3 %	79,3 %	84,5 %	84,5 %	82,8 %
SAB					100 %	86,2 %	77,6 %	77,6 %	75,9 %	75,9 %	79,3 %
TAC						100 %	77,6 %	77,6 %	74,1 %	72,4 %	79,3 %
LAG							100 %	100 %	86,2 %	87,9 %	96,6 %
LAJ								100 %	86,2 %	87,9 %	96,6 %
LCA									100 %	100 %	86,2 %
LCE										100 %	86,2 %
MOP											100 %

CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe;

LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia

En negrita y grisado se destacan las comparaciones entre cepas de un mismo arnavirus

En negrita y sombreado amarillo claro se destacan los límites inferior y superior, entre arnavirus diferentes, de cada recuadro.

Tabla 13.c.1. Identidad de secuencias aminoacídicas. - Glicoproteína G1

	NUEVO MUNDO						VIEJO MUNDO				
	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
CD1	100 %	92,7 %	17,0 %	14,6 %	21,7 %	44,5 %	19,0 %	19,0 %	15,5 %	17,3 %	19,1 %
MC2		100 %	17,0 %	14,0 %	20,6 %	40,6 %	17,4 %	17,4 %	14,1 %	15,9 %	17,6 %
OLI			100 %	33,9 %	22,9 %	24,2 %	27,7 %	26,9 %	24,8 %	23,5 %	26,2 %
PIC				100 %	21,8 %	19,0 %	29,6 %	30,0 %	29,2 %	26,1 %	29,2 %
SAB					100 %	20,6 %	17,9 %	18,9 %	17,3 %	17,2 %	17,5 %
TAC						100 %	18,4 %	18,4 %	17,7 %	18,6 %	15,9 %
LAG							100 %	98,0 %	45,8 %	46,2 %	73,0 %
LAJ								100 %	45,8 %	46,2 %	72,5 %
LCA									100 %	91,8 %	44,6 %
LCE										100 %	43,7 %
MOP											100 %

Tabla 13.c.2. Homología de secuencias aminoacídicas. - Glicoproteína G1

	NUEVO MUNDO						VIEJO MUNDO				
	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
CD1	100 %	93,8 %	48,0 %	44,6 %	48,8 %	74,9 %	44,1 %	44,1 %	40,6 %	42,7 %	45,2 %
MC2		100 %	46,3 %	43,7 %	47,3 %	70,3 %	40,4 %	40,4 %	38,0 %	39,7 %	44,3 %
OLI			100 %	62,6 %	53,8 %	44,5 %	51,3 %	51,3 %	53,9 %	51,7 %	50,7 %
PIC				100 %	50,9 %	43,0 %	55,5 %	55,5 %	53,1 %	52,2 %	56,6 %
SAB					100 %	49,5 %	53,3 %	53,8 %	44,6 %	43,0 %	52,1 %
TAC						100 %	45,8 %	45,8 %	41,4 %	42,7 %	43,3 %
LAG							100 %	100 %	67,9 %	69,3 %	89,0 %
LAJ								100 %	68,4 %	68,9 %	89,0 %
LCA									100 %	96,6 %	66,2 %
LCE										100 %	66,7 %
MOP											100 %

CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe;

LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia

En negrita y grisado se destacan las comparaciones entre cepas de un mismo arnavirus

En negrita y sombreado amarillo claro se destacan los límites inferior y superior, entre arnavirus diferentes, de cada recuadro.

Tabla 13.d.1. Identidad de secuencias aminoacídicas. - Glicoproteína G2

	NUEVO MUNDO						VIEJO MUNDO				
	CDI	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
CDI	100 %	99,2 %	61,2 %	55,5 %	74,6 %	81,4 %	57,2 %	57,2 %	54,2 %	52,1 %	53,4 %
MC2		100 %	62,0 %	56,4 %	75,4 %	82,2 %	57,6 %	57,6 %	54,7 %	52,5 %	53,8 %
OLI			100 %	59,9 %	60,3 %	59,9 %	57,0 %	56,5 %	57,8 %	56,5 %	57,4 %
PIC				100 %	55,9 %	51,7 %	55,6 %	55,1 %	53,2 %	53,4 %	56,8 %
SAB					100 %	73,3 %	55,1 %	55,1 %	49,6 %	48,3 %	54,2 %
TAC						100 %	52,1 %	52,1 %	51,3 %	50,4 %	53,0 %
LAG							100 %	98,3 %	70,2 %	68,8 %	80,3 %
LAJ								100 %	70,2 %	68,8 %	79,8 %
LCA									100 %	94,9 %	68,1 %
LCE										100 %	66,2 %
MOP											100 %

Tabla 13.d.2. Homología de secuencias aminoacídicas. - Glicoproteína G2

	NUEVO MUNDO						VIEJO MUNDO				
	CDI	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
CDI	100 %	100 %	83,5 %	79,7 %	92,8 %	94,9 %	78,8 %	78,4 %	77,5 %	76,3 %	77,1 %
MC2		100 %	83,5 %	79,7 %	92,8 %	94,9 %	78,4 %	78,0 %	77,1 %	75,9 %	76,7 %
OLI			100 %	83,1 %	83,1 %	82,7 %	78,1 %	77,6 %	78,9 %	77,6 %	78,5 %
PIC				100 %	80,5 %	78,4 %	77,8 %	77,4 %	75,3 %	75,2 %	78,2 %
SAB					100 %	90,7 %	78,8 %	78,8 %	78,0 %	77,5 %	75,9 %
TAC						100 %	78,4 %	78,0 %	77,1 %	75,9 %	75,9 %
LAG							100 %	99,1 %	87,7 %	87,6 %	92,7 %
LAJ								100 %	87,7 %	87,6 %	92,7 %
LCA									100 %	98,7 %	88,9 %
LCE										100 %	88,9 %
MOP											100 %

CDI: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe;

LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia

En negrita y grisado se destacan las comparaciones entre cepas de un mismo arenavirus

En negrita y sombreado amarillo claro se destacan los límites inferior y superior, entre arenavirus diferentes, de cada recuadro.

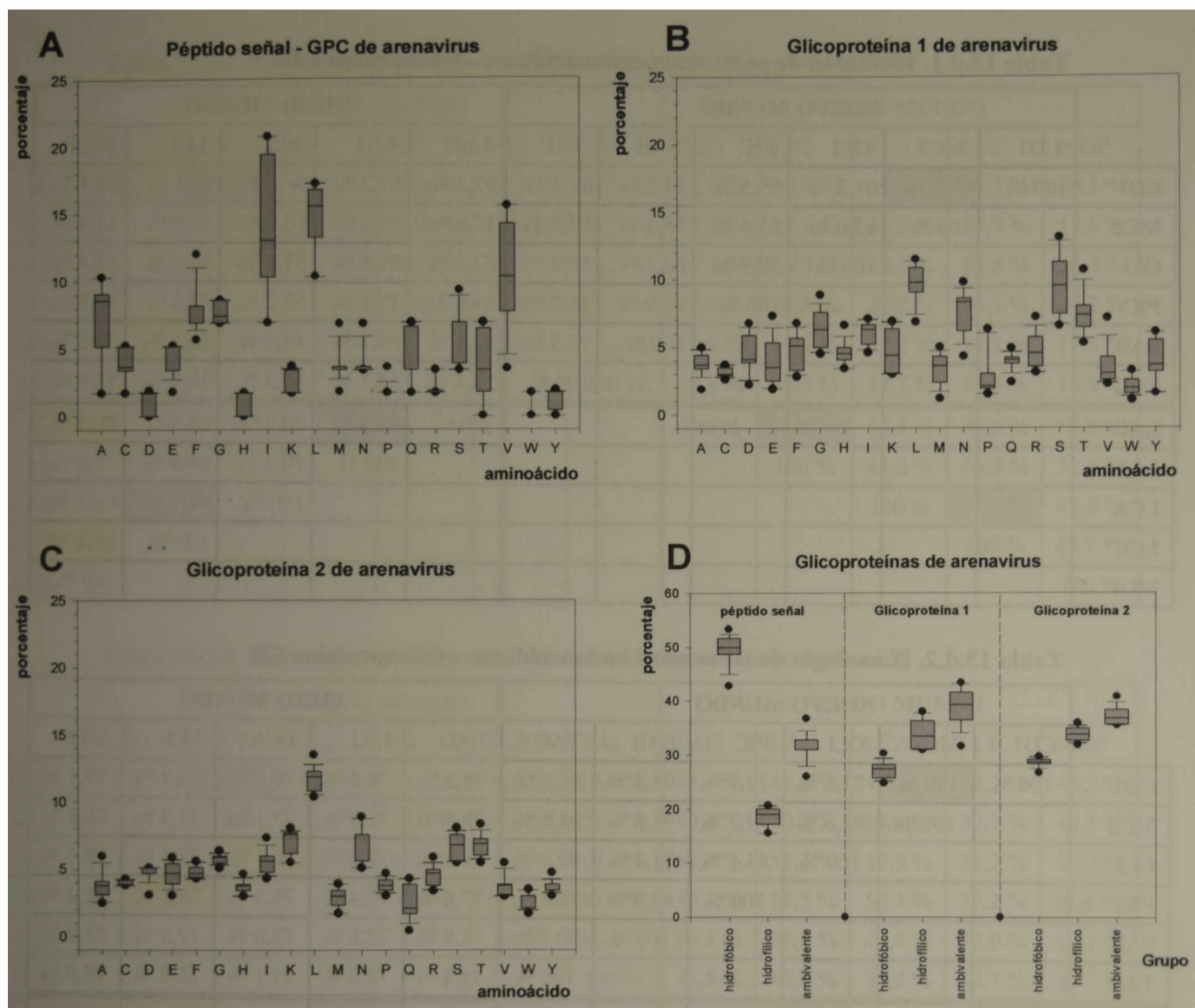


Figura 13.6. Gráficos de cajas mostrando las frecuencias y dispersión de los distintos aminoácidos en las proteínas GPC de los arnavirus. Los límites inferior y superior de cada caja corresponden al 25^{vo} percentil y al 75^{vo} percentil, respectivamente. Los límites inferior y superior de cada barra corresponden al 10^{mo} percentil y al 90^{vo} percentil, respectivamente. Los puntos inferior y superior corresponden al 5^o percentil y al 95^o percentil, respectivamente. La línea horizontal dentro de la caja corresponde a la mediana (50^{vo} percentil) de los valores. La distancia desde la mediana al límite superior de la caja indica la proporción de frecuencias presente en el rango 50^{vo} percentil - 75^{vo} percentil. La distancia desde la mediana al límite inferior de la caja indica la proporción de frecuencias presente en el rango 25^{vo} percentil - 50^{vo} percentil. **A.** Péptido señal. Se observa que hay una gran proporción de aminoácidos hidrofóbicos (en particular isoleucina, leucina y valina) y ambivalentes, mientras que los aminoácidos hidrofílicos son muy escasos. Por otra parte, la amplitud de las cajas nos indica que, en la mayor parte de los aminoácidos, los porcentajes de cada uno, en los distintos arnavirus, presentan gran dispersión. **B.** Glicoproteína G1. Los porcentajes de cada uno, en los distintos arnavirus, no difieren en gran medida. **C.** Glicoproteína G2. Es el péptido que presenta mayor conservación en cuanto a la frecuencia de los diferentes aminoácidos. Es destacable la alta proporción de leucina *versus* el resto. Por otra parte, la amplitud de las cajas nos indica que, en la mayor parte de los aminoácidos, los porcentajes de cada uno, en los distintos arnavirus, están bastante conservados. **D.** Frecuencias y distribución de los aminoácidos agrupados de acuerdo a Karlin *et al.* (1990). En el péptido señal se resaltan las peculiaridades ya mencionadas, muy alta proporción de residuos hidrofóbicos y escasa proporción de hidrofílicos. Mientras que, en G1 y G2, las frecuencias de los distintos

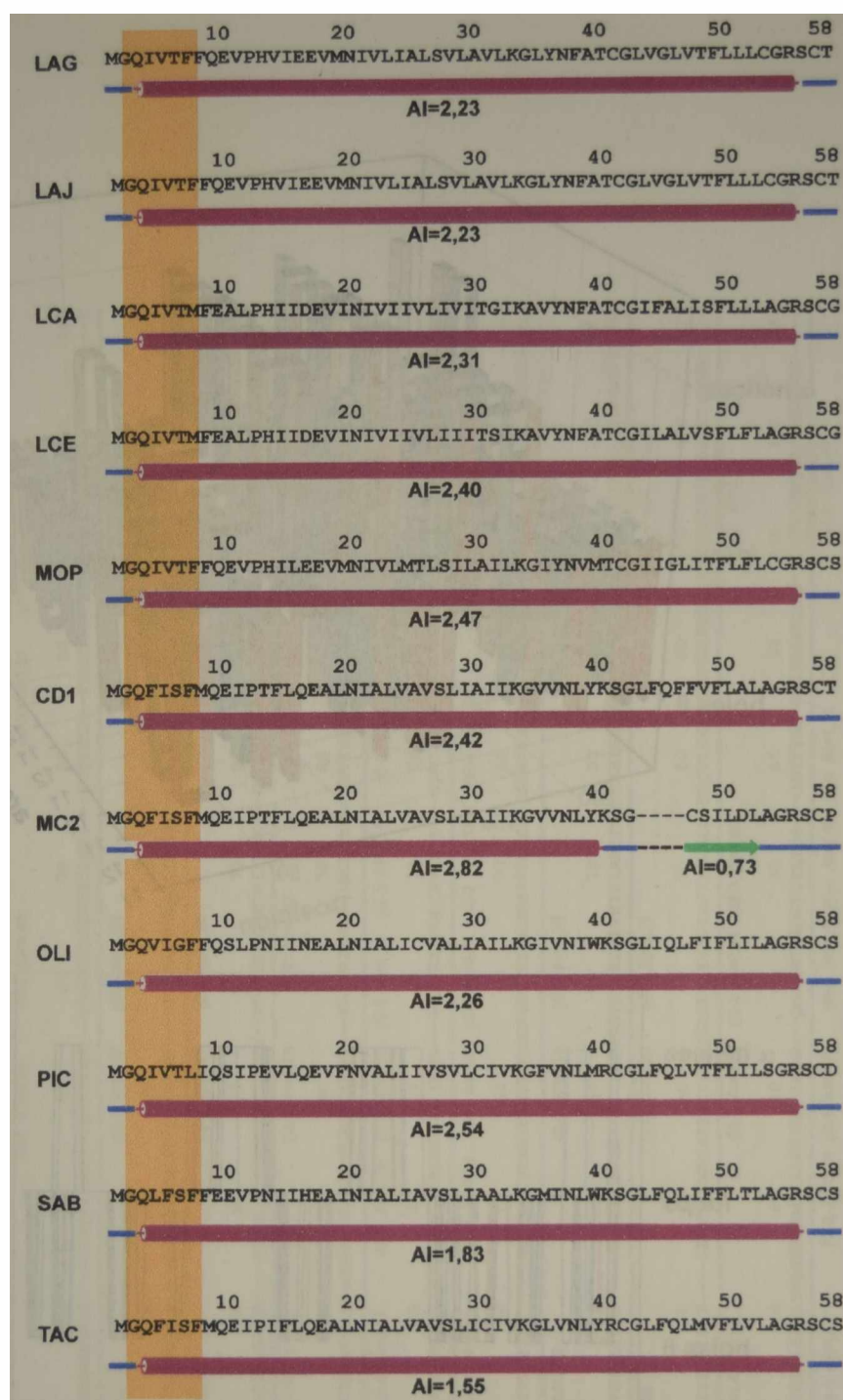


Figura 13.7. Comparación de estructuras secundarias predichas para el péptido señal de los arnavirus. En todos los casos se representa sólo el consenso de la predicción de estructura secundaria obtenido con el servidor JPRED (<http://circinus.ac.uk/jpred>). (ver Capítulo 17). Las predicciones para cada arnavirus se realizaron en forma independiente. Se observa que, con excepción de MC2, para todos los arnavirus analizados la predicción es la misma y corresponde a una α hélice de 53 residuos de longitud. En MC2 la predicción corresponde a una α hélice de 38 residuos de longitud, separada por 3 residuos de una estructura β de 6 aminoácidos. En todos los casos, los extremos fueron predichos como zonas sin estructura definida. Debajo de cada estructura se indica el AI (índice de anifipaticidad), calculado de acuerdo a Cornette *et al.* (1987) (ver Apéndice 5), con un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis). Aún cuando en la mayoría de los arnavirus el AI es superior a 2, lo cual sugiere anifipaticidad, de acuerdo a un análisis detallado de los períodos predominantes (no mostrado aquí, ver Apéndice 7) y a la alta frecuencia de aminoácidos hidrofóbicos y ambivalentes, se debe concluir que estas estructuras son preferentemente no anifipáticas. Con sombreado anaranjado se indica el sitio de miristoilación conservado. LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia; CD1: Junin-Candid #1; MC2: Junin-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe.

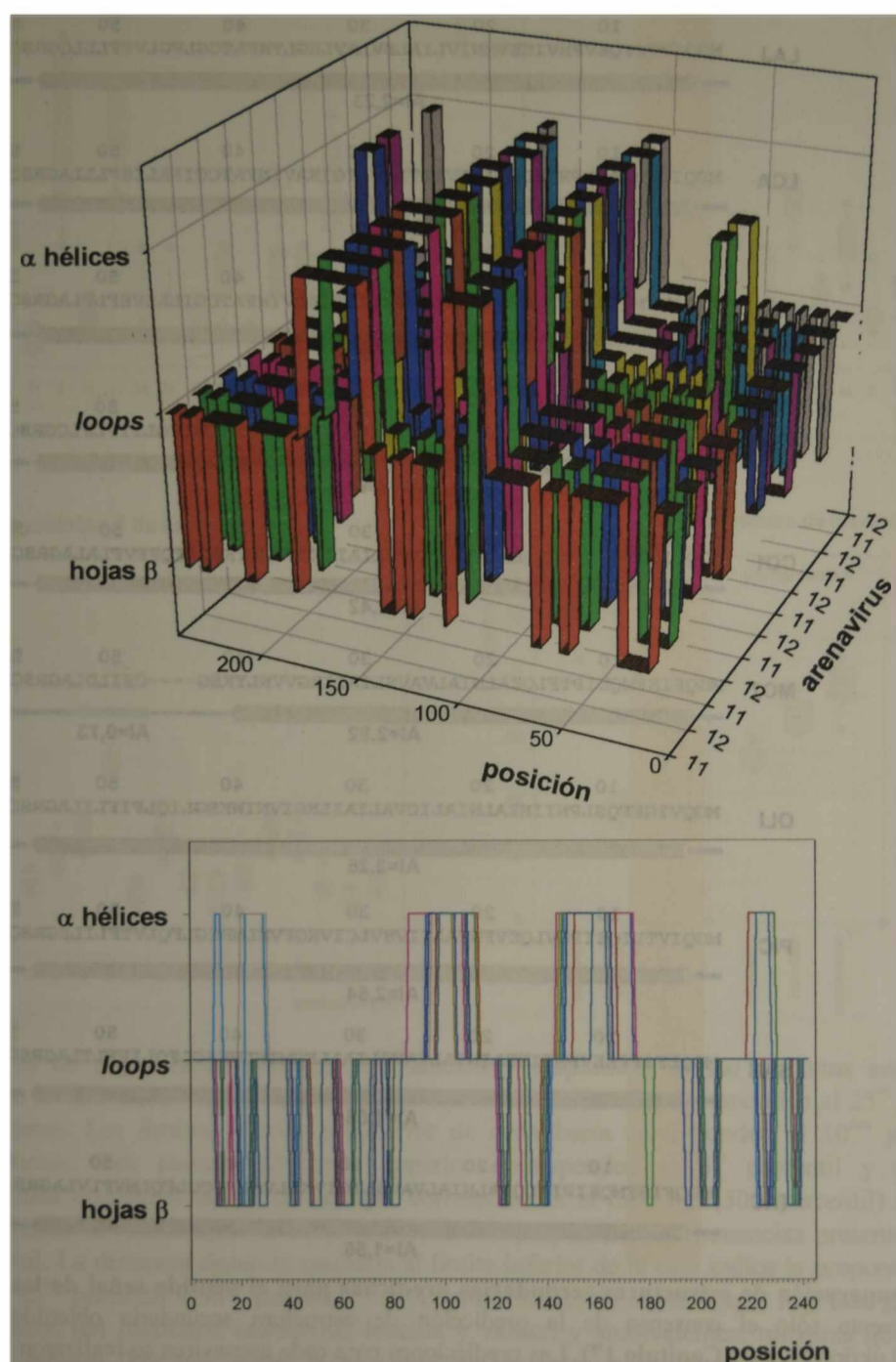


Figura 13.8. Gráficos de distribución de las estructuras secundarias en las glicoproteínas G1 de los arenavirus. Las estructuras secundarias predichas con el servidor JPRED corresponden a 3 categorías: α hélices, hojas β y *loops* (ver también **Figura 13.9**). Con la finalidad de construir el gráfico, a cada uno de los tipos estructurales se le asignó un valor arbitrario (α hélices= +1; hojas β = -1; *loops*= 0), con un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis). La representación, posición por posición, se realizó en función del alineamiento de secuencias y estructuras de todos los arenavirus analizados. En el panel superior se muestran las curvas individuales para cada arenavirus en un gráfico tridimensional. En el panel inferior, se representan todas las curvas individuales solapadas en un gráfico bidimensional, en el cual se puede observar mejor la variabilidad existente en las glicoproteínas G1. LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE- LCM-WE; MOP: Mopeia; CDI: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe.

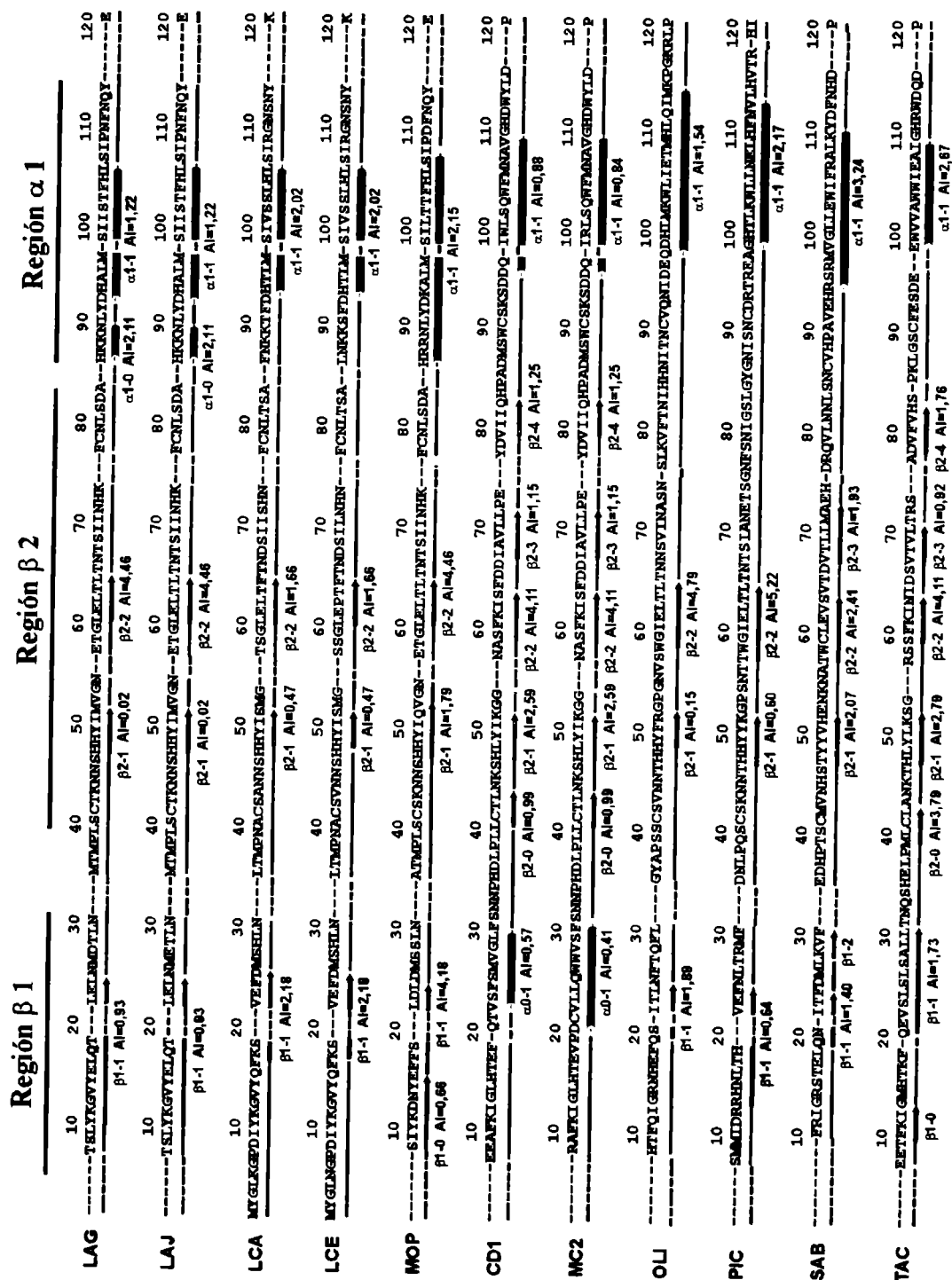


Figura 13.9. Comparación de estructuras secundarias predichas para las glicoproteínas G1 de los arnavirus. En todos los casos se representa sólo el consenso de la predicción de estructura secundaria obtenido con el servidor JPRED (<http://circinus.ac.uk/jpred>) (ver Apéndice 8). Las predicciones para cada arnavirus se realizaron en forma independiente. Se observa que las predicciones son diferentes para la mayor parte de los arnavirus analizados, con la excepción de las distintas cepas de un mismo arnavirus entre sí (Lassa, LCM y Junin). A pesar de las diferencias, se pueden reconocer diferentes regiones donde predominan las estructuras tipo α hélice o tipo hoja β. En función de esto, se han definido 4 regiones de α hélice. Las denominaciones de las distintas estructuras se han numerado de acuerdo a la región. (continúa en la página siguiente).

	Región β 3					Región α 2					Región β 4								
	130	140	150	160	170	180	190	200	210	220	230	240							
LAG	AMSCDFNGK--ISVOTMLSRBYAGDAN--	HCQVYANGVLTQFEMARGGSYIALDSGRGWD---	CIMTSYQYLLIQHTWDSQCFBFBSPISGTYLGLLSGRTRDIYSRRLL																
β 3-0				α 2-1 AI=2,85															
LAJ	AMSCDFNGK--ISVOTMLSRBYAGDAN--	HCQVYANGVLTQFEMARGGSYIALDSGRGWD---	CIMTSYQYLLIQHTWDSQCFBFBSPISGTYLGLLSGRTRDIYSRRLL																
β 3-0				α 2-1 AI=2,85															
LCA	AVBCDFRNG--ITIQYMLSRBYAGDAN--	HCQVYANGVLTQFEMARGGSYIALDSGRGWD---	CIMTSYQYLLIQHTWDSQCFBFBSPISGTYLGLLSGRTRDIYSRRLL																
β 3-0				α 2-1 AI=2,70															
LCE	AVBCDFRNG--ITIQYMLSRBYAGDAN--	HCQVYANGVLTQFEMARGGSYIALDSGRGWD---	CIMTSYQYLLIQHTWDSQCFBFBSPISGTYLGLLSGRTRDIYSRRLL																
β 3-0				α 2-1 AI=3,22															
MOP	AMSCDFNGK--ISVOTMLSRBYAGDAN--	HCQVYANGVLTQFEMARGGSYIALDSGRGWD---	CIMTSYQYLLIQHTWDSQCFBFBSPISGTYLGLLSGRTRDIYSRRLL																
β 3-0				α 2-1 AI=3,20															
CD1	PFLCNRNKT--EGEIFQVPTKTKGIDEN--	YAKAFKGMHLYREYFDS	CIDGKLCIMRQAPTSWPLQCFLDHVNTHLFTGRNIQL--PR-RSLK																
β 3-1 AI=0,99				α 2-1 AI=2,92															
MC2	PFLCNRNKT--EGEIFQVPTKTKGIDEN--	YAKAFKGMHLYREYFDS	CIDGKLCIMRQAPTSWPLQCFLDHVNTHLFTGRNIQL--PR-RSLK																
β 3-1 AI=0,99				α 2-1 AI=2,92															
OLI	PILCEKRG--LLELYMNTLASREKHSYFWSQYLLGLSKLSSKLFYDQCORADCMQDSEHSILKCNYSKSGEYFYLLIQHTWDSQCFBFBSPISGTYLGLLSGRTRDIYSRRLL	HCQVYANGVLTQFEMARGGSYIALDSGRGWD---	CIMTSYQYLLIQHTWDSQCFBFBSPISGTYLGLLSGRTRDIYSRRLL																
β 3-1 AI=1,81				α 2-1 AI=2,28															
PIC	GARKYVGGVLIQYMLSRBYAGDAN--	RELLASLAQIIQDPKIANVGGKFN--NCSGDTCKLNT--	CEGZHYNFLIQHTWDSQCFBFBSPISGTYLGLLSGRTRDIYSRRLL																
β 3-0				α 2-1 AI=2,34															
SAB	TPLCQKQTSVWETRWQINITEFGSGK--	FEDTILQRLGVLEGSRIAFSN--	IQDGLKRRFLIRNISTWKNQCEMHHVNSMIMLZANAGRSS--GSRAPL																
β 3-1 AI=2,64				α 2-1 AI=1,99															
TAC	GULLCRNKT--EGKLIQINLSRADGNVH--	YGRLENGLDHYRGEPEP--	CFEQGLIKIQEDPFTDCKADHTTFRLSRQKSLA--VG-RTIK																
β 3-1 AI=0,59				α 2-1 AI=1,83															

(Viene de la página anterior.)
 A su vez, en cada región se tomó el criterio de asignar como primer estructura a la conservada en posición en la mayor parte de los arnavirus en el alineamiento. En el caso de la existencia de estructuras del mismo tipo previas a la conservada, se las denominó con el número 0. Por otra parte, en todos los arnavirus, se han predicho zonas de estructura tipo loop que, en realidad, pueden ser simplemente loops o contener β turns (no analizados). Para facilitar la comparación, se han introducido gaps, en base a un alineamiento múltiple de secuencias (ver Apéndice 6), tanto en las estructuras secundarias como en las secuencias. En todas las estructuras formadas por 4 o más residuos se ha calculado el AI de acuerdo a Cornette *et al.* (1987) (ver Apéndice 5). El mismo se indica debajo de cada estructura. LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia; CD1: Junin-Candid #1; MC2: Junin-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabia; TAC: Tacaribe.

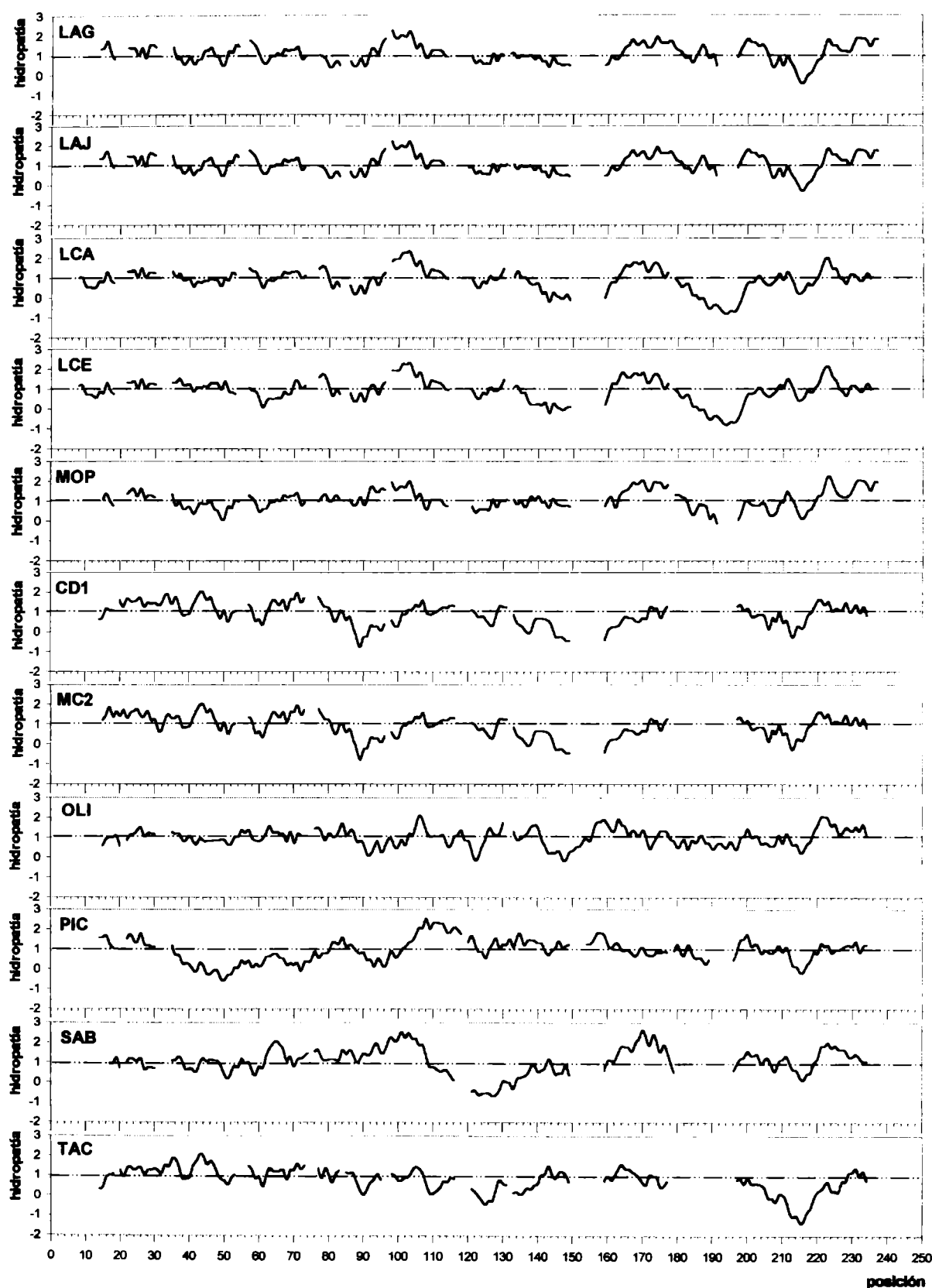


Figura 13.9.B. Perfiles de hidropatía de las glicoproteínas G1 de los arnavirus. Los perfiles fueron calculados utilizando ventanas sucesivas de 15 residuos. Con la finalidad de poder comparar entre sí los distintos gráficos, se introdujeron *gaps* en los perfiles respectivos, con un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis). Las posiciones de los *gaps* derivan de un alineamiento múltiple (ver Apéndice 6, Figura Ap.6.2). La línea de guiones y puntos ubicada en hidropatía 1, sirve solamente como referencia; los picos por encima de la línea corresponden a regiones de mayor hidrofobicidad y los que están por debajo a regiones más hidrofílicas. LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE- LCM-WE; MOP: Mopeia; CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe. Para los cálculos de hidrofobicidad se utilizó la tabla PRIFT (Cornette *et al.*, 1987).

Tabla 13.e.1. Similitud entre perfiles de hidropatía de las glicoproteínas G1 de arnavirus

	NUEVO MUNDO						VIEJO MUNDO				
	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
CD1	100 %	95,9 %	57,3 %	59,8 %	64,8 %	85,2 %	71,7 %	71,9 %	74,2 %	72,9 %	71,7 %
MC2		100 %	56,0 %	58,0 %	63,5 %	82,3 %	70,2 %	70,4 %	73,5 %	72,3 %	70,4 %
OLI			100 %	62,5 %	66,9 %	52,6 %	60,0 %	60,2 %	64,4 %	63,6 %	58,2 %
PIC				100 %	76,1 %	58,3 %	60,7 %	60,6 %	60,2 %	60,4 %	63,7 %
SAB					100 %	65,7 %	65,2 %	65,0 %	65,3 %	65,6 %	68,4 %
TAC						100 %	73,2 %	73,2 %	73,0 %	72,6 %	75,2 %
LAG							100 %	97,9 %	79,2 %	78,5 %	86,2 %
LAJ								100 %	79,1 %	78,5 %	86,4 %
LCA									100 %	96,4 %	84,7 %
LCE										100 %	85,0 %
MOP											100 %

Tabla 13.e.2. Distancia entre perfiles de hidropatía de las glicoproteínas G1 de arnavirus

	NUEVO MUNDO						VIEJO MUNDO				
	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
CD1	0,0 %	4,1 %	42,7 %	40,2 %	35,2 %	14,8 %	28,3 %	28,1 %	25,8 %	27,1 %	28,3 %
MC2		0,0 %	44,0 %	42,0 %	36,6 %	17,7 %	29,8 %	29,6 %	26,5 %	27,7 %	29,6 %
OLI			0,0 %	37,5 %	33,1 %	47,4 %	40,0 %	39,8 %	35,6 %	36,4 %	41,8 %
PIC				0,0 %	23,9 %	41,7 %	39,3 %	39,4 %	39,9 %	39,6 %	36,3 %
SAB					0,0 %	34,3 %	34,8 %	35,0 %	34,7 %	34,4 %	31,6 %
TAC						0,0 %	26,8 %	26,8 %	27,0 %	27,4 %	24,8 %
LAG							0,0 %	2,1 %	20,9 %	21,5 %	13,9 %
LAJ								0,0 %	20,9 %	21,5 %	13,6 %
LCA									0,0 %	3,6 %	15,3 %
LCE										0,0 %	15,0 %
MOP											0,0 %

CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe;

LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia

En **negrita y grisado** se destacan las comparaciones entre cepas de un mismo arnavirus

En **negrita y sombreado amarillo claro** se destacan los límites inferior y superior, entre distintos arnavirus, de cada recuadro.

Nota: Los porcentajes de similitud de perfiles de hidropatía fueron calculados utilizando ventanas de 15 valores (ventanas parcialmente solapadas, desplazadas 1 valor). A continuación, se aplicó la transformada de Fourier a cada una de las ventanas y se realizó la comparación de manera análoga a las comparaciones de perfiles de LCC (ver **Capítulo 18. Comparación cuantitativa de LCCs**) (Ghiringhelli, P.D., 2002, esta Tesis).

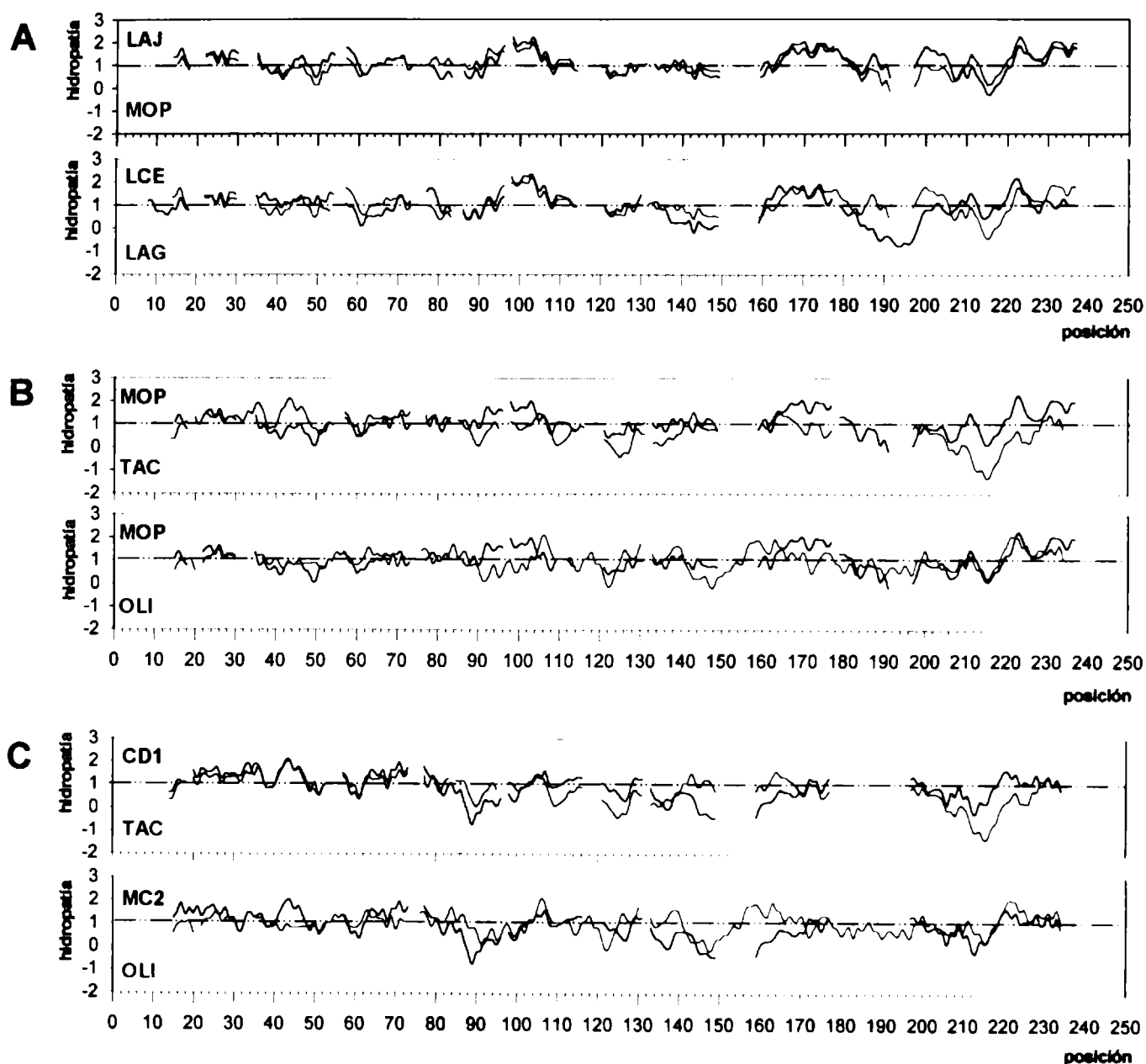


Figura 13.9.C. Comparación de perfiles de hidropatía de pares seleccionados de glicoproteínas G1 de arnavirus. Los perfiles fueron calculados utilizando ventanas sucesivas de 15 residuos. Con la finalidad de poder comparar entre sí los distintos gráficos, se introdujeron *gaps* en los perfiles respectivos, con un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis). Las posiciones de los *gaps* derivan de un alineamiento múltiple (ver Apéndice 6, Figura Ap.6.2). La línea de guiones y puntos ubicada en hidropatía 1, sirve solamente como referencia; los picos por encima de la línea corresponden a regiones de mayor hidrofobicidad y los que están por debajo a regiones más hidrofílicas. **A.** Los perfiles superpuestos corresponden a los pares de mayor similitud (LAJ/MOP; 86,4 %) y de menor similitud (LCE/LAG; 78,5 %) del grupo de los arnavirus del Viejo Mundo. **B.** Los perfiles superpuestos corresponden a los pares de mayor similitud (MOP/TAC; 75,2 %) y de menor similitud (MOP/OLI; 58,2 %) de la comparación entre los arnavirus del Viejo Mundo y los del Nuevo Mundo. **C.** Los perfiles superpuestos corresponden a los pares de mayor similitud (CD1/TAC; 85,2 %) y de menor similitud (MC2/OLI; 56,0 %) del grupo de los arnavirus del Nuevo Mundo (ver Tabla 13.e.1). LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE- LCM-WE; MOP: Mopeia; CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe.

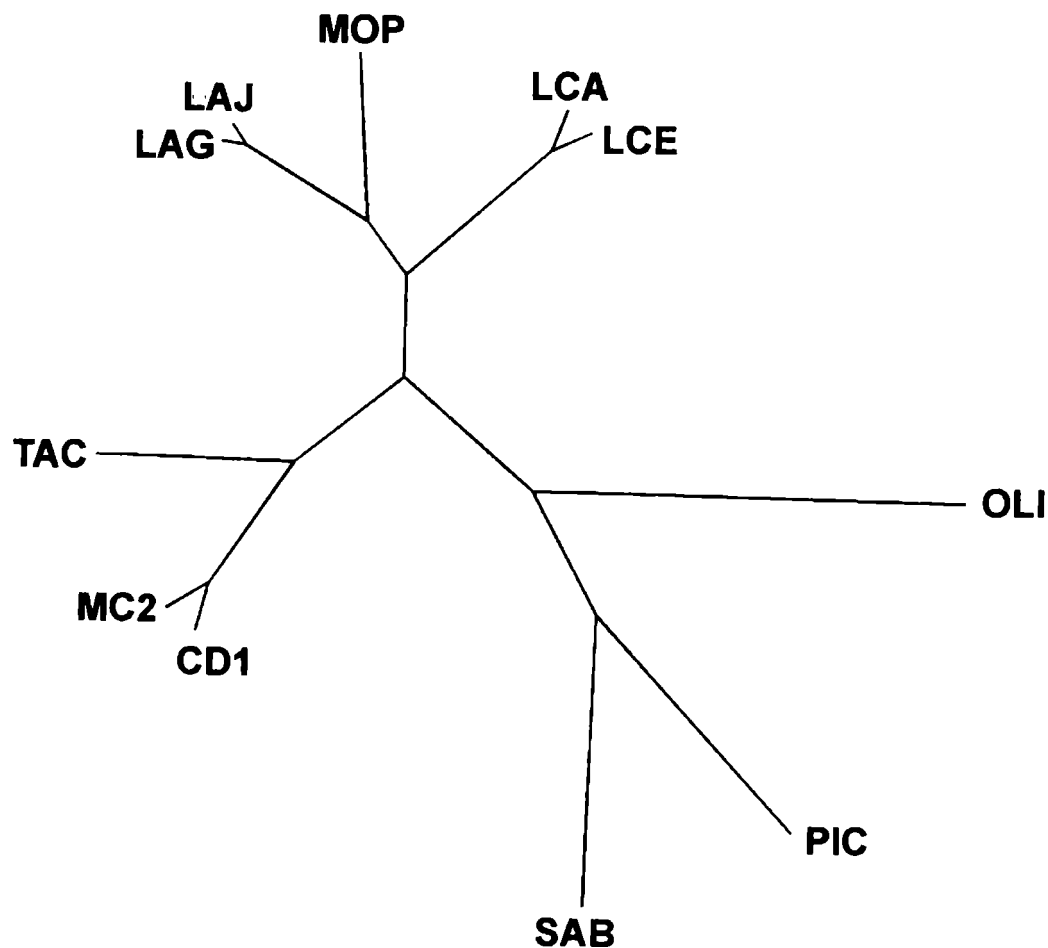


Figura 13.9.D. Cladograma correspondiente a relaciones estructurales entre las G1 de los arnavirus. El cladograma está basado en datos de similitud de los perfiles de hidropatía, y fue calculado utilizando el método UPGMA (*Unweighted Pair Group Arithmetic Average Clustering*). Como los datos de similitud representan la proximidad entre cada par de secuencias, la tabla de valores utilizada fue calculada asumiendo como distancias filogenéticas los valores resultantes del cálculo: $100 - v_n$ (donde v_n = valor individual) para cada par de secuencias (Tabla 13.e.2). Estas tablas de distancias fueron analizadas con la rutina NEIGHBOR (método UPGMA) y el cladograma fue dibujado con la rutina DRAWTREE, del paquete de programas PHYLIP (*Phylogeny Inferencing Package*, v 3.5, Felsenstein, J., 1995). Las longitudes de las distintas ramas en el cladograma están en relación con la distancia evolutiva. CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia.

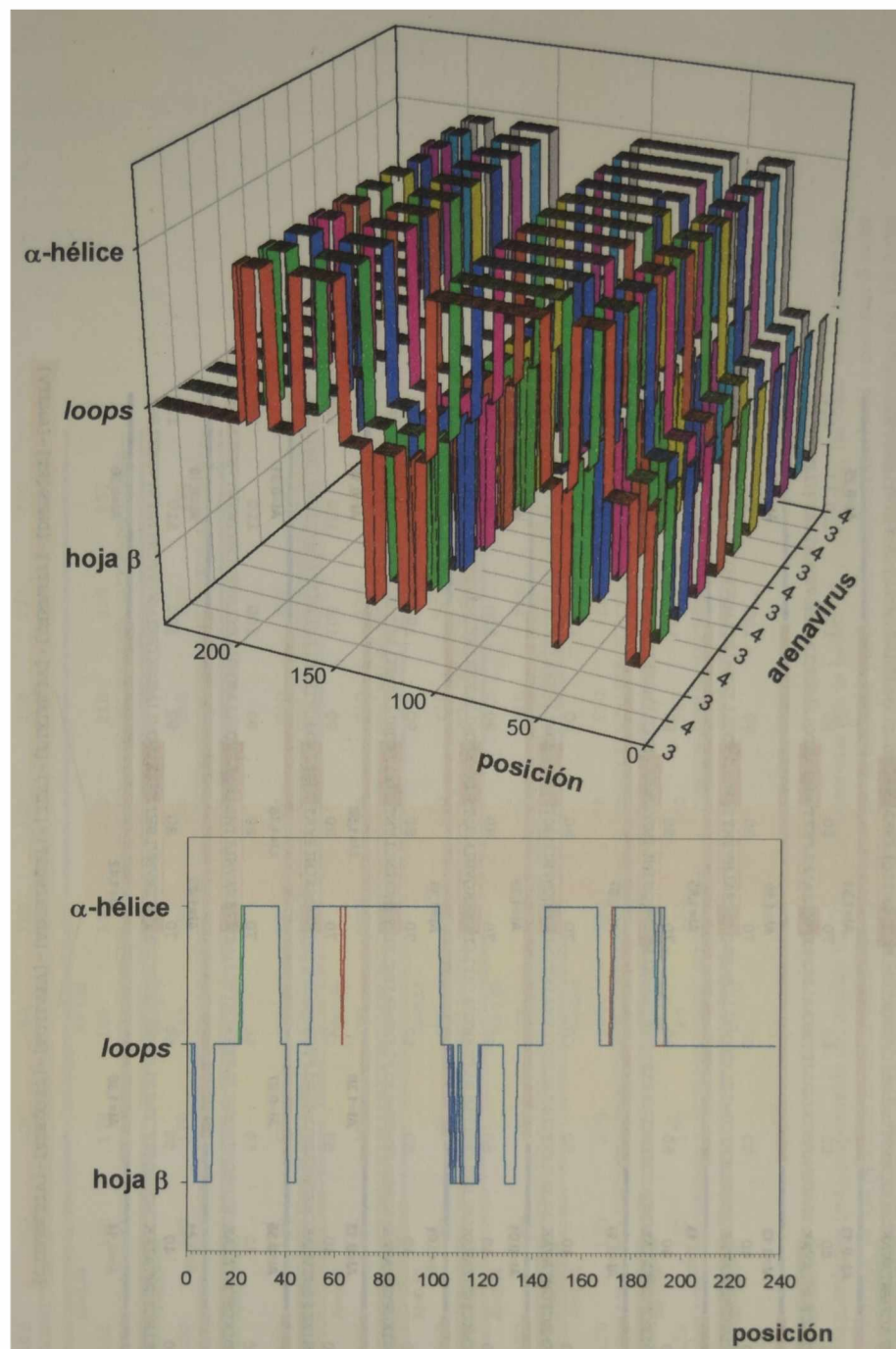


Figura 13.10. Gráficos de distribución de las estructuras secundarias en las glicoproteínas G2 de los arnavirus. Las estructuras secundarias predichas con el servidor JPRED corresponden a 3 categorías: α hélices, hojas β y *loops* (ver también **Figura 13.11**). Con la finalidad de construir el gráfico, a cada uno de los tipos estructurales se le asignó un valor arbitrario (α hélices= +1; hojas β = -1; *loops*= 0), con un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis). La representación, posición por posición, se realizó en función del alineamiento de secuencias y estructuras de todos los arnavirus analizados. En el panel superior se muestran las curvas individuales para cada arnavirus en un gráfico tridimensional. En el panel inferior, se representan todas las curvas individuales solapadas en un gráfico bidimensional, en el cual se puede observar mejor el alto grado de conservación de las estructuras predichas. LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE- LCM-WE; MOP: Mopeia; CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe.

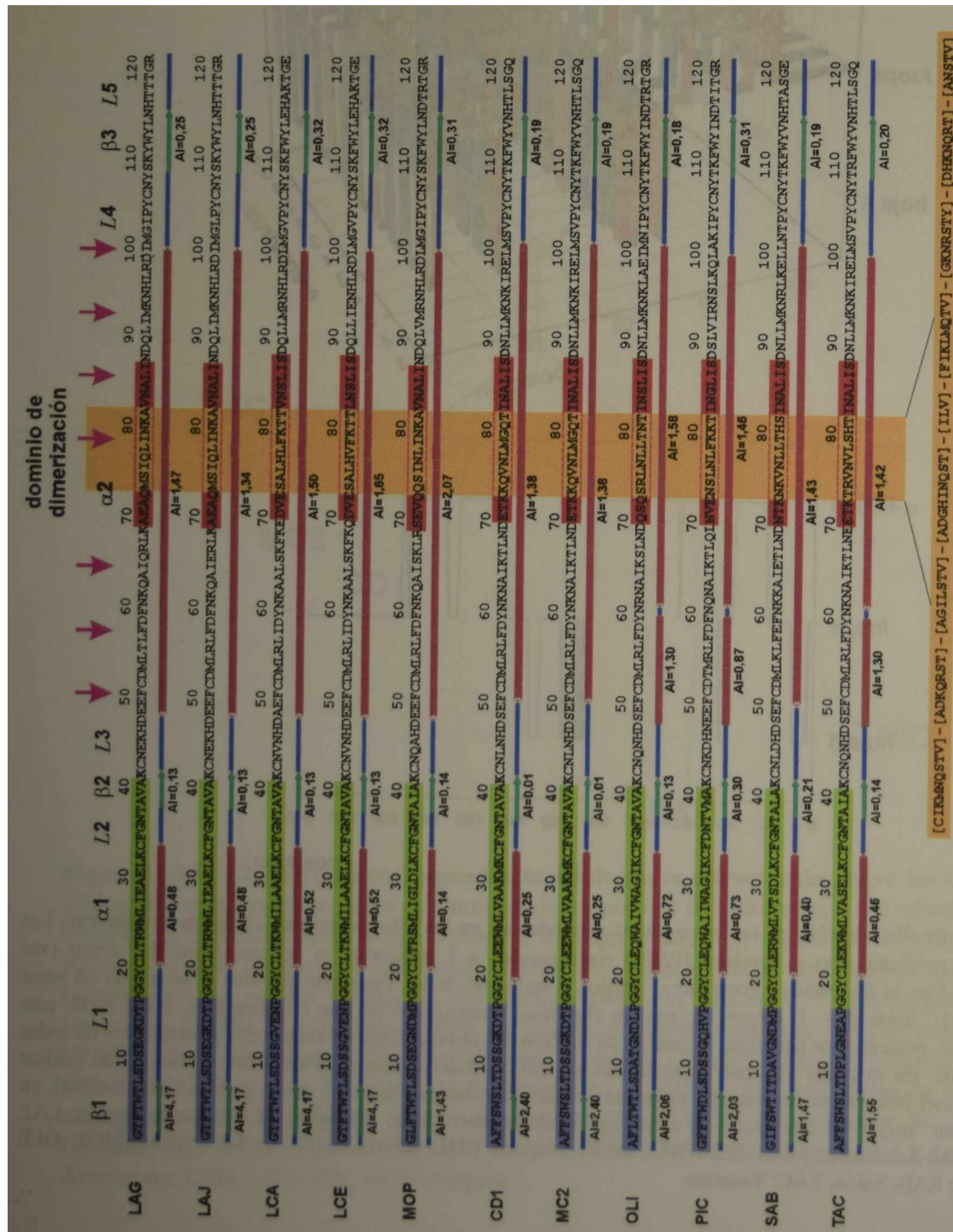
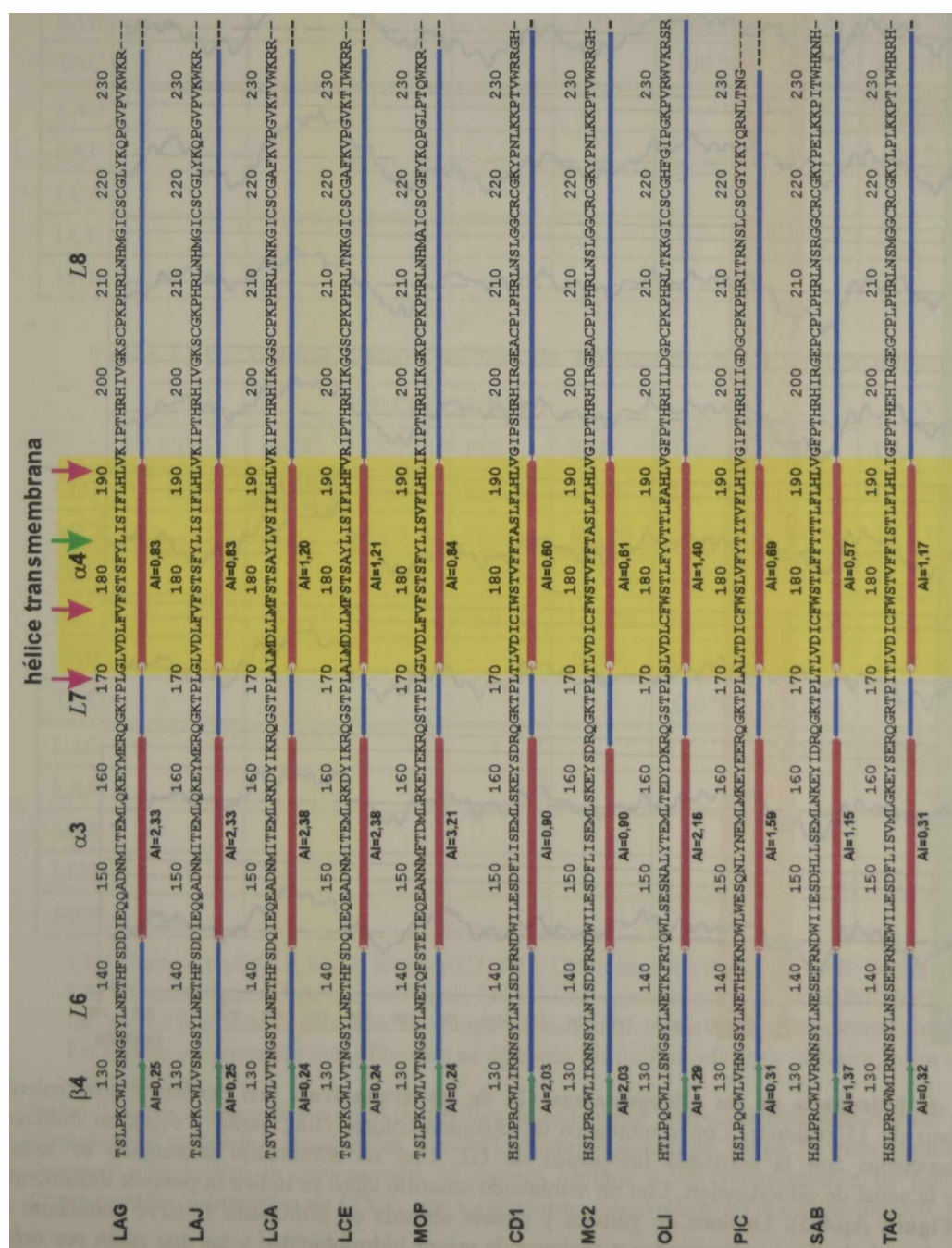


Figura 13.11.A. Comparación de estructuras secundarias predichas para las glicoproteínas G2 de los arnavirus. En todos los casos se representa sólo el consenso de la predicción de estructura secundaria obtenido con el servidor JPRED (<http://cirrus.us.uk/jpred>) (ver Apéndice 8). Las predicciones para cada arnavirus se realizaron en forma independiente. Se observa que, salvo ligeras diferencias, las predicciones son iguales para todos los arnavirus analizados. En función de esto, se han identificado 4 regiones de hoja β , 4 regiones de α hélice y 8 regiones definidas como *loops*. Todas las estructuras se han numerado secuencialmente. Con sombreados de diferentes colores (hila, verde claro y rojo), sobre las secuencias, se indican tres regiones aminoacídicas potencialmente asociadas con la actividad fusogénica de G2. Con un sombreado naranja claro se señala el nonapéptido potencialmente responsable de estabilizar la asociación tetramérica de la porción extracelular de G2 (señal de dimerización) y, en el recuadro inferior un consenso de la señal de dimerización, ampliado de acuerdo a los sitios hallados en la G2 y las proteínas N (ver Capítulo 14). Las flechas de color fucsia indican la presencia de aminoácidos hidrofóbicos repetidos cada 7 residuos (*heptad repeats*); según lo descripto (Yu *et al.*, 1996), en un entorno acuoso, los *heptad repeats* colaboran en la estabilización de interacciones proteína-proteína. (Continúa en la página siguiente).



(Viene de la página anterior)

Con un sombreado amarillo se indica la hélice transmembrana ($\alpha 4$), responsable del anclaje de G2 y, conjuntamente con la hélice $\alpha 2$, responsable de la estabilidad del tetramero de G2. Por otra parte, en todos los arnavirus, se han predicho zonas de estructura tipo *loop* que, en realidad, pueden ser simplemente *loops* o contener β turns (no analizados). Dada la mayor similitud de las secuencias polipeptídicas, no ha sido necesario introducir *gaps*, salvo en el extremo C-terminal (ver Apéndice 6). En todas las estructuras formadas por 4 o más residuos se ha calculado el AI de acuerdo a Cornette *et al.* (1987) (ver Apéndice 5). El mismo se indica debajo de cada estructura. Las flechas color fucsia en la región transmembrana indican aminoácidos hidrofóbicos repetidos cada 7 posiciones, *heptad repeats*, (en verde se señala una columna no conservada totalmente). En este contexto, los *heptad repeats* podrían colaborar en la estabilización de G2 en la bicapa lipídica. LAG: Lassa-Nigeria; LAJ: Lassa-Nigeria; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia; CDI: Junin-Candid #1; MC2: Junin-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabia; TAC: Tacaribe.

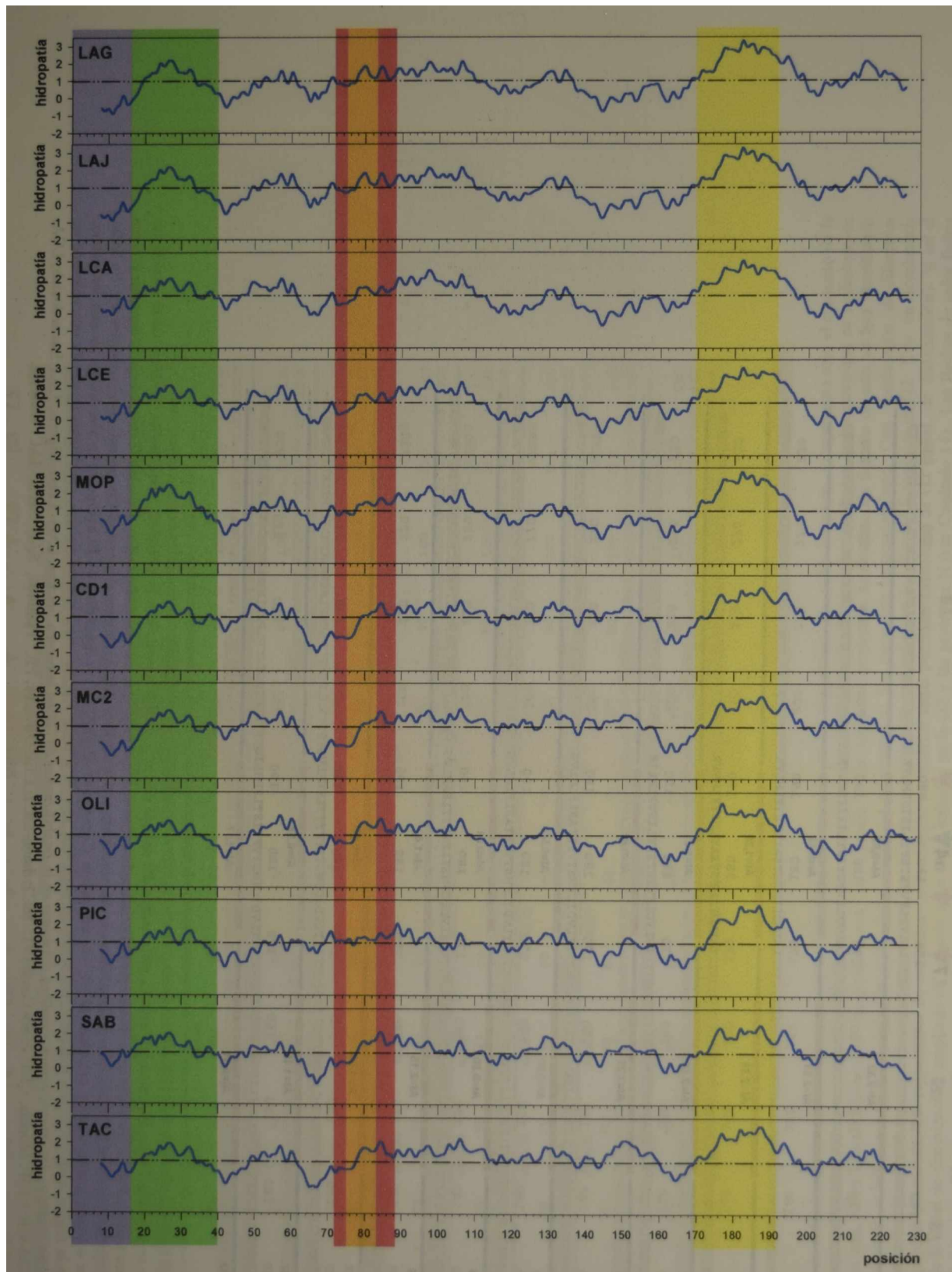


Figura 13.11.B. Perfiles de hidropatía de las glicoproteínas G2 de los arenavirus. Los perfiles fueron calculados utilizando ventanas sucesivas de 15 residuos. Con sombreados de distintos colores (lila, verde y rojo), se indican tres regiones potencialmente asociadas con la actividad fusogénica de G2. Con un sombreado anaranjado se señala el nonapéptido que constituye la señal de dimerización. Con un sombreado amarillo claro se indica la porción transmembrana de G2 (ver Apéndice 6, Figura Ap.6.3). La línea de guiones y puntos ubicada en hidropatía 1, sirve solamente como referencia; los picos por encima de la línea corresponden a regiones de mayor hidrofobicidad y los que están por debajo a regiones más hidrofílicas. LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE- LCM-WE; MOP: Mopeia; CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe. Para los cálculos de hidrofobicidad se utilizó la tabla PRIFT (Cornette *et al.*, 1987) y un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis).

Tabla 13.f.1. Similitud entre perfiles de hidropatía de las glicoproteínas G2 de arnavirus

	NUEVO MUNDO						VIEJO MUNDO				
	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
CD1	100 %	99,4 %	86,4 %	84,8 %	91,1 %	88,8 %	82,7 %	82,8 %	86,7 %	86,6 %	81,2 %
MC2		100 %	86,7 %	85,2 %	91,2 %	89,2 %	83,0 %	83,2 %	87,0 %	86,9 %	81,6 %
OLI			100 %	85,8 %	87,2 %	88,3 %	85,0 %	85,3 %	87,7 %	88,1 %	84,3 %
PIC				100 %	85,8 %	86,6 %	84,4 %	83,9 %	85,5 %	85,4 %	83,7 %
SAB					100 %	89,8 %	83,6 %	83,5 %	86,6 %	86,7 %	81,3 %
TAC						100 %	86,0 %	86,1 %	86,6 %	86,5 %	83,0 %
LAG							100 %	97,9 %	86,3 %	86,0 %	87,8 %
LAJ								100 %	86,3 %	86,0 %	85,1 %
LCA									100 %	97,9 %	85,6 %
LCE										100 %	85,1 %
MOP											100 %

Tabla 13.f.2. Distancia entre perfiles de hidropatía de las glicoproteínas G2 de arnavirus

	NUEVO MUNDO						VIEJO MUNDO				
	CD1	MC2	OLI	PIC	SAB	TAC	LAG	LAJ	LCA	LCE	MOP
CD1	0,0 %	0,6 %	13,6 %	15,2 %	8,9 %	11,3 %	17,3 %	17,2 %	13,3 %	13,4 %	18,8 %
MC2		0,0 %	13,3 %	14,8 %	8,8 %	10,8 %	17,0 %	16,8 %	13,0 %	13,1 %	18,4 %
OLI			0,0 %	14,2 %	12,9 %	11,8 %	15,0 %	14,7 %	12,4 %	11,9 %	15,7 %
PIC				0,0 %	14,2 %	13,4 %	15,6 %	16,1 %	14,5 %	14,7 %	16,3 %
SAB					0,0 %	10,2 %	16,4 %	16,5 %	13,4 %	13,3 %	18,7 %
TAC						0,0 %	14,0 %	13,9 %	13,4 %	13,6 %	17,0 %
LAG							0,0 %	2,1 %	13,7 %	14,0 %	12,2 %
LAJ								0,0 %	13,7 %	14,0 %	12,6 %
LCA									0,0 %	2,1 %	14,4 %
LCE										0,0 %	14,9 %
MOP											0,0 %

CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe;

LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia

En negrita y grisado se destacan las comparaciones entre cepas de un mismo arnavirus

En negrita y sombreado amarillo claro se destacan los límites inferior y superior, entre arnavirus distintos, de cada recuadro.

Nota: Los porcentajes de similitud de perfiles de hidropatía fueron calculados utilizando ventanas de 15 valores (ventanas parcialmente solapadas, desplazadas 1 valor). A continuación, se aplicó la transformada de Fourier a cada una de las ventanas y se realizó la comparación de manera análoga a las comparaciones de perfiles de LCC (ver Capítulo 18, Comparación cuantitativa de LCCs), con un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis).

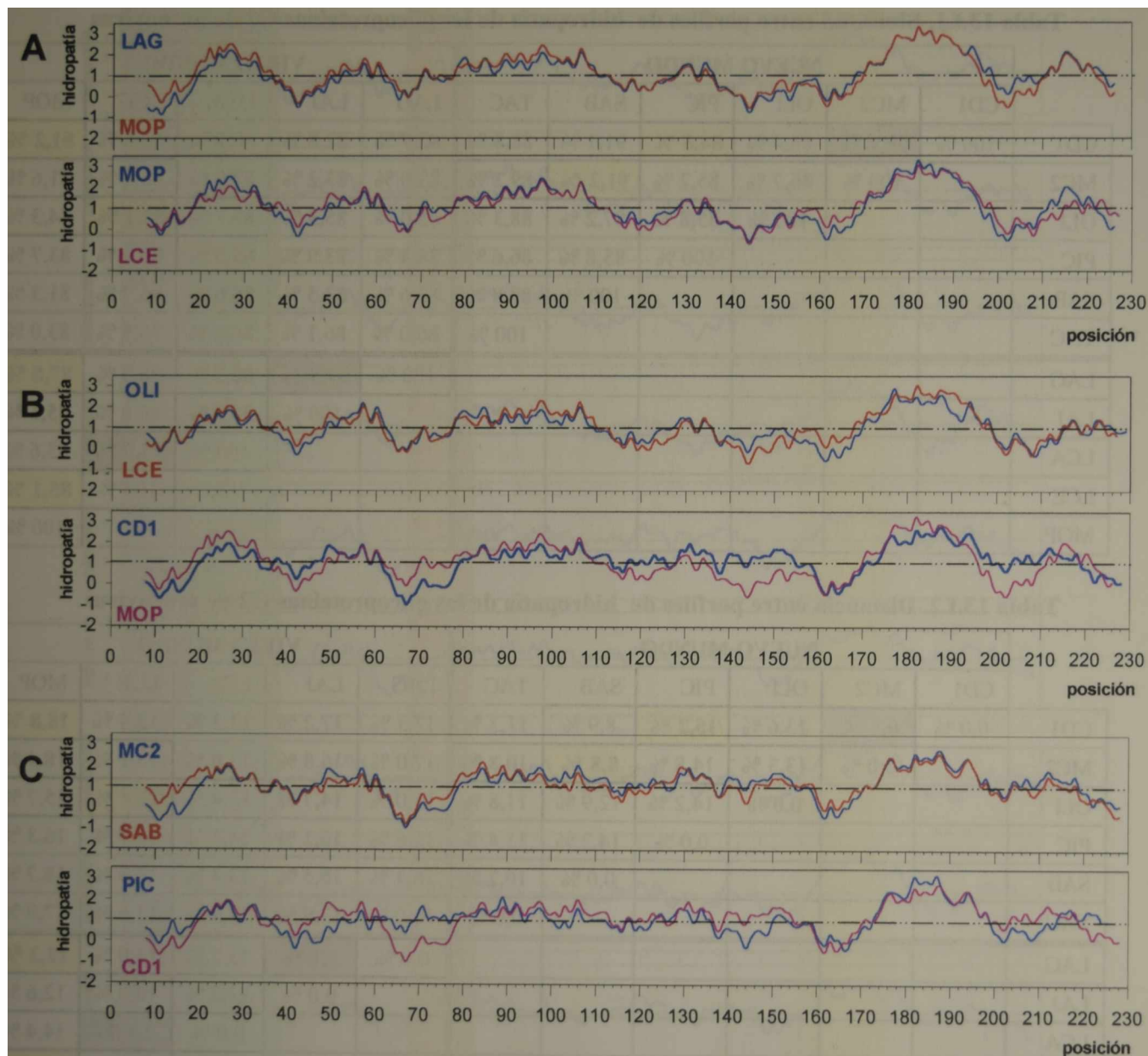


Figura 13.11.C. Comparación de perfiles de hidropatía de pares seleccionados de glicoproteínas G2 de arnavirus. Los perfiles fueron calculados utilizando ventanas sucesivas de 15 residuos. Con la finalidad de poder comparar entre sí los distintos gráficos, se introdujeron *gaps* en los perfiles respectivos. Las posiciones de los *gaps* derivan de un alineamiento múltiple (ver Apéndice 6, Figura Ap.6.2) y fueron ajustadas con un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis). La línea de guiones y puntos ubicada en hidropatía 1, sirve solamente como referencia; los picos por encima de la línea corresponden a regiones de mayor hidrofobicidad y los que están por debajo a regiones más hidrofílicas. **A.** Los perfiles superpuestos corresponden a los pares de mayor similitud (LAI/MOP; 87,8 %) y de menor similitud (MOP/LCE; 85,1 %) del grupo de los arnavirus del Viejo Mundo. **B.** Los perfiles superpuestos corresponden a los pares de mayor similitud (OLI/LCE; 88,1 %) y de menor similitud (CD1/MOP; 81,2 %) de la comparación entre los arnavirus del Viejo Mundo y los del Nuevo Mundo. **C.** Los perfiles superpuestos corresponden a los pares de mayor similitud (MC2/SAB; 91,2 %) y de menor similitud (PIC/CD1; 84,8 %) del grupo de los arnavirus del Nuevo Mundo (ver **Tabla 13.f.1**). LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE- LCM-WE; MOP: Mopeia; CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe.

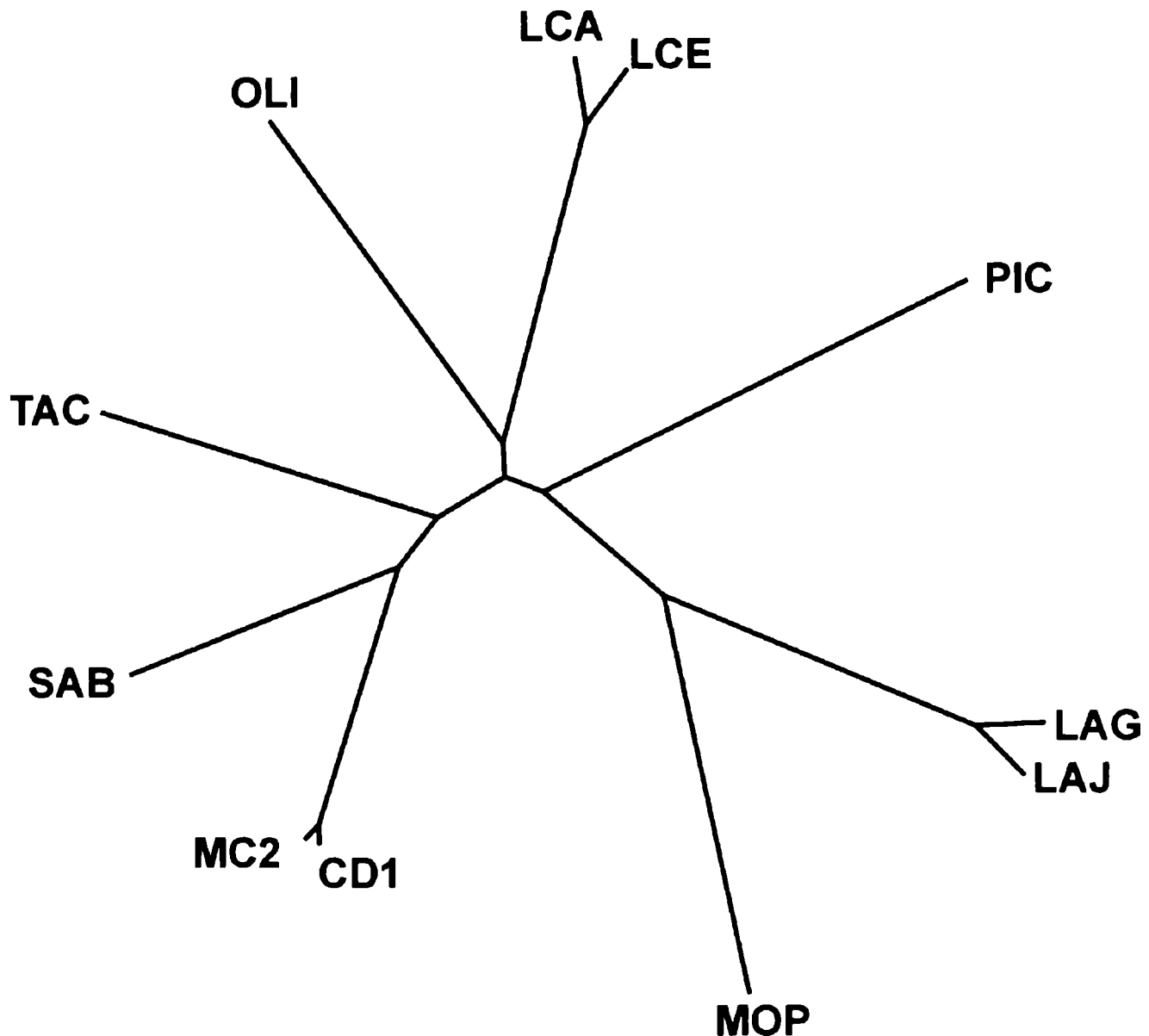


Figura 13.11.D. Cladograma correspondiente a relaciones estructurales entre las G2 de los arnavirus. El cladograma está basado en datos de similitud de los perfiles de hidropatía, y fue calculado utilizando el método UPGMA (*Unweighted Pair Group Arithmetic Average Clustering*). Como los datos de similitud representan la proximidad entre cada par de secuencias, la tabla de valores utilizada fue calculada asumiendo como distancias filogenéticas los valores resultantes del cálculo: $100 - v_n$ (donde v_n = valor individual) para cada par de secuencias (**Tabla 13.f.2**). Estas tablas de distancias fueron analizadas con la rutina NEIGHBOR (método UPGMA) y el cladograma fue dibujado con la rutina DRAWTREE, del paquete de programas PHYLIP (*Phylogeny Inferencing Package*, v 3.5, Felsenstein, J., 1995). Las longitudes de las distintas ramas en el cladograma están en relación con la distancia evolutiva. CD1: Junín-Candid #1; MC2: Junín-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia.

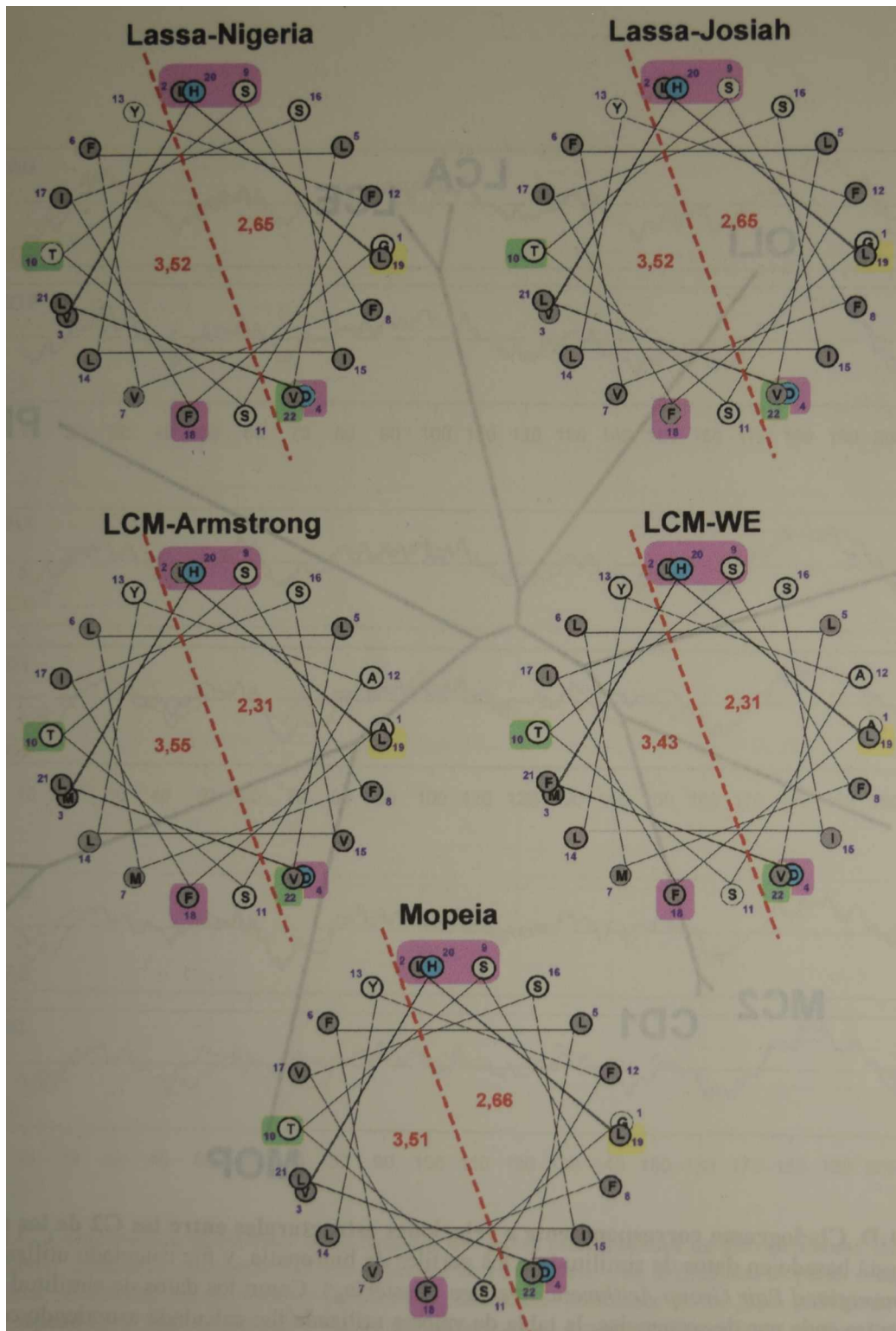


Figura 13.12. Representaciones tipo *helical wheel* de las α hélices transmembrana de los arnavirus del Viejo Mundo. Se muestran los diagramas individuales de cada virus. Los aminoácidos fueron numerados secuencialmente desde el extremo N-terminal de la estructura. Basado en Karlín *et al.* (1990), se hace referencia al diferente grado de hidrofobicidad del residuo con distinto tipo de fondo (fondo celeste: residuos hidrofílicos; fondo grisado: residuos hidrofóbicos; fondo blanco: los residuos ambivalentes). La línea de guiones roja corresponde a un plano hipotético que divide la α hélice en dos hemihélices. Los valores indicados en rojo corresponden a la hidrofobicidad media de la respectiva hemihélice. Como se puede observar, 20 de los 22 residuos que forman las hélices son hidrofóbicos o ambivalentes y, por lo tanto, las hélices son netamente hidrofóbicas. Sin embargo, la distribución de los residuos hidrofóbicos no es exactamente uniforme, y el plano de división en hemihélices permite identificar dos regiones de diferente hidrofobicidad media. Además, las dos hemihélices presentan una marcada asimetría en la conservación de residuos en todos los arnavirus: la hemihélice menos hidrofóbica presenta 6 residuos altamente conservados en tipo y posición, mientras que la hemihélice más hidrofóbica presenta sólo 2 residuos conservados. El grado de conservación se indica con fondos sombreados de distinto color (fondo rosa: 100%, fondo amarillo pálido: ~90%, fondo verde claro: ~80%). Por otra parte, algunos de los residuos tienen escasa probabilidad de estar presentes en la hemihélice que interactúa exclusivamente con los lípidos de membrana, como el aspartato (D). En función de estas consideraciones, se puede suponer que los residuos con mayor grado de conservación son cruciales para la formación de los tetrámeros. Para los cálculos de hidrofobicidad se utilizó la tabla PRIFT (Cornette *et al.*, 1987, 1995) y un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis).

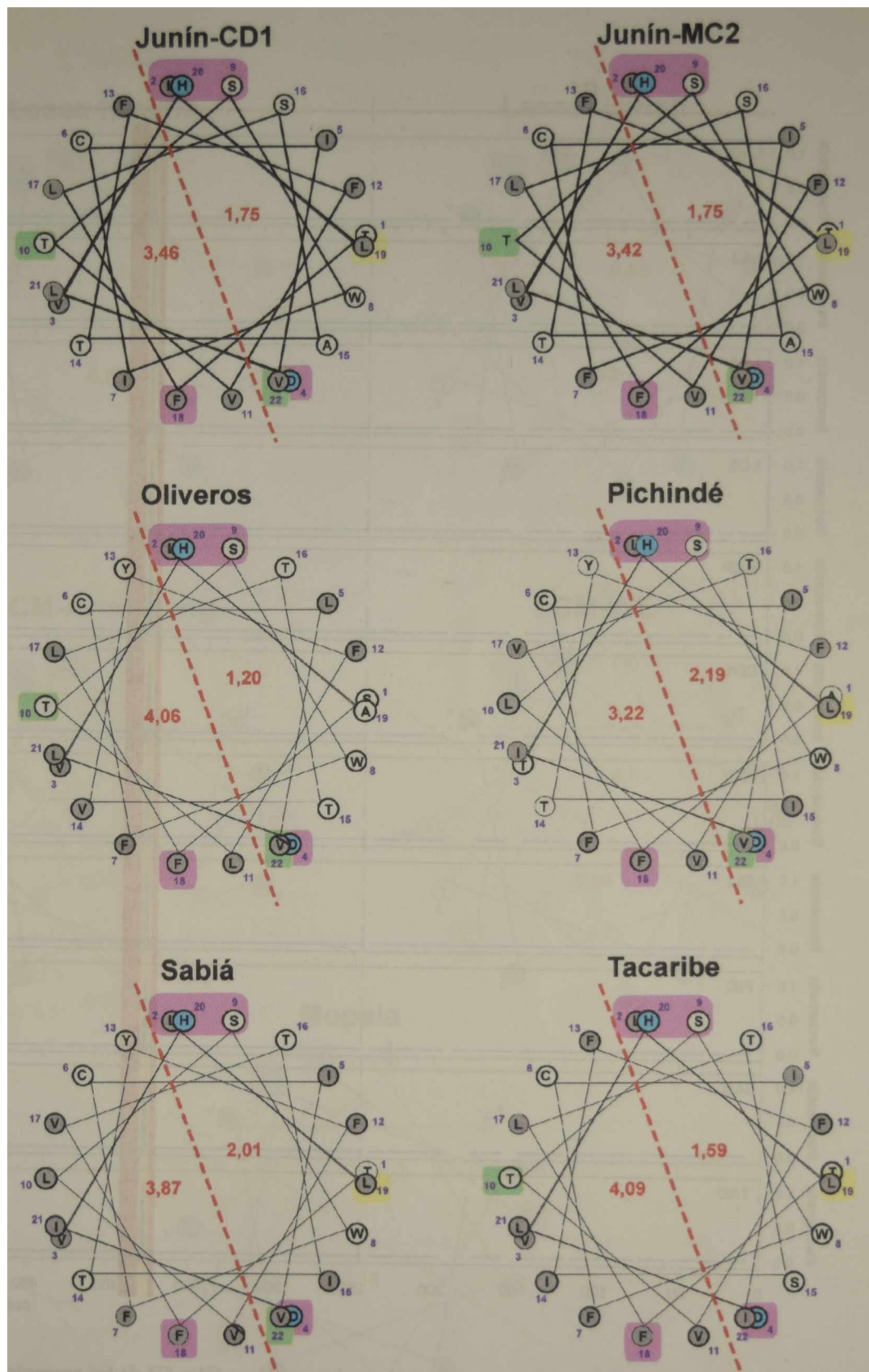


Figura 13.13. Representaciones tipo *helical wheel* de las α hélices transmembrana de los arnavirus del Nuevo Mundo. Se muestran los diagramas individuales de cada virus. Los aminoácidos fueron numerados secuencialmente desde el extremo N-terminal de la estructura. Basado en Karlin *et al.* (1990), se hace referencia al diferente grado de hidrofobicidad del residuo con distinto tipo de fondo (fondo celeste: residuos hidrofílicos; fondo grisado: residuos hidrofóbicos; fondo blanco: los residuos ambivalentes). La línea de guiones roja corresponde a un plano hipotético que divide la α hélice en dos hemihélices. Los valores indicados en rojo corresponden a la hidrofobicidad media de la respectiva hemihélice. Como se puede observar, 20 de los 22 residuos que forman las hélices son hidrofóbicos o ambivalentes y, por lo tanto, las hélices son netamente hidrofóbicas. Sin embargo, la distribución de los residuos hidrofóbicos no es exactamente uniforme, y el plano de división en hemihélices permite identificar dos regiones de diferente hidrofobicidad media. Además, las dos hemihélices presentan una marcada asimetría en la conservación de residuos en todos los arnavirus: la hemihélice menos hidrofóbica presenta 6 residuos altamente conservados en tipo y posición, mientras que la hemihélice más hidrofóbica presenta sólo 2 residuos conservados. El grado de conservación se indica con fondos sombreados de distinto color (fondo rosa: 100%, fondo amarillo pálido: ~90%, fondo verde claro: ~80%). Por otra parte, algunos de los residuos tienen escasa probabilidad de estar presentes en la hemihélice que interactúa exclusivamente con los lípidos de membrana, como el aspartato (D). En función de estas consideraciones, se puede suponer que los residuos con mayor grado de conservación son cruciales para la formación de los tetrámeros. Para los cálculos de hidrofobicidad se utilizó la tabla PRIFT (Cornette *et al.*, 1987, 1995) y un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis).

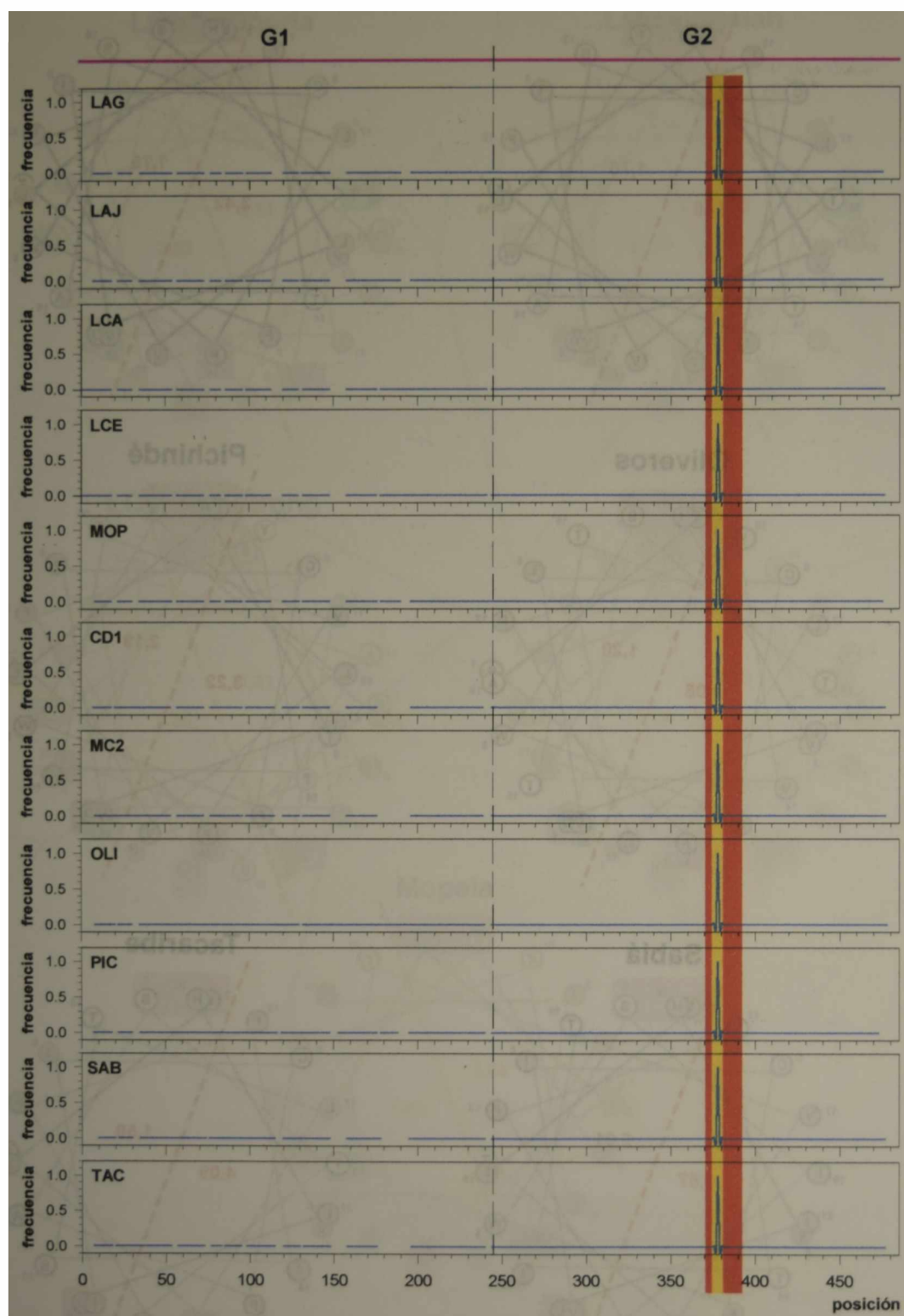


Figura 13. 14. Frecuencia de sitios potenciales de dimerización en las proteínas G1 y G2 de los arnavirus. Los sitios indicados corresponden a todos los nonapéptidos que coinciden con un consenso general definido en esta tesis: [CIKMNSTV]-[ADKQRST]-[AIGLSTV]-[AHIDGNQST]-[ILV]-[FILKMQTV]-[GKNRSTY]-[DHKNQRT]-[ANSTV]. El gráfico fue construido utilizando un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis), asignando un valor arbitrario de +1 a la posición central del nonapéptido coincidente con el consenso y 0 a la posición central de todos los nonapéptidos no coincidentes. Para correlacionar con un alineamiento múltiple (ver Apéndice 6, Figuras Ap.6.2 y Ap.6.3), se han introducido *gaps*. Con un sombreado anaranjado se indican los nonapéptidos identificados con el PROSITE, para algunos arnavirus (OLI y TAC), y extrapolados para el resto. Aún cuando este consenso es más amplio que el definido en el PROSITE, no se identifica ningún nonapéptido adicional. Con un sombreado rojo se indica la región correspondiente al tercer potencial dominio fusogénico de G2 (ver Apéndice 6, Figuras Ap.6.2 y Ap.6.3). LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE- LCM-WE; MOP: Mopeia; CD1: Junin-Candid #1; MC2: Junin-MC2; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe.

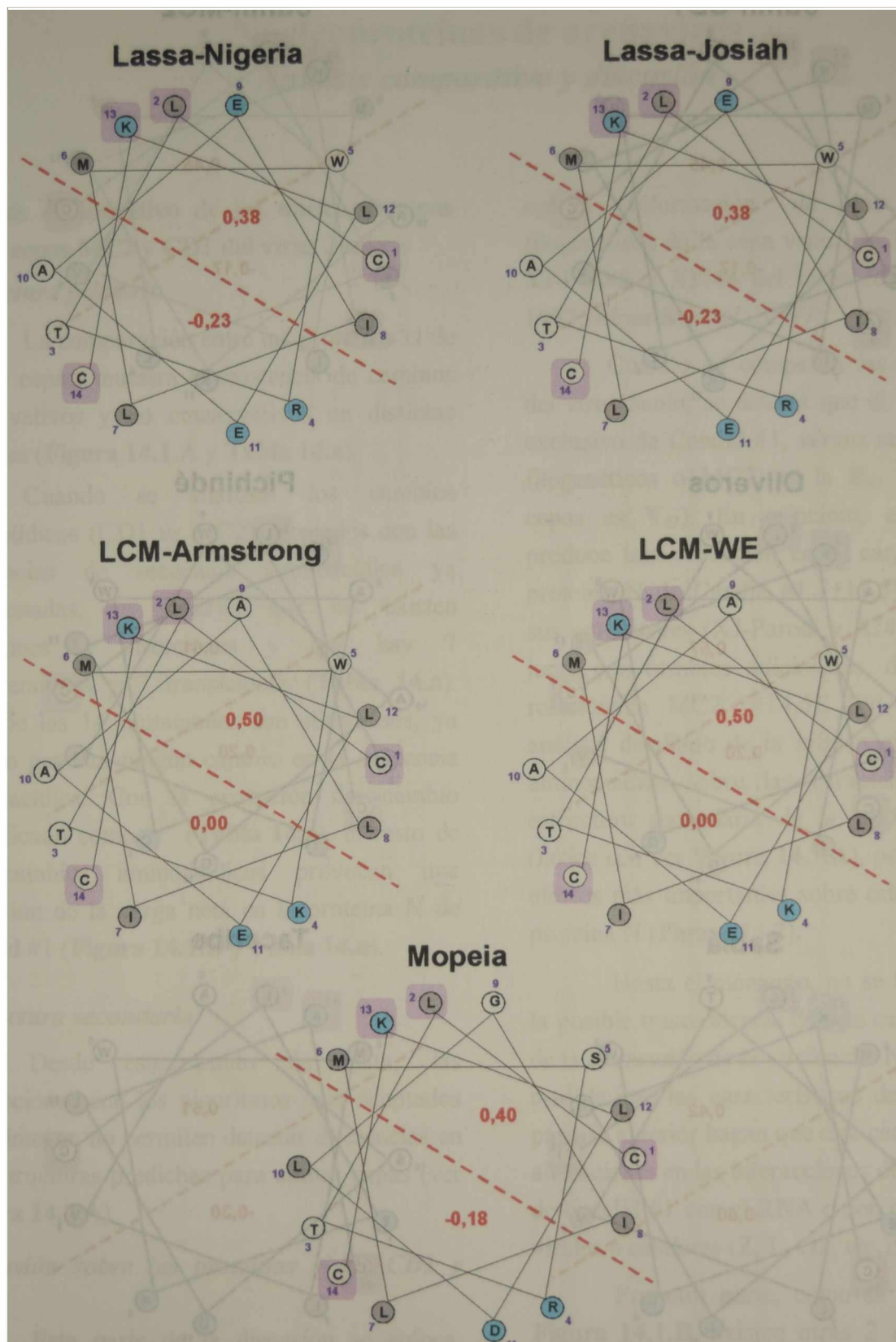


Figura 13.15. Representaciones tipo *helical wheel* de las α hélices $\alpha 1$ de los arnavirus del Viejo Mundo. Se muestran los diagramas individuales de cada virus. Los aminoácidos fueron numerados secuencialmente desde el extremo N-terminal de la estructura. Basado en Karlin *et al.* (1990), se hace referencia al diferente grado de hidrofobicidad del residuo con distinto tipo de fondo (fondo celeste: residuos hidrofílicos; fondo grisado: residuos hidrofóbicos; fondo blanco: residuos ambivalentes). La línea de guiones roja corresponde a un plano hipotético que divide la α hélice en dos hemihélices. Los valores indicados en rojo corresponden a la hidrofobicidad media de la respectiva hemihélice. Como se puede observar las hélices tienen una predominancia de aminoácidos hidrofóbicos o ambivalentes. Sin embargo, la distribución de los residuos no es exactamente uniforme, y el plano de división en hemihélices permite identificar dos regiones de diferente hidrofobicidad media. Además, las dos hemihélices presentan una marcada asimetría en la conservación de residuos en todos los arnavirus: la hemihélice más hidrofóbica presenta 3 residuos altamente conservados en tipo y posición, mientras que la hemihélice menos hidrofóbica presenta sólo 1 residuo conservado (señalado con sombreado de color rosa). Para comparar los valores de hidrofobicidad con la bibliografía, en este gráfico se utilizó la tabla de Eisenberg normalizada (Cornette *et al.*, 1987, 1995) y un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis).

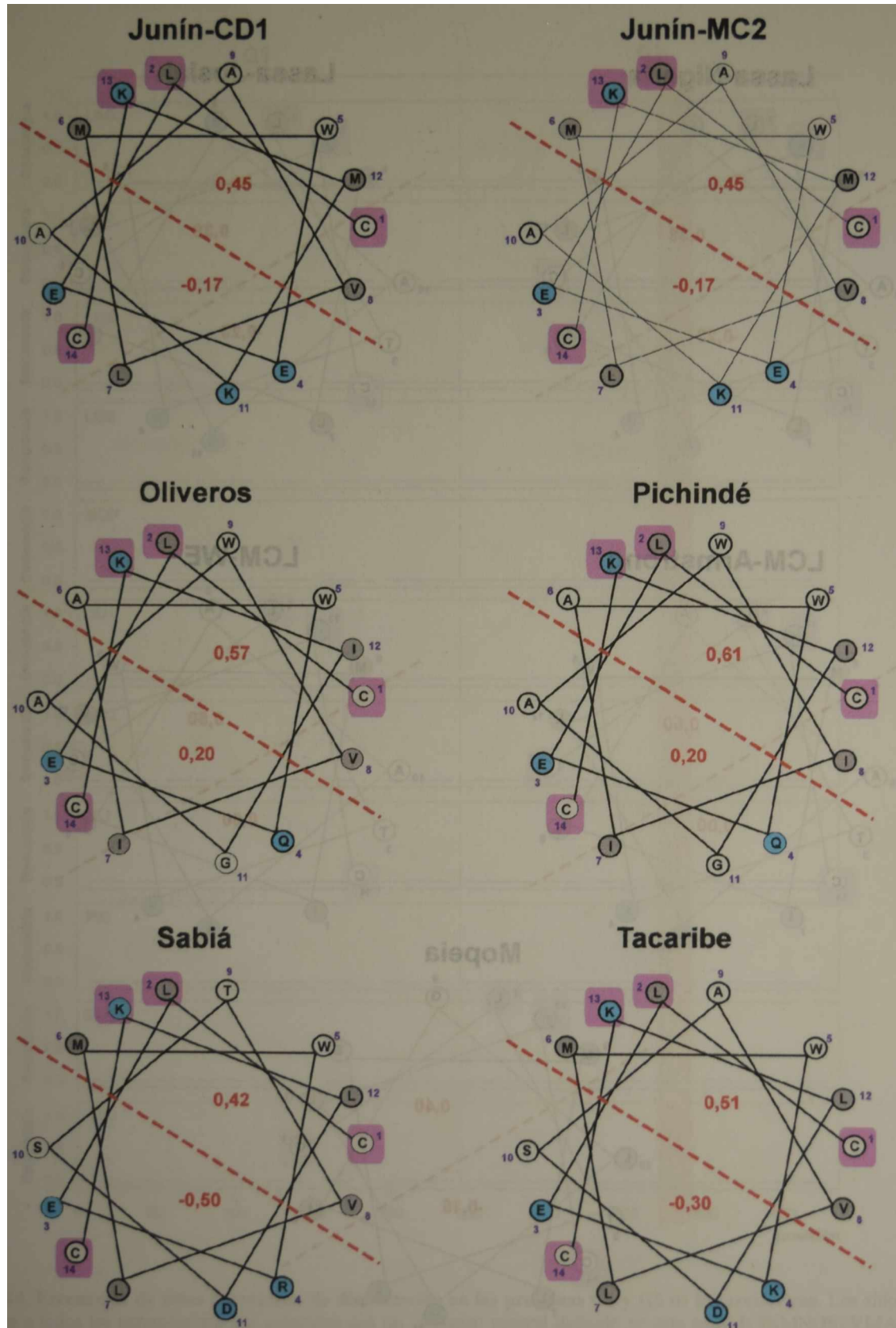


Figura 13.16. Representaciones tipo *helical wheel* de las α hélices $\alpha 1$ de los arnavirus del Nuevo Mundo. Se muestran los diagramas individuales de cada virus. Los aminoácidos fueron numerados secuencialmente desde el extremo N-terminal de la estructura. Basado en Karlín *et al.* (1990), se hace referencia al diferente grado de hidrofobicidad del residuo con distinto tipo de fondo (fondo celeste: residuos hidrofílicos; fondo grisado: residuos hidrofóbicos; fondo blanco: residuos ambivalentes). La línea de guiones roja corresponde a un plano hipotético que divide la α hélice en dos hemihélices. Los valores indicados en rojo corresponden a la hidrofobicidad media de la respectiva hemihélice. Como se puede observar las hélices tienen una predominancia de aminoácidos hidrofóbicos o ambivalentes. Sin embargo, la distribución de los residuos no es exactamente uniforme, y el plano de división en hemihélices permite identificar dos regiones de diferente hidrofobicidad media. Además, las dos hemihélices presentan una marcada asimetría en la conservación de residuos en todos los arnavirus: la hemihélice más hidrofóbica presenta 3 residuos altamente conservados en tipo y posición, mientras que la hemihélice menos hidrofóbica presenta sólo 1 residuo conservado (señalado con sombreado de color rosa). Para comparar los valores de hidrofobicidad con la bibliografía, en este gráfico se utilizó la tabla de Eisenberg normalizada (Cornette *et al.*, 1987, 1995) y un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis).

Capítulo 14

Nucleoproteínas de arnavirus

Análisis comparativo y discusión

Análisis comparativo de las nucleoproteínas de las cepas MC2 y CD1 del virus Junín

Estructura primaria

La comparación entre las proteínas N de ambas cepas, muestra la existencia de cambios conservativos y no conservativos en distintas regiones (**Figura 14.1.A** y **Tabla 14.a**).

Cuando se analizan los cambios nucleotídicos (CD1 vs MC2) asociados con las diferencias de secuencia aminoacídica ya mencionadas, se observa que no existen deleciones o inserciones y que hay 7 transversiones y 7 transiciones (**Tabla 14.a**). Tres de las 14 mutaciones son silenciosas, ya que no generan ningún cambio en la secuencia aminoacídica. Con la excepción del cambio identificado como *f* (**Tabla 14.a**), el resto de los cambios aminoacídicos provocan una variación de la carga neta en la proteína N de Candid #1 (**Figura 14.1.B** y **Tabla 14.a**).

Estructura secundaria

Desde este punto de vista, las predicciones con los algoritmos más aceptados actualmente, no permiten detectar diferencias en las estructuras predichas para ambas cepas (ver **Figura 14.7.A**).

Discusión sobre las proteínas N de CD1 y MC2

Esta parte de la discusión se enfoca, principalmente, sobre la información de las estructuras primarias y secundarias de las proteínas N de la cepa salvaje MC2 (Ghiringhelli *et al.*, 1991) y la cepa vacunal CD1 (Ghiringhelli *et al.*, 1997) del virus Junín. Para una mejor comprensión de la posible importancia de algunos cambios observados, se

agrega información de los antecesores filogenéticos de la cepa vacunal CD1, las cepas XJ-Parodi y XJ#44 del virus Junín (Albariño, 1997; Albariño *et al.*, 1997).

Cuando se comparan las cuatro cepas del virus Junín, se detecta que el único cambio exclusivo de Candid #1, *versus* sus antecesores filogenéticos o MC2, es la E₄₇ (en las otras cepas es V₄₇). En principio, esta mutación produce una reducción en la carga neta de la proteína N de Candid #1 (+11,5), respecto de sus antecesores (XJ-Parodi y XJ#44, +12,5) o, junto con cambios adicionales, de la cepa no relacionada MC2 (+14,5). Sin embargo, un análisis detallado de la influencia del cambio aminoacídico sobre las características de la estructura secundaria de la cual forma parte (hélice α_3 , ver **Figura 14.7.A**), permite detectar efectos más importantes sobre esta región de la proteína N (**Figura 14.2**).

Hasta el momento, no se ha establecido la posible trascendencia de este cambio respecto de la atenuación de la virulencia. No obstante, es posible que las características descritas en el párrafo anterior hagan que este cambio derive en alteraciones en las interacciones de la proteína N de Candid #1 con el RNA o con otras proteínas virales o celulares (Z, L, G2, etc.).

Por otra parte, como se observa en la **Figura 14.1.B**, existen otras 2 regiones en la proteína N de Candid #1 que presentan diferencias notorias en la carga neta local *versus* la cepa MC2. Sin embargo, dado que en los antecesores filogenéticos XJ-Parodi y XJ#44 existe el mismo perfil local que en Candid #1, estas regiones deberían considerarse no relacionadas con la atenuación.

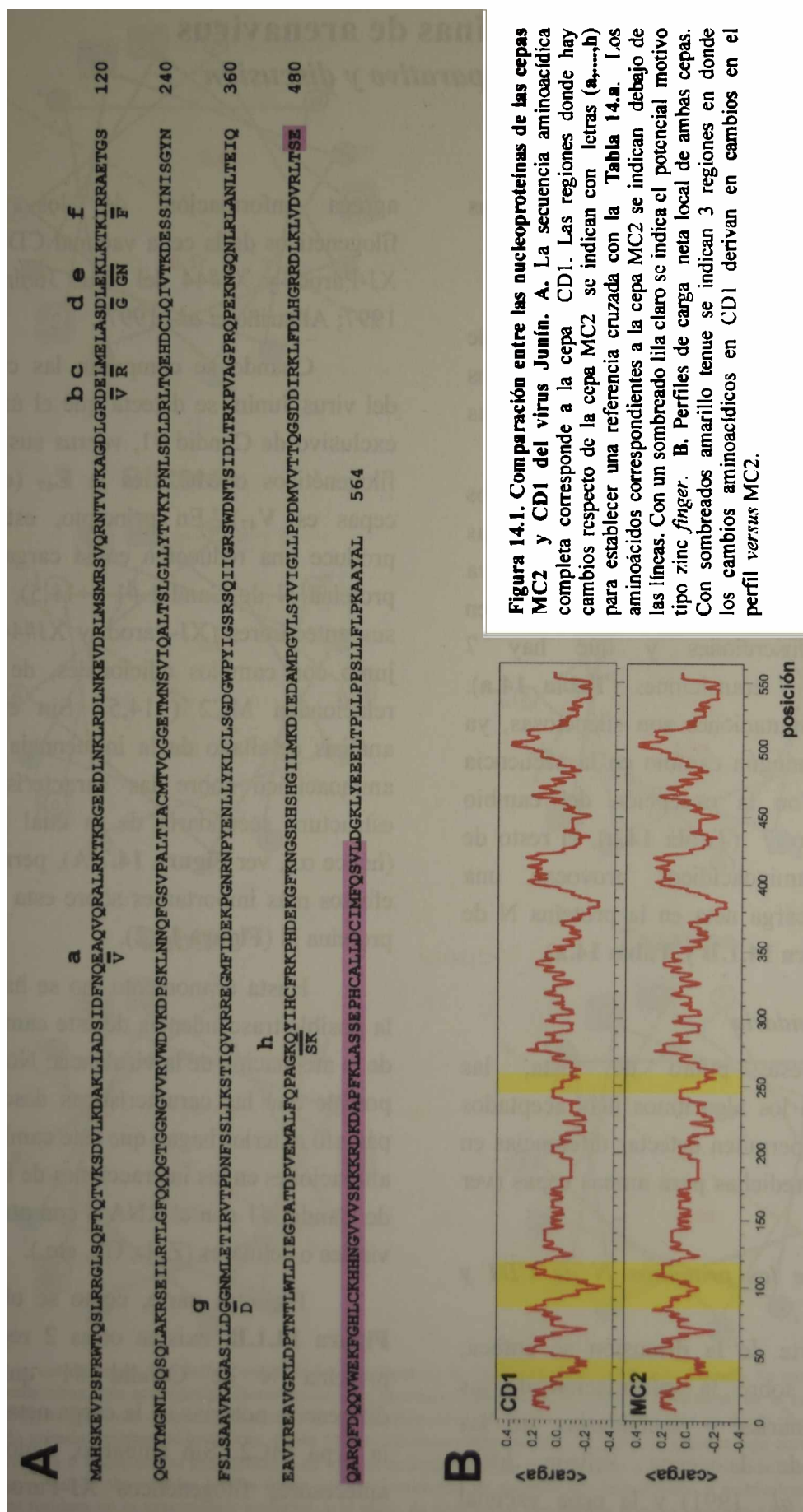


Figura 14.1. Comparación entre las nucleoproteínas de las cepas MC2 y CD1 del virus Junín. A. La secuencia aminoacídica completa corresponde a la cepa CD1. Las regiones donde hay cambios respecto de la cepa MC2 se indican con letras (a,....h) para establecer una referencia cruzada con la Tabla 14.a. Los aminoácidos correspondientes a la cepa MC2 se indican debajo de las líneas. Con un sombreado ilia claro se indica el potencial motivo tipo zinc finger. B. Perfiles de carga neta local de ambas cepas. Con sombreados amarillo tenue se indican 3 regiones en donde los cambios aminoacídicos en CD1 derivan en cambios en el perfil versus MC2.

Tabla 14.a. Cambios en las secuencias de las nucleoproteínas de las cepas MC2 y CD1 del virus Junín

Posición	Nucleótido		Codón		Aminoácido	
	MC2	CD1	MC2	CD1	MC2	CD1
221	U	A	GUG	GAG	V	E ^a
342	G	A	ACG	ACA	T	T
371	U	A	GUU	GAU	V	D ^b
377	G	U	CGG	CUG	R	L ^c
395	G	A	GGC	GAC	G	D ^d
401	G	A	GGG	GAG	G	E ^e
405	U	A	AAU	AAA	N	K ^e
418	U	A	UUU	AUU	F	I ^f
471	G	A	CAG	CAA	Q	Q
848	A	G	GAU	GGU	D	G ^g
1277 / 1278	GC	A	AGC	AAG	S	K ^h
1279	A	C	AAG	CAG	K	Q ^h
1536	C	U	UUC	UUU	F	F
Carga neta a pH 7,0					+ 14,5	+ 11,5

a, b, c, d, e, f, g, h letras que identifican regiones de cambios (para facilitar la referencia cruzada con la Figura 14.1.A).

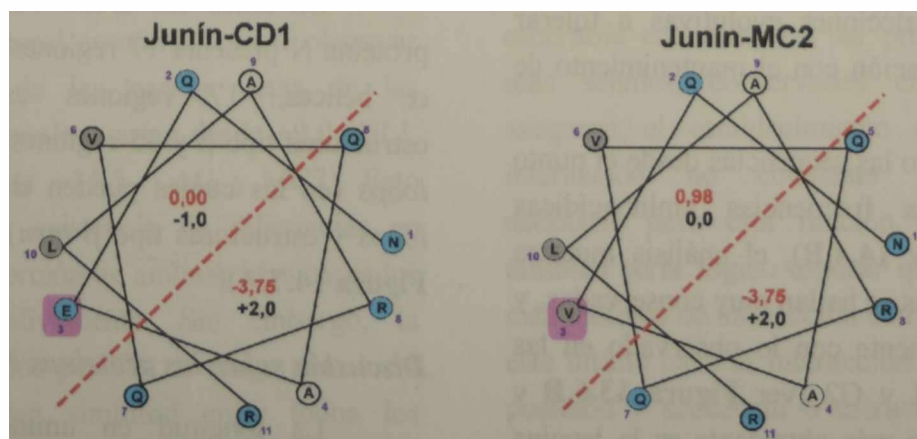


Figura 14.2. Representaciones tipo *helical wheel* de las α hélices $\alpha 3$ de las cepas Candid #1 y MC2 del virus Junín. Se muestran los diagramas individuales de cada cepa. Los aminoácidos fueron numerados secuencialmente desde el extremo N-terminal de la estructura. Basado en Karlin *et al.* (1990), se hace referencia al diferente grado de hidrofobicidad del residuo con distinto tipo de fondo (fondo celeste: residuos hidrofílicos; fondo grisado: residuos hidrofóbicos; fondo blanco: residuos ambivalentes). La línea de guiones roja corresponde a un plano hipotético que divide la α hélice en dos hemihélices. Los valores indicados en rojo corresponden a la hidrofobicidad media y los valores indicados en negro corresponden a la carga neta local de las respectivas hemihélices. La hemihélice más hidrofóbica es la que se ve afectada por el cambio $V_{47} \rightarrow E_{47}$ presente en CD1 (señalado con sombreado de color rosa). Para el cálculo de hidrofobicidades, se utilizó la tabla PRIFT (Cornette *et al.*, 1987, 1995) y un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, *est. Fesis*). Para el cálculo de carga neta, se asignó +1 a la R y -1 a la E.

Análisis comparativo de las proteínas N de los arnavirus

Estructura primaria

La comparación entre las proteínas N de los arnavirus, permite observar la existencia de diferencias significativas en cuanto al grado de conservación aminoacídica en distintas regiones peptídicas (**Figura 14.3**).

Un análisis más detallado, basado en alineamientos múltiples (ver **Apéndice 6**), de la similitud aminoacídica, muestra que:

- En las proteínas N se observa que la identidad varía entre 47,6 % (MC2-LAJ) y 52,3 % (SAB-LCA) y la homología entre 71,2 % (MC2-LAG) y 75,9 % (SAB-LCA), cuando se tienen en cuenta los arnavirus del Nuevo Mundo vs. los del Viejo Mundo. Por otra parte, cuando se consideran por separado los dos grandes grupos de arnavirus, en los del Nuevo Mundo la identidad varía entre 54,4 % (MC2-PIC ó MAC-PIC) y 86,9 % (CD1-MAC) y la homología entre 78,5 % (MC2-PIC) y 96,1 % (CD1-MAC); mientras que en los del Viejo Mundo la identidad varía entre 60,6 % (LAG-LCA) y 73,4 % (LAG -MOP) y la homología entre 80,6 % (LAG-LCA) y 91,1 % (LAJ-MOP) (**Tablas 14.b.1 y 14.b.2**).

En síntesis, el grado de conservación global de las secuencias aminoacídicas en las proteínas N es sorprendentemente alto (*versus* lo esperado por azar); lo cual, probablemente, sea el reflejo de restricciones evolutivas a tolerar cambios en asociación con el mantenimiento de sus funciones.

Estudiando las secuencias desde el punto de vista de las frecuencias aminoacídicas (**Figura 14.4.A y 14.4.B**), el análisis muestra que las frecuencias se hallan muy conservadas, y que, coincidentemente con lo observado en las glicoproteínas G1 y G2 (ver **Figura 13.6.B y C**), el aminoácido más abundante es la leucina (L). Por otra parte, cuando se agrupan los aminoácidos en hidrofóbicos, hidrofílicos y ambivalentes (Karlin *et al.*, 1990), las

dispersiones también son pequeñas. En particular, se observa una alta predominancia de aminoácidos hidrofílicos y ambivalentes y una abundancia significativamente menor de aminoácidos hidrofóbicos.

Una característica bastante llamativa, probablemente asociada con la organización estructural de las proteínas N, es la baja frecuencia de algunos aminoácidos, en particular las cisteínas y las histidinas. Cuando se analiza la distribución de estos dos residuos aminoacídicos, se puede observar que la distribución no es aleatoria (**Figura 14.5.A, B y C**). Aproximadamente desde el residuo 400 hasta el final de las proteínas N (tercio C-terminal), se encuentra la región con mayor concentración de C y H. Entre los residuos 190 a 320 de todas las proteínas N existe otra región con una concentración local de C y H relativamente alta, tres de estos residuos (2 C y 1 H) están totalmente conservados en posición cuando se analiza un alineamiento múltiple. Por último, la mayor parte de los arnavirus presentan 1 ó 2 de estos residuos distribuidos en los 100 primeros aminoácidos.

Estructura secundaria

La predicción de estructuras secundarias y el análisis de las mismas, muestran que existe un grado excepcionalmente alto de conservación estructural.

En todos los arnavirus analizados, la proteína N presenta 17 regiones predichas como α hélices, 17 regiones estimadas como estructuras tipo β y 35 regiones predichas como *loops* (en las cuales pueden existir verdaderos *loops* y estructuras tipo β *turn*) (**Figura 14.6 y Figura 14.7.A**).

Discusión sobre las proteínas N de arnavirus

La longitud en aminoácidos de las proteínas N varía desde 558 residuos en las cepas de LCM y en Oliveros hasta 570 residuos en Lassa-Nigeria, en Mopeia y en Tacaribe, lo

cual es coincidente con la gran conservación de secuencias ya mencionada (ver **Tablas 14.b**).

Desde un punto de vista estructural, las proteínas N de los diferentes arnavirus tienen, en general, una distribución aproximadamente regular de las tres categorías de aminoácidos: hidrofóbicos, hidrofílicos y ambivalentes, lo cual redundaría en una mayor probabilidad promedio de exposición al solvente y sugiere que, en todos los casos, podría ser una proteína soluble (análisis general no mostrado). Esta información es coincidente con la predicción computacional realizada para las proteínas N de las cepas del virus Junín (ver **Capítulos 7 y 11**), que sugiere que las Ns son proteínas con una conformación de tipo globular.

Un análisis detallado de los perfiles de hidropatía de las nucleoproteínas (**Figura 14.7.B**), permite observar que los mismos son muy parecidos. La similitud de perfiles de hidropatía varía entre 76,3 % (OLI-MOP) y 82,9 % (PIC-LCA), cuando se tienen en cuenta los arnavirus del Nuevo Mundo vs. los del Viejo Mundo. Por otra parte, cuando se consideran por separado los dos grandes grupos de arnavirus, en los del Nuevo Mundo la similitud varía entre 80,0 % (PIC-TAC) y 92,7 % (CD1-MAC); mientras que en los del Viejo Mundo la similitud varía entre 79,3 % (LAJ-LCE) y 87,9 % (LAJ-MOP) (**Tabla 14.c.1, Figuras 14.7.C y 14.7.D**).

Una prueba adicional de la existencia de un patrón estructural común se puede observar en un análisis de las cargas netas de las proteínas N, las cuales varían desde +9,0 (OLI, PIC y TAC) hasta +14,5 (LAG y MC2). Esto último indica la existencia de diferencias en cuanto a la proporción de aminoácidos cargados positiva o negativamente. Sin embargo, la comparación de los perfiles de carga neta local muestra una gran similitud entre todos los arnavirus analizados (**Figura 14.8**).

El conjunto de características mencionado en los párrafos anteriores es

coincidente con la existencia de un patrón común observada en las predicciones de estructura secundaria, tanto respecto del tipo de estructura predicha como respecto de algunas de las características fisicoquímicas (AI: índice de anfipaticidad, hidrofobicidad media local, etc.) (ver **Figura 14.7.A**).

En síntesis, en cualquier tipo de comparación que se realice, las similitudes entre las nucleoproteínas de todos los arnavirus son mayores que las diferencias.

El patrón estructural uniforme para todas las proteínas N, probablemente, está en directa relación con dos de las funciones biológicas descriptas para la misma:

1. El rol estructural en la constitución de las nucleocápsides virales.
2. El rol regulatorio en el proceso de transcripción/replicación (función de antiterminadora; Tortorici *et al.*, 2001b).

Respecto del primer punto, es importante recordar que, en las micrografías electrónicas de viriones desarmados mediante shock osmótico, las nucleocápsides se observan como estructuras circulares, formadas por hebras contorsionadas de unos 12-15 nm de diámetro, en las cuales es posible distinguir subunidades globulares de 4-5 nm (Young and Howard, 1983). Además, el RNA presente en las nucleocápsides está protegido del ataque por RNasas. En base a estas consideraciones, es esperable encontrar en las proteínas N una o más señales, conservadas en posición, que aseguren el establecimiento adecuado de la interacción no covalente proteína-proteína necesaria para esta función. Por otra parte, también sería lógico esperar que presenten 1 o más motivos de interacción con RNA; aunque en este último caso la interacción independiente de posición y secuencia o estructura nucleotídica (necesaria para formar las nucleocápsides) puede estar asociada simplemente con la distribución de cargas netas locales.

Respecto del segundo punto, se pueden considerar varias alternativas:

- ⇒ que las proteínas N presenten una región de secuencia y/o estructura que le permita interactuar específicamente con las estructuras tipo *hairpin loop* intergénicas, independientemente de si interactúan con otras proteínas (L, Z, o proteínas de la célula).
- ⇒ que las proteínas N presenten una región de secuencia y/o estructura que le permita interactuar específicamente con las proteínas del complejo replicativo, y su presencia y/o actividad permita desarmar las estructuras tipo *hairpin loop* intergénicas, facilitando la replicación.
- ⇒ que existan ambos tipos de motivos y ambos sean necesarios para la actividad regulatoria.

En cualquiera de las alternativas subyace una pregunta adicional. Existe alguna señal (interacción con otra molécula o una modificación postraducciona) asociada con el cambio de rol de las proteínas N, desde el papel estructural al papel regulatorio? Desde este punto de vista, en distintos sistemas virales se ha descrito la importancia de la existencia de ciclos de fosforilación/defosforilación, por parte de quinasas virales y/o celulares, asociados con la regulación de los procesos de transcripción y replicación viral. Así es como, en el virus herpes simplex, la presencia de inhibidores de las tirosinquinasa disminuye significativamente la replicación (Yura *et al.*, 1997); en el virus de la hepatitis delta, la presencia en forma individual de inhibidores de caseinquinasa II o de PKC interfiere significativamente con la replicación pero no con los procesos transcripcionales y el ensamblaje de cápsides vacías (Yeh *et al.*, 1996); en el virus Sendai la proteína P se encuentra fosforilada aparentemente por la PKC- ζ (Huntley *et al.*, 1997); en el virus de la rabia la proteína N es fosforilada por la caseinquinasa II (Anzai *et al.*, 1997); la proteína

N del virus de sarampión (Measles) presenta dos formas alternativas de fosforilación (Gombart *et al.*, 1995; Segev *et al.*, 1995) y la proteína N del virus Marburg posee múltiples formas de fosforilación por diversas quinasas celulares (Löftering *et al.*, 1999). Por otra parte, en diversos sistemas virales se ha demostrado la existencia de distintos estados de fosforilación que puede tener relación con el ejercicio de diferentes funciones. Así es como en el virus de la estomatitis vesicular (VSV), los ciclos de fosforilación/defosforilación de la proteína P están íntimamente asociados con la regulación de la transcripción (Barik and Banerjee, 1992).

Por último, es bien conocida la participación de las tirosinquinasa en la regulación del desarrollo (Blaschke *et al.*, 1991).

En el caso de los arnavirus, un análisis exhaustivo de las proteínas de N con la base de datos PROSITE predice la existencia de diferentes potenciales consensos: varios sitios de N-glicosilación; múltiples sitios de fosforilación por PKC y caseinquinasa II; un sitio de fosforilación por tirosinquinasa; en algunos de los arnavirus, un sitio de fosforilación por proteinquinasa dependientes de cAMP o cGMP; varios sitios de miristoilación; en algunos casos un sitio de amidación; y, en varios de los arnavirus, al menos una señal de dimerización.

En los párrafos subsiguientes se presenta un ensayo analítico de la posible funcionalidad de algunos de los sitios consenso hallados, y se presenta la existencia de sitios no detectados con la base de datos PROSITE.

Respecto a los sitios de N-glicosilación, la cantidad y posición en las proteínas N es sumamente variable. No obstante, nunca se ha podido demostrar experimentalmente que las proteínas N estén glicosiladas, por lo cual estos sitios deberían descartarse.

Respecto de los sitios de fosforilación, es importante mencionar que existen algunas publicaciones que indican la presencia de N

fosforilada en LCM (Howard and Buchmeier, 1983; Bruns *et al.*, 1983b, 1986). En particular, en la publicación de Howard & Buchmeier (1983), se detecta la presencia de treoninas (T) y serinas (S) fosforiladas. La responsabilidad de la actividad de quinasa ha sido atribuida a una proteína viral (probablemente la L) (Howard and Buchmeier, 1983; Bruns *et al.*, 1986, 1990). Sin embargo, no se puede descartar la presencia de fosforilaciones adicionales derivadas de la actividad de quinasas celulares.

La mayor parte de los sitios de fosforilación encontrados, corresponden a sitios con alta frecuencia de encuentro por azar, ya que están formados por unos pocos aminoácidos (3 ó 4). En la **Figura 14.9.A** se puede apreciar la abundancia y distribución de los sitios de fosforilación por proteinquinasa C y por caseinquinasa II. Sólo un sitio consenso para cada quinasa se encuentra conservado en posición en todos los arnavirus estudiados.

Por otra parte, uno de los sitios potenciales de fosforilación merece especial atención. El sitio de fosforilación por tirosinquinasa es único en todas las proteínas N analizadas, y se encuentra exactamente conservado en posición (**Figura 14.9.A**). Los consensos de fosforilación por tirosinquinasa descritos en el PROSITE corresponden a dos octapéptidos alternativos: [KR]-x(2)-[DE]-x(3)-Y o [KR]-x(3)-[DE]-x(2)-Y.

En este punto conviene realizar algunas consideraciones estadísticas. La población total de octapéptidos posibles es de $2,56 \times 10^{10}$ miembros, mientras que la fracción de esa población que comprende los dos consensos es de $2,56 \times 10^7$ y la probabilidad de encontrar uno de ellos es 1×10^{-3} (frecuencias iguales para los distintos aminoácidos). Si esta probabilidad se corrige con las frecuencias de los distintos aminoácidos, existentes en N, la probabilidad disminuye a $6,24 \times 10^{-4}$, debido a la baja frecuencia de Y (su frecuencia es 0,02 y lo esperable por azar es 0,05). Esto último implica

que, para una proteína de ~585 aminoácidos (promedio de los alineamientos múltiples de N), la frecuencia esperada de hallazgo de octapéptidos sería de 0,36 sitios/proteína.

En función del análisis realizado en los párrafos anteriores se puede hipotetizar que el cambio de rol de las proteínas N (de un papel estructural a un papel regulatorio) puede estar asociado con la presencia o ausencia de fosforilaciones en una o más posiciones, por la acción de quinasas celulares y/o virales.

Además, dado que se ha publicado una posible relación entre la miristoilación en sitios internos de la proteína y el patrón de fosforilación por PKC en sistemas eucarióticos (Utsumi *et al.*, 1994), no se puede dejar de considerar la existencia de múltiples sitios potenciales de adición de residuos miristato (**Figura 14.9.A**), sólo dos de ellos se encuentran conservados en posición en todos los arnavirus analizados.

En última instancia, la existencia real de alguna (o todas) estas modificaciones postraduccionales debería ser demostrada experimentalmente.

Por otra parte, sobre los sitios consenso de amidación, se debe tener en cuenta que no existen datos experimentales que soporten su empleo durante el ciclo viral y, además, no están presentes en todas las proteínas N. Desde este punto de vista, es factible considerar que la presencia de este sitio en algunos de los arnavirus se debe exclusivamente al azar.

Respecto de las señales de dimerización, es importante considerar que pueden ser las señales que aseguren la interacción proteína N-proteína N, o proteína N-proteína viral o celular.

Los nonapéptidos que podrían actuar como dominios de dimerización fueron detectados en 7 de las 12 proteínas N analizadas: en Oliveros (₂₇TKGDVLRDA₃₅ y ₁₉₈TTLGLLYTV₂₀₆), en Lassa- Nigeria (₃₂₂SRTSIVGRA₃₃₀), en Mopeia (₃₂₂SRTSIVGRA₃₃₀), en Junín-Candid #1 (₃₁₇SRSQIIGRS₃₂₅), en

Junín-MC2 (³¹⁷SRSQIIGRS₃₂₅), en Machupo (³¹⁷SRSQILGRS₃₂₅) y en Tacaribe (³¹⁷SRSQIMGRS₃₂₅) (ver Apéndice 6, Figura Ap6.4).

Dada la escasa conservación de secuencia entre los ocho motivos hallados en las N (ver también Capítulo 13, sitios hallados en las glicoproteínas G2), se definió un consenso ampliado que comprende a todos los nonapéptidos de este tipo hallados en las proteínas N de arnavirus: [CIKMNQSTV]-[ADKQRST]-[AIGLSTV]-[AHIDGNQST]-[ILV]-[FILKMQTV]-[GKNRSTY]-[DHKNQRT]-[ANSTV].

En este punto conviene realizar algunas consideraciones estadísticas. La población total de nonapéptidos posibles es de $5,1 \times 10^{11}$ miembros, mientras que la fracción de esa población que coincide con el consenso ampliado es de $2,3 \times 10^7$ y la probabilidad de encontrar uno de ellos es $4,5 \times 10^{-5}$ (frecuencias iguales para los distintos aminoácidos). Si esta probabilidad se corrige con las frecuencias de los distintos aminoácidos, existentes en N, la probabilidad aumenta a $1,8 \times 10^{-4}$. Esto último implica que, para una proteína de ~585 aminoácidos (promedio de los alineamientos múltiples de N), la frecuencia esperada de hallazgo de nonapéptidos sería de 0,10 sitios/proteína.

Teniendo en cuenta las consideraciones anteriores, se realizó una búsqueda exhaustiva en las nucleoproteínas de los arnavirus, con el consenso ampliado.

En todos los arnavirus se encontraron 3 sitios, ubicados en las mismas posiciones (Figura 14.9.B) (ver también Figuras 14.7.A, B y C, 14.8 y Apéndice 6, Figura Ap6.4). En Machupo, Pichindé y Sabiá aparece, además, un cuarto sitio.

Por otra parte, con la finalidad de evaluar la confiabilidad de esta predicción, se realizaron dos controles: se buscó la existencia de nonapéptidos coincidentes con el consenso en un conjunto de 20 secuencias aminoacídicas no

relacionadas y se generó un conjunto de 20 secuencias con un orden aleatorio de los aminoácidos, basadas en las secuencias de N de los arnavirus y manteniendo las frecuencias promedio características. En ninguno de los dos conjuntos se encontraron sitios coincidentes con el consenso.

En función de lo anterior, se puede postular que los nonapéptidos hallados con el PROSITE en algunos arnavirus y los hallados con el consenso ampliado en los otros arnavirus podrían ser los responsables de estabilizar las interacciones mutuas para formar las nucleocápsides y/o las interacciones con proteínas virales o celulares para cumplir el rol funcional.

Por último, un análisis comparativo detallado de un alineamiento múltiple de las proteínas N de los arnavirus, basado principalmente en la distribución de cisteínas e histidinas (ver Figura 14.5) permite predecir la interacción de las mismas con el catión Zinc, en un motivo tipo *zinc finger*. Estos motivos no han sido detectados con la base de datos PROSITE y son diferentes a las familias de *zinc fingers* ya descritas (Pabo and Sauer, 1992).

En 1989, Salvato & Shimomaye, publicaron un trabajo en el cual demostraron la capacidad de unión de zinc por parte de la proteína Z de LCM. Dado que en los ensayos de *zinc blot* emplearon extractos proteicos completos de viriones, también pudieron observar la unión de zinc por parte de la N.

En 1993, Ghiringhelli *et al.* demostraron la capacidad de unión de zinc por parte de la proteína N del virus Junín-MC2, expresada como recombinante en el sistema baculovirus-células de insecto (ver Capítulo 8).

En función de estos resultados experimentales y teniendo en cuenta que, en un alineamiento múltiple de las proteínas N de los arnavirus, justamente los virus LCM y Junín corresponden a los dos extremos en cuanto a la longitud de la secuencia aminoacídica, se puede

postular que en todos los arnavirus la región C-terminal de la proteína N puede interactuar con el catión zinc.

Por otra parte, como se puede observar en la **Figura 14.10**, en todos los *loops* mayores de los potenciales *zinc fingers* se encuentra un conjunto de 4 o más aminoácidos básicos, precedidos por un grupo de 2 ó 3 aminoácidos hidrofóbicos. Estos conjuntos de aminoácidos básicos e hidrofóbicos podrían participar de la interacción específica con el RNA viral durante el proceso de replicación/transcripción.

Actualmente, todavía no existe una confirmación experimental que demuestre fehacientemente si existe una dependencia estricta entre la interacción con zinc y la interacción con RNA. Algunos resultados preliminares, obtenidos con la proteína N recombinante del virus Junín-MC2 (ver **Capítulo 8**), sugieren que la interacción de la proteína N con el RNA es diferente en presencia o ausencia de zinc. Es posible que la coordinación del catión zinc asegure una conformación local adecuada para la interacción con el RNA. Sin embargo, se deberían realizar experimentos adicionales que permitan

confirmar o rechazar estas observaciones. Además, deberían realizarse experimentos de interacción proteína N-RNA y proteína N-zinc con proteínas modificadas mediante mutagénesis dirigida, para confirmar o rechazar la intervención de los aminoácidos predichos.

En la **Figura 14.11** se presenta un diagrama esquemático planar de la posible disposición de los motivos estructurales que conforman la región tipo *zinc finger* de la proteína N de los virus Junín-MC2 y Junín-Candid #1. Este diagrama debería ser confirmado o rechazado mediante modelado molecular y/o mediante determinaciones experimentales de la estructura. Es de destacar que, basado en las predicciones de estructura secundaria, en la homogeneidad de patrón estructural encontrada y en las metodologías de reconocimiento de plegamiento (*folding recognition*), la estructura local de la región C-terminal (Dominio 3) de todas las proteínas N analizadas resulta equivalente.

Finalmente, en la **Figura 14.12** se presentan 2 modelos alternativos sobre las posibles bases moleculares de la transición transcripción/replicación.

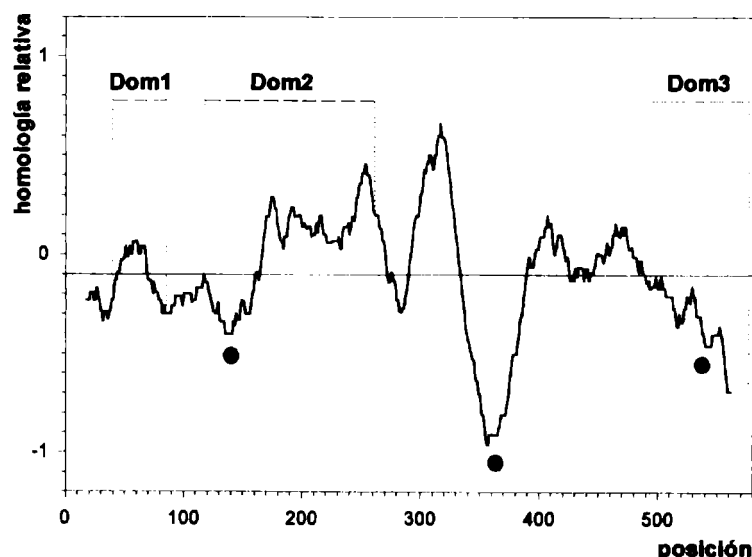


Figura 14.3. Regiones conservadas y regiones variables en las nucleoproteínas de los arnavirus Junín-Candid #1, Junín-MC2, Machupo, Oliveros, Pichindé, Sabiá, Tacaribe, Lassa-Josiah, Lassa-Nigeria, LCM-Armstrong, LCM-WE y Mopeia. El gráfico fue calculado con la rutina HOMGRAPH (Ghiringhelli, P.D., 2002, esta tesis), utilizando una ventana de 35 residuos y una tabla de equivalencias de +1,0 para las identidades, +0,5 y +0,25 para los diferentes grados de cambios conservativos (Schwartz and Dayhoff, 1979), y -1,0 para la ausencia de identidad. La línea llena horizontal indica la media de los valores de homología relativa. Los picos por encima de la media corresponden a las regiones conservadas y los picos por debajo corresponden a las regiones variables. En el gráfico se notan grandes contrastes, se observan algunas regiones conservadas (40-70; 170-270; 290-335 y 390-490) y regiones con escasa conservación; en particular, las señaladas con puntos de color fucsia corresponden a zonas en las cuales, además de las diferencias de secuencia, existen múltiples gaps. Dom1, Dom2 y Dom3, corresponden a tres potenciales dominios de interacción con RNA. (Ver también Figura Ap6.4).

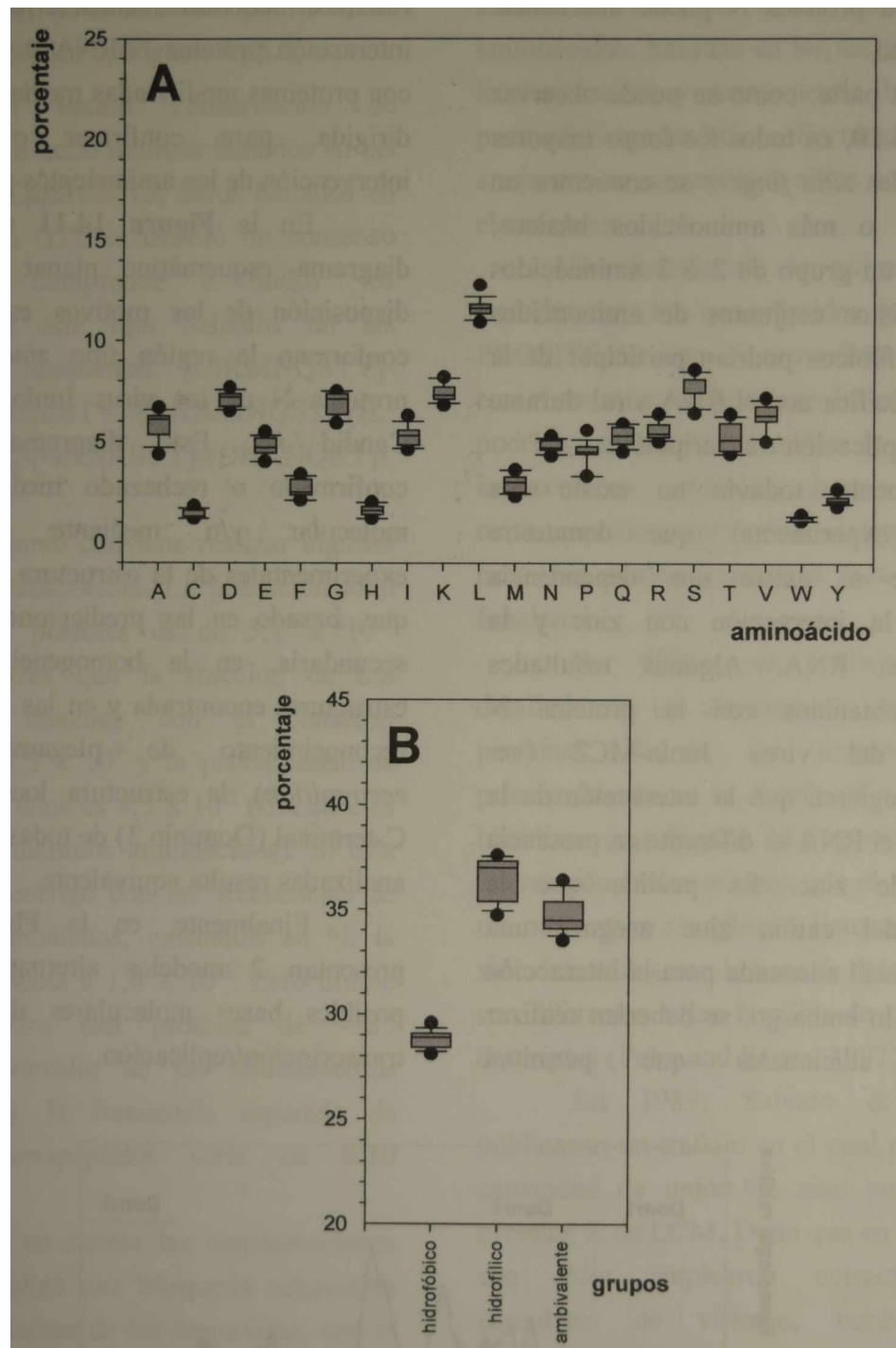


Figura 14.4. Gráficos de cajas mostrando las frecuencias y dispersión de los distintos aminoácidos en las proteínas N de los arnavirus. Los límites inferior y superior de cada caja corresponden al 25^{vo} percentil y al 75^{vo} percentil, respectivamente. Los límites inferior y superior de cada barra corresponden al 10^{mo} percentil y al 90^{vo} percentil, respectivamente. Los puntos inferior y superior corresponden al 5^{to} percentil y al 95^{vo} percentil, respectivamente. La línea horizontal dentro de la caja corresponde a la mediana (50^{vo} percentil) de los valores. La distancia desde la mediana al límite superior de la caja indica la proporción de frecuencias presente en el rango 50^{vo} percentil - 75^{vo} percentil. La distancia desde la mediana al límite inferior de la caja indica la proporción de frecuencias presente en el rango 25^{vo} percentil - 50^{vo} percentil. **A.** Proteínas N. Se observa que, aún cuando el aminoácido más abundante es hidrofóbico (L), la mayor proporción de los aminoácidos son hidrofílicos y ambivalentes. Por otra parte, la amplitud de las cajas nos indica que, en la mayor parte de los aminoácidos, los porcentajes de cada uno, en los distintos arnavirus, presentan escasa dispersión. **B.** Frecuencias y distribución de los aminoácidos agrupados de acuerdo a Karlin *et al.* (1990). En el gráfico, se resaltan las peculiaridades ya mencionadas, el orden de abundancia es: hidrofílicos, ambivalentes, hidrofóbicos..

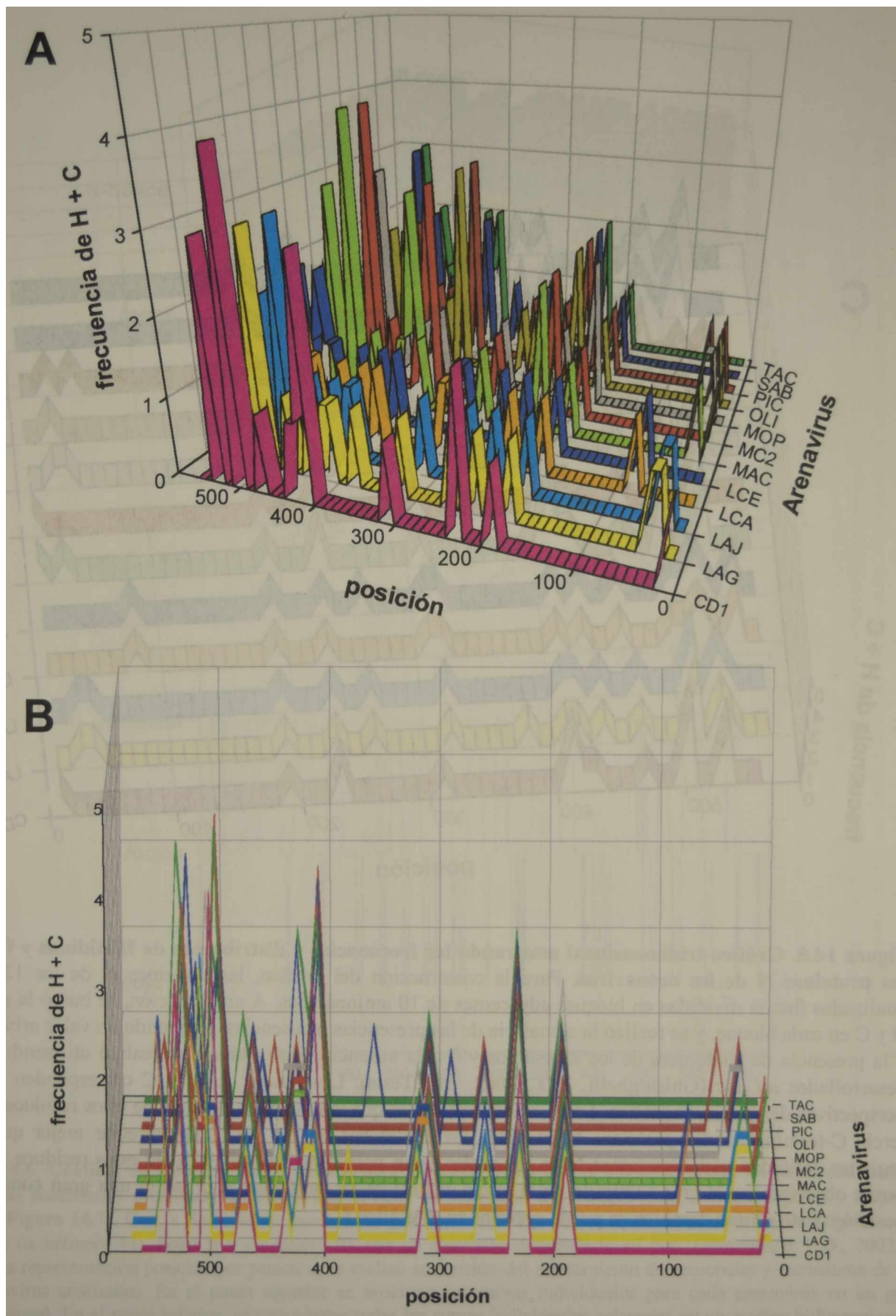


Figura 14.5. Gráficos tridimensionales mostrando las frecuencias y distribución de Histidinas y Cisteínas en las proteínas N de los arnavirus. (Ver leyenda en la página siguiente)

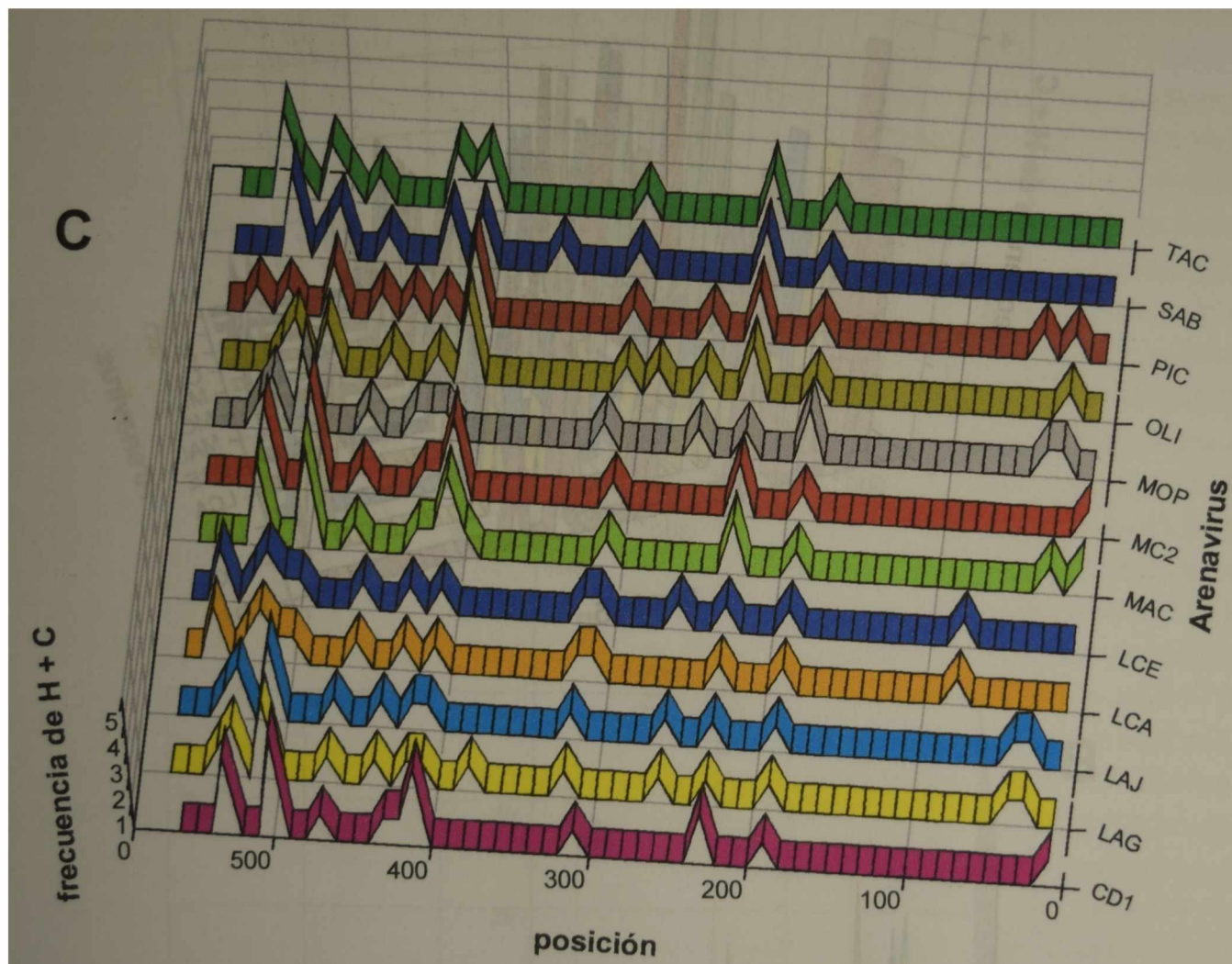


Figura 14.5. Gráfico tridimensional mostrando las frecuencias y distribución de Histidinas y Cisteínas en las proteínas N de los arnavirus. Para la construcción del gráfico, las proteínas N de los 12 arnavirus analizados fueron divididas en bloques adyacentes de 10 aminoácidos. A continuación, se buscó la presencia de H y C en cada bloque, y se realizó la sumatoria de las presencias o ausencias, asignando un valor arbitrario de +1 a la presencia de cualquiera de los 2 residuos y 0 a la ausencia. Este análisis se realizó utilizando algoritmos desarrollados *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis). Las figuras A, B y C corresponden a diferentes perspectivas de un mismo gráfico. En A. se observa mejor que la mayor abundancia de estos residuos existe en el tercio C-terminal de las proteínas N en todos los arnavirus. En B. Se puede apreciar mejor que existen 2 regiones definidas en las proteínas N en las cuales están presentes más del 90% de estos residuos. Además, se puede observar que el grupo ubicado en el tercio central de las proteínas N presenta una gran conservación en posición. Esto último, también se puede apreciar en el gráfico C.

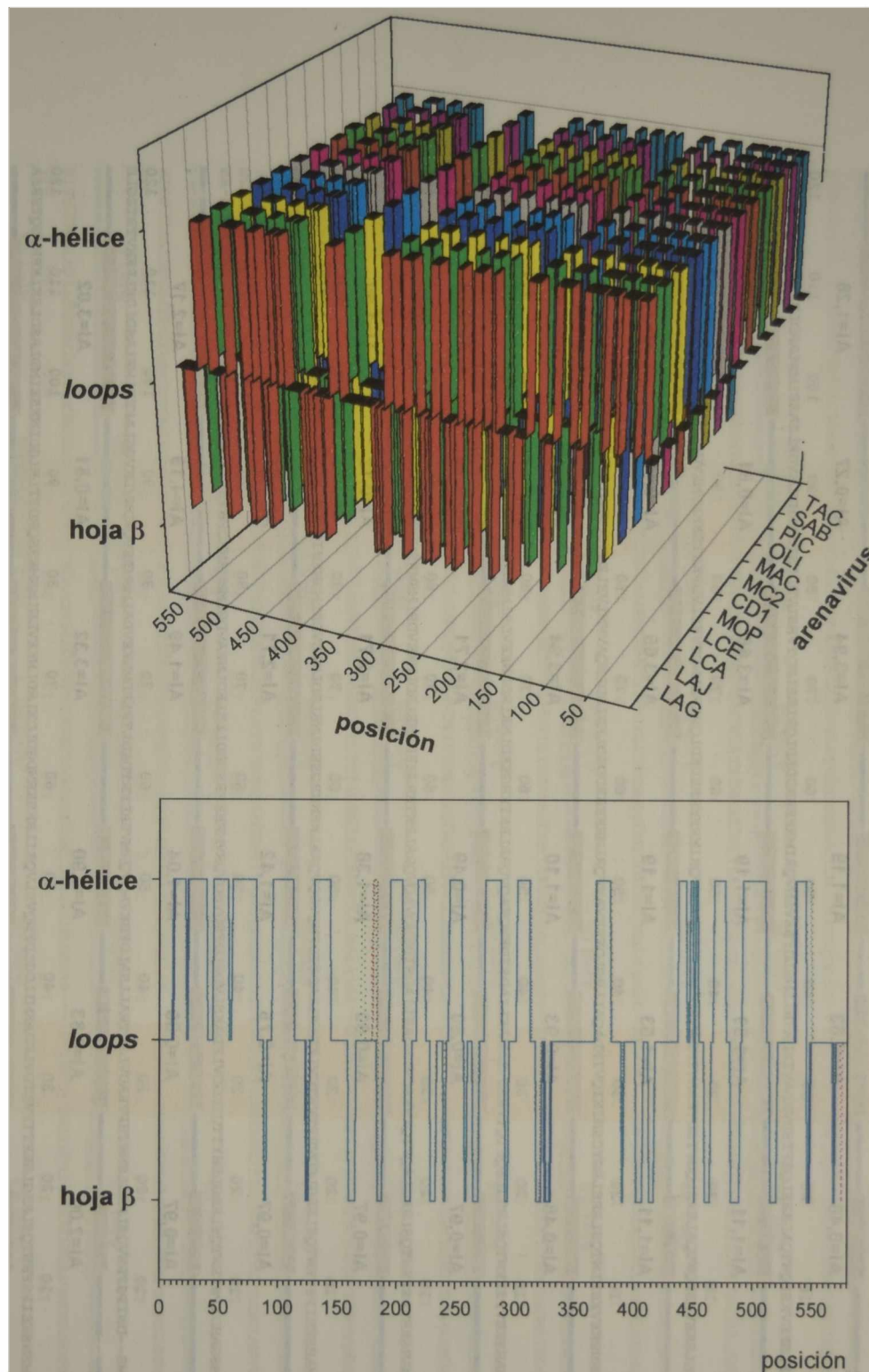


Figura 14.6. Gráficos de distribución de las estructuras secundarias en las nucleoproteínas de los arenavirus. Las estructuras secundarias predichas con el servidor JPRED corresponden a 3 categorías: α hélices, hojas β y *loops* (ver también **Figura 14.7**). Con la finalidad de construir el gráfico, a cada uno de los tipos estructurales se le asignó un valor arbitrario (α hélices= +1; hojas β = -1; *loops*= 0), con un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis). La representación posición por posición, se realizó en función del alineamiento de secuencias y estructuras de todos los arenavirus analizados. En el panel superior se muestran las curvas individuales para cada arenavirus en un gráfico tridimensional. En el panel inferior, se representan todas las curvas individuales solapadas en un gráfico bidimensional, en el cual se puede observar mejor la homogeneidad de patrón estructural existente en las nucleoproteínas. LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE- LCM-WE; MOP: Mopeia; CD1: Junín-Candid #1; MC2: Junín-MC2; MAC: Machupo; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe.

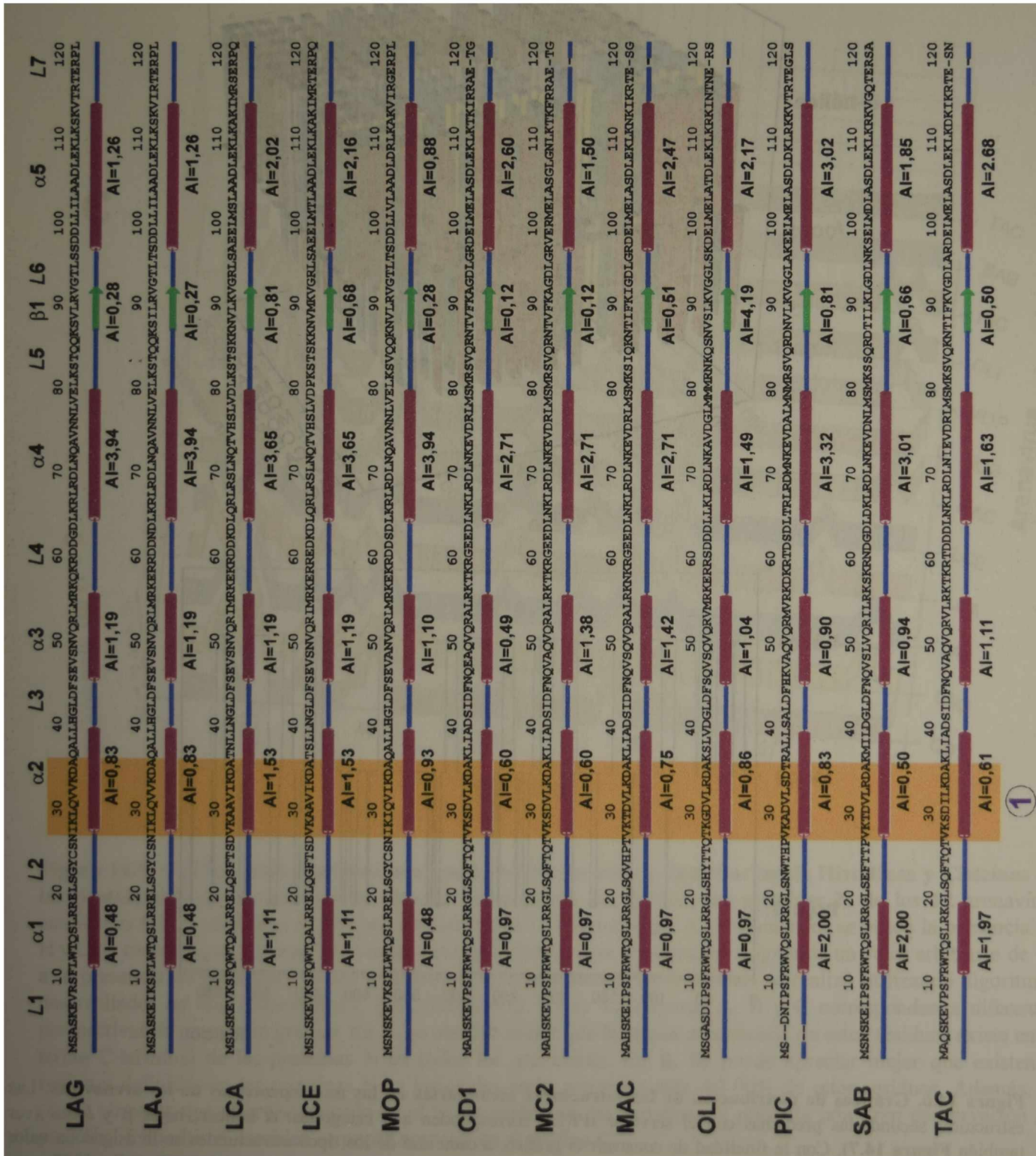


Figura 14.7.A. Comparación de estructuras secundarias predichas para las nucleoproteínas de los arnavirus. (Ver leyenda en la última página)

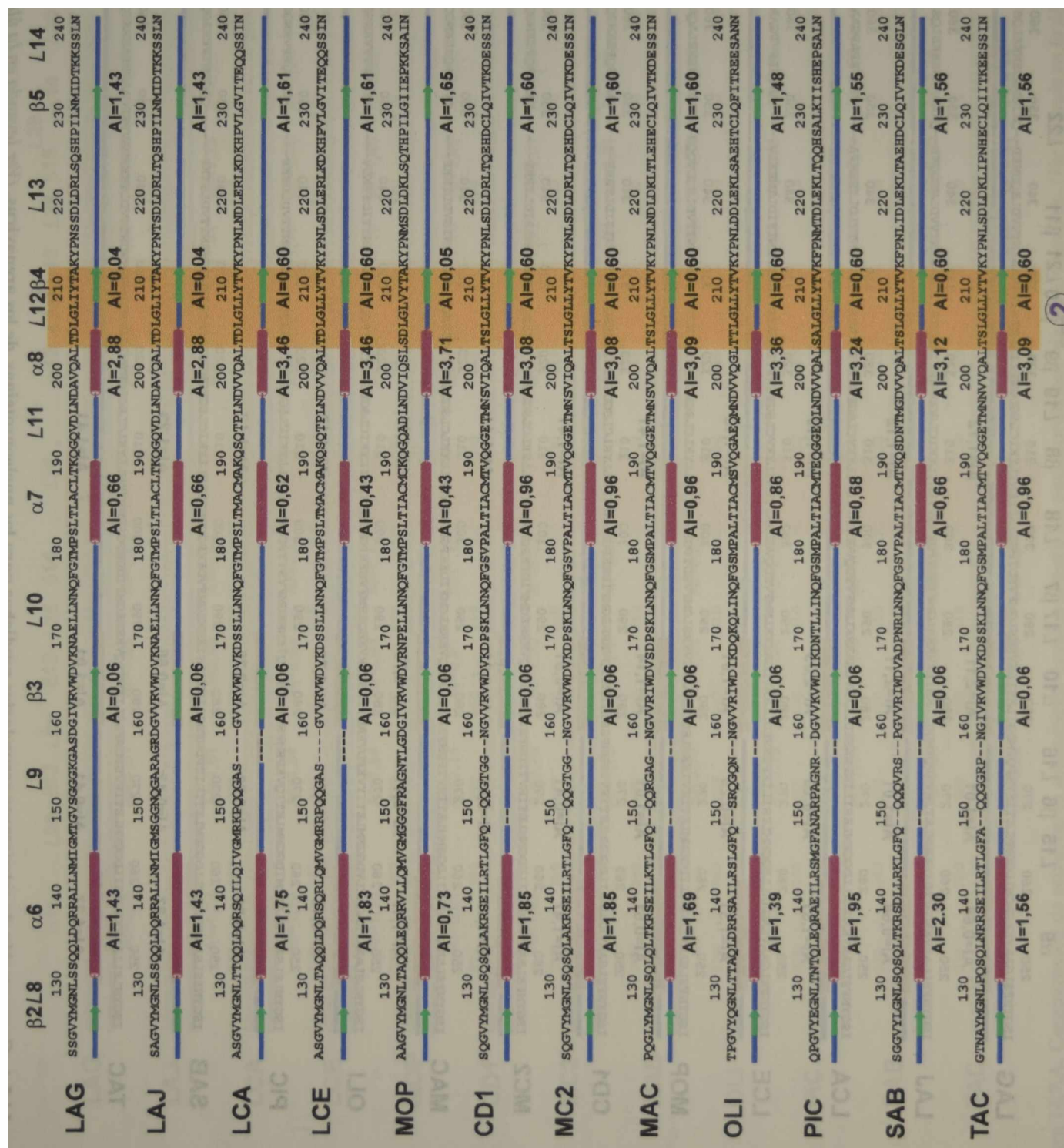


Figura 14.7.A. Comparación de estructuras secundarias predichas para las nucleoproteínas de los arnavirus. (Ver leyenda en la última página)

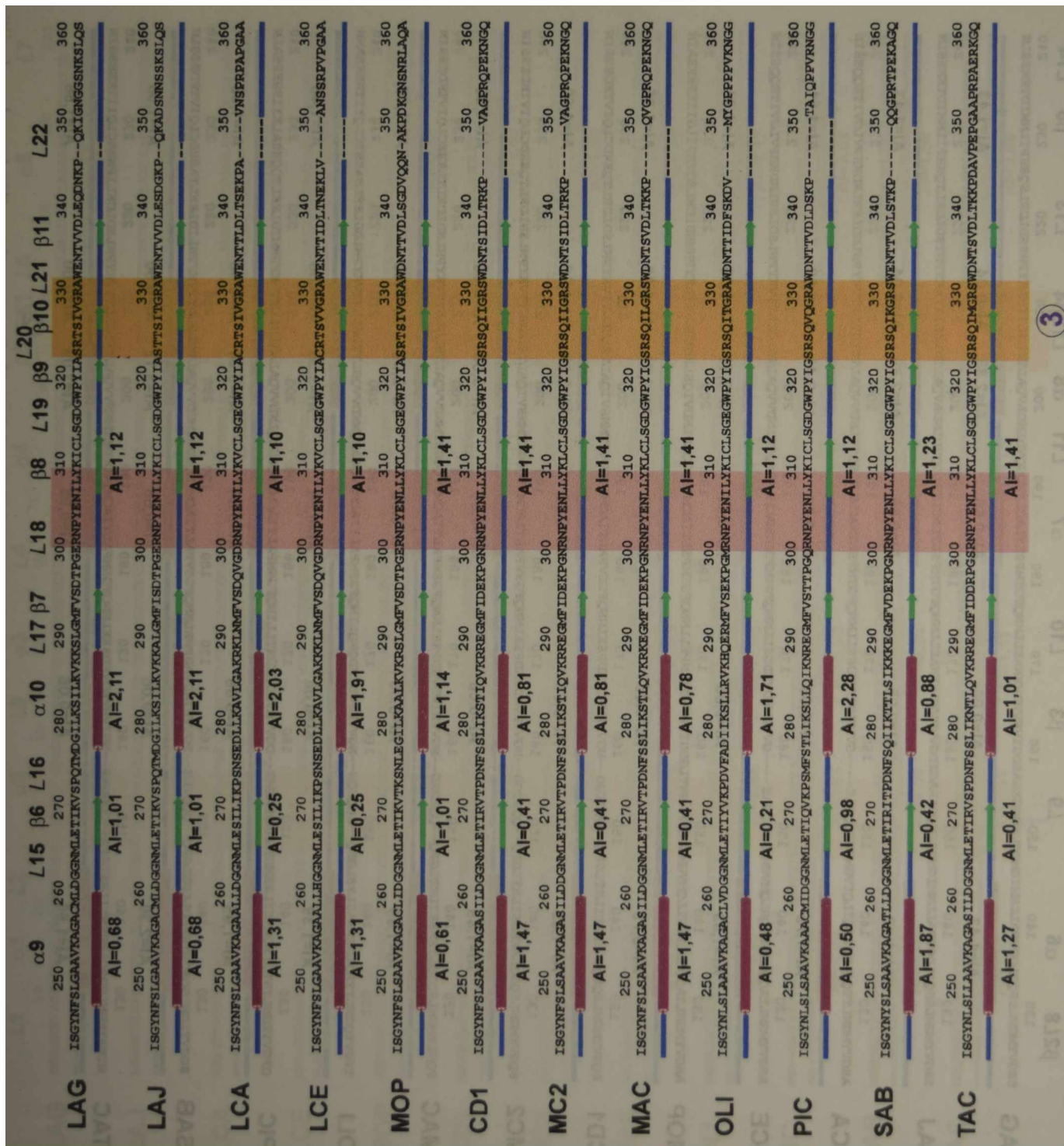


Figura 14.7.A. Comparación de estructuras secundarias predichas para las nucleoproteínas de los arnavirus. (Ver leyenda en la última página)

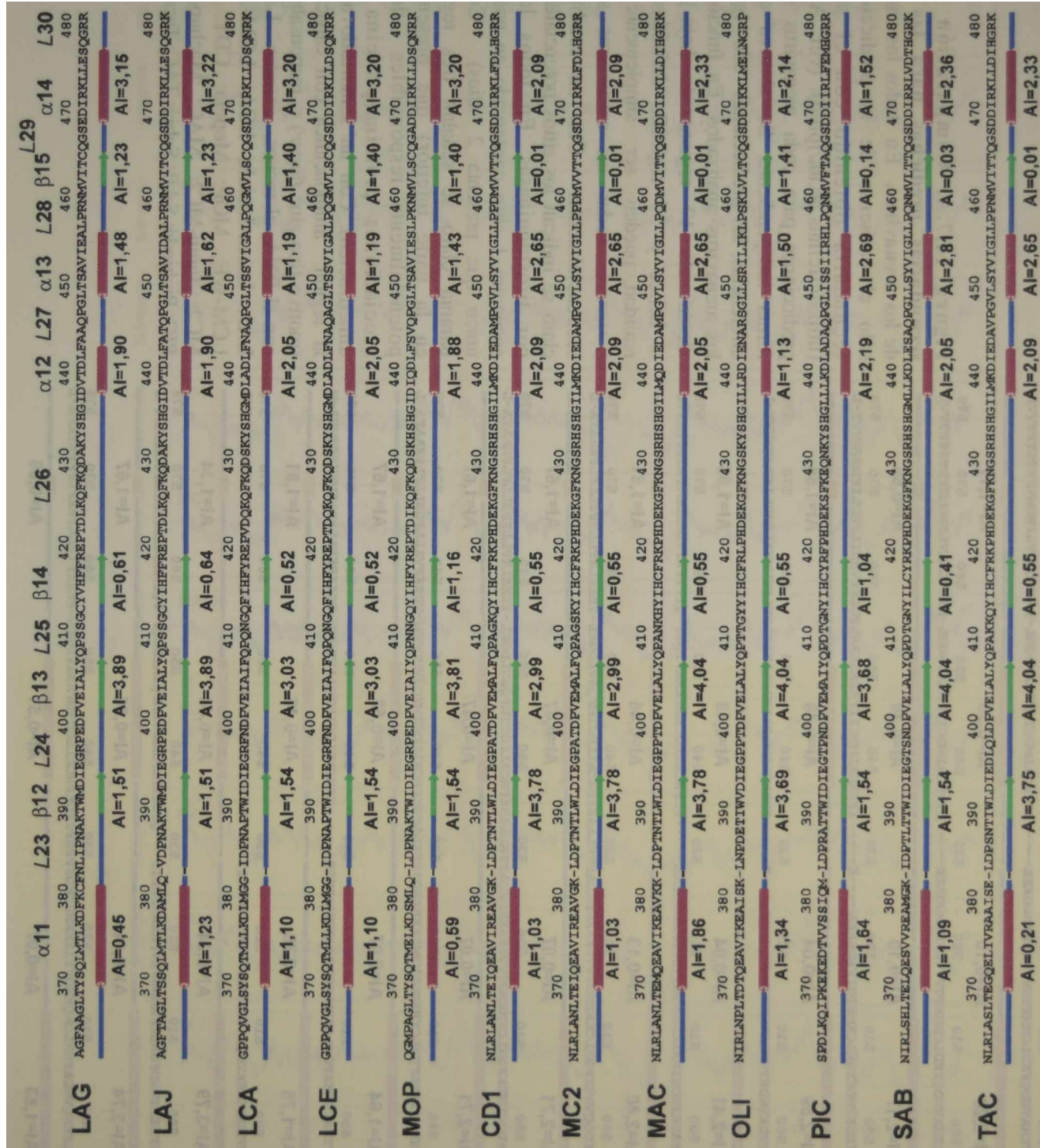


Figura 14.7.A. Comparación de estructuras secundarias predichas para las nucleoproteínas de los arnavirus. (Ver leyenda en la última página)



Figura 14.7.A. Comparación de estructuras secundarias predichas para las nucleoproteínas de los arnavirus. En todos los casos se representa sólo el consenso de la predicción de estructura secundaria obtenido con el servidor JPRED (<http://circinus.ac.uk/jpred>) (ver Apéndice 8). Las predicciones para cada arnavirus se realizaron en forma independiente. Se observa que, salvo ligeras diferencias, las predicciones son iguales para todos los arnavirus analizados. En función de esto, se han identificado 17 regiones de hoja β, 17 regiones de α hélice y 35 regiones definidas como *loop* (en realidad, pueden ser simplemente *loops* o contener β *turns*). Todas las estructuras se han numerado secuencialmente. Con un sombreado color rosa claro se indican los sitios potenciales de fosforilación por tirosininas. En todos los arnavirus se encuentra 1 sitio conservado en posición (PIC es el único que presenta 2 sitios). Con un sombreado naranja claro se señalan tres regiones (numeradas en la parte inferior) que presentan nonapeptidos potencialmente responsables de estabilizar las asociaciones proteína-proteína (señales de dimerización). Con un sombreado amarillo se indica la región aminoacídica con capacidad de unir zinc (motivo tipo zinc *finger*) y, responsable de la interacción con el RNA, al menos, *in vitro*. LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia; CD1: Junin-Candid #1; MC2: Junin-MC2; MAC: Machupo; OLI: Oliveros; PIC: Pichindé; SAB: Sabia; TAC: Tacaribe.

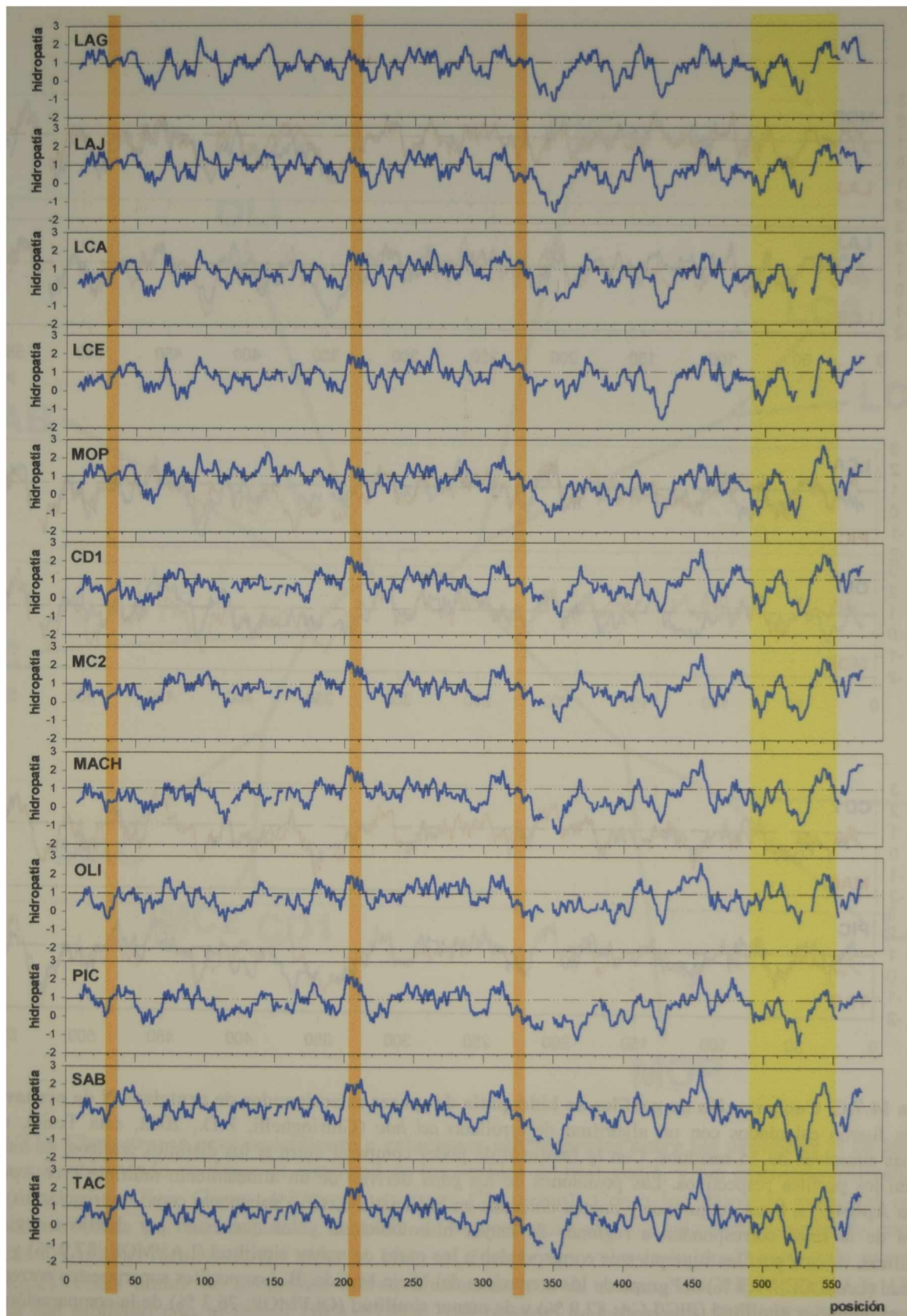


Figura 14.7.B. Perfiles de hidropatía de las proteínas N de los arnavirus. Los perfiles fueron calculados con un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis), utilizando ventanas sucesivas de 15 residuos. Con un sombreado anaranjado se señalan los nonapéptidos que constituyen las señales de dimerización. Con un sombreado amarillo claro se indica la región correspondiente al zinc *finger* (ver Apéndice 6, Figura Ap.6.4). La línea de guiones y puntos ubicada en hidropatía 1, sirve solamente como referencia; los picos por encima de la línea corresponden a regiones de mayor hidrofobicidad y los que están por debajo a regiones más hidrofílicas. LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE- LCM-WE; MOP: Mopeia; CD1: Junín-Candid #1; MC2: Junín-MC2; MACH: Machupo; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe.

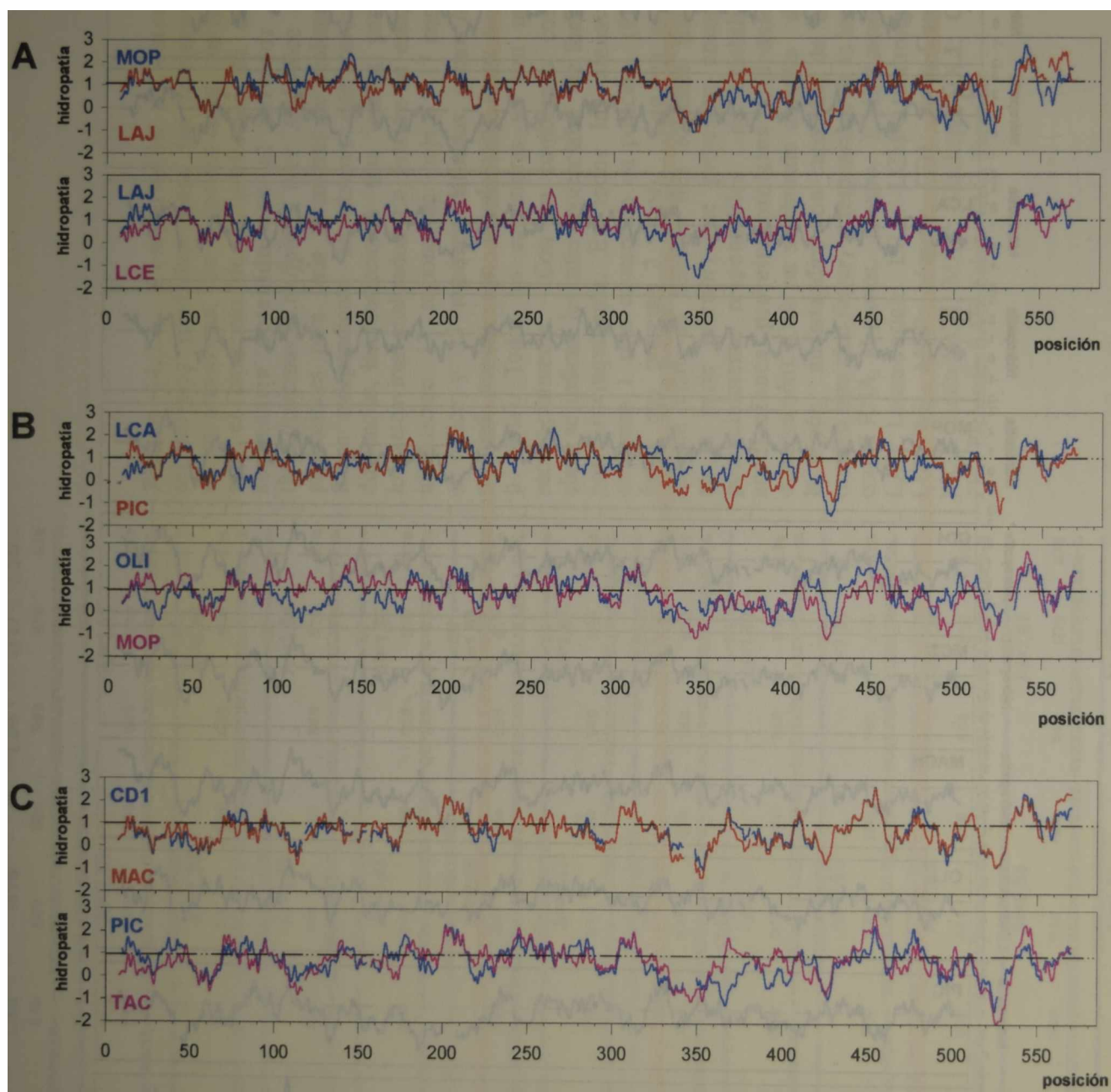


Figura 14.7.C. Comparación de perfiles de hidropatía de pares seleccionados de proteínas N de arnavirus. Los perfiles fueron calculados con un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis), utilizando ventanas sucesivas de 15 residuos. Con la finalidad de poder comparar entre sí los distintos gráficos, se introdujeron *gaps* en los perfiles respectivos. Las posiciones de los *gaps* derivan de un alineamiento múltiple (ver Apéndice 6, Figura Ap6.4). La línea de guiones y puntos ubicados en hidropatía 1, sirve solamente como referencia; los picos por encima de la línea corresponden a regiones de mayor hidrofobicidad y los que están por debajo a regiones más hidrofílicas. **A.** Los perfiles superpuestos corresponden a los pares de mayor similitud (LAJ/MOP; 87,9 %) y de menor similitud (LAJ/LCE; 79,3 %) del grupo de los arnavirus del Viejo Mundo. **B.** Los perfiles superpuestos corresponden a los pares de mayor similitud (PIC/LCA; 82,9 %) y de menor similitud (OLI/MOP; 76,3 %) de la comparación entre los arnavirus del Viejo Mundo y los del Nuevo Mundo. **C.** Los perfiles superpuestos corresponden a los pares de mayor similitud (CD1/MAC; 92,7 %) y de menor similitud (PIC/TAC; 80,0 %) del grupo de los arnavirus del Nuevo Mundo (ver Tabla 14.c.1). LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE- LCM-WE; MOP: Mopeia; CD1: Junín-Candid #1; MC2: Junín-MC2; MAC: Machupo; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe.

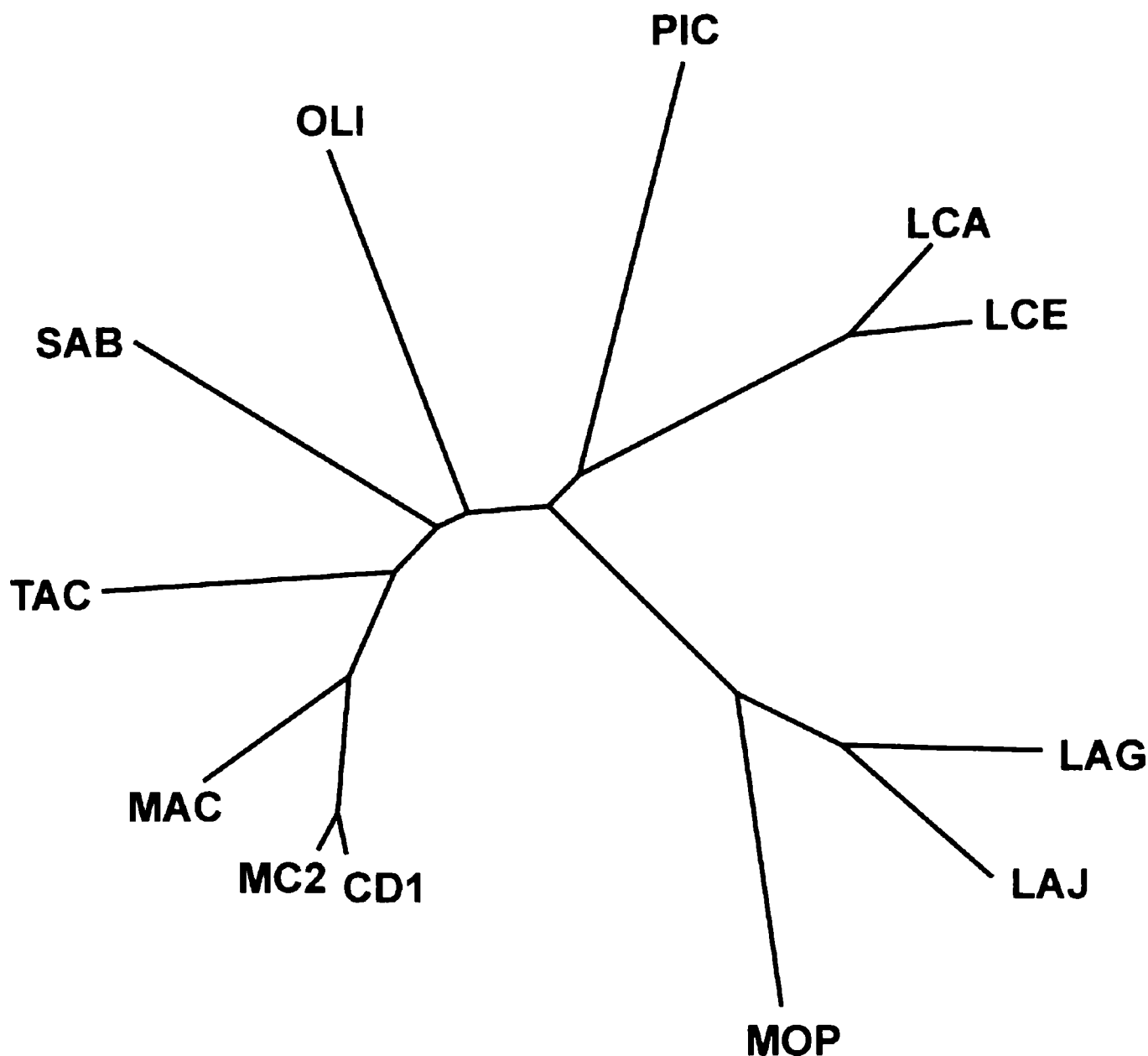


Figura 14.7.D. Cladograma correspondiente a relaciones estructurales entre las proteínas N de los arnavirus. El cladograma está basado en datos de similitud de los perfiles de hidropatía, y fue calculado utilizando el método UPGMA (*Unweighted Pair Group Arithmetic Average Clustering*). Como los datos de similitud representan la proximidad entre cada par de secuencias, la tabla de valores utilizada fue calculada asumiendo como distancias filogenéticas los valores resultantes del cálculo: $100 - v_n$ (donde v_n = valor individual) para cada par de secuencias (**Tabla 14.c.2**). Estas tablas de distancias fueron analizadas con la rutina NEIGHBOR (método UPGMA) y el cladograma fue dibujado con la rutina DRAWTREE, del paquete de programas PHYLIP (*Phylogeny Inferencing Package*, v 3.5, Felsenstein, J., 1995). Las longitudes de las distintas ramas en el cladograma están en relación con la distancia evolutiva. CD1: Junín-Candid #1; MC2: Junín-MC2; MAC: Machupo; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe; LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE: LCM-WE; MOP: Mopeia

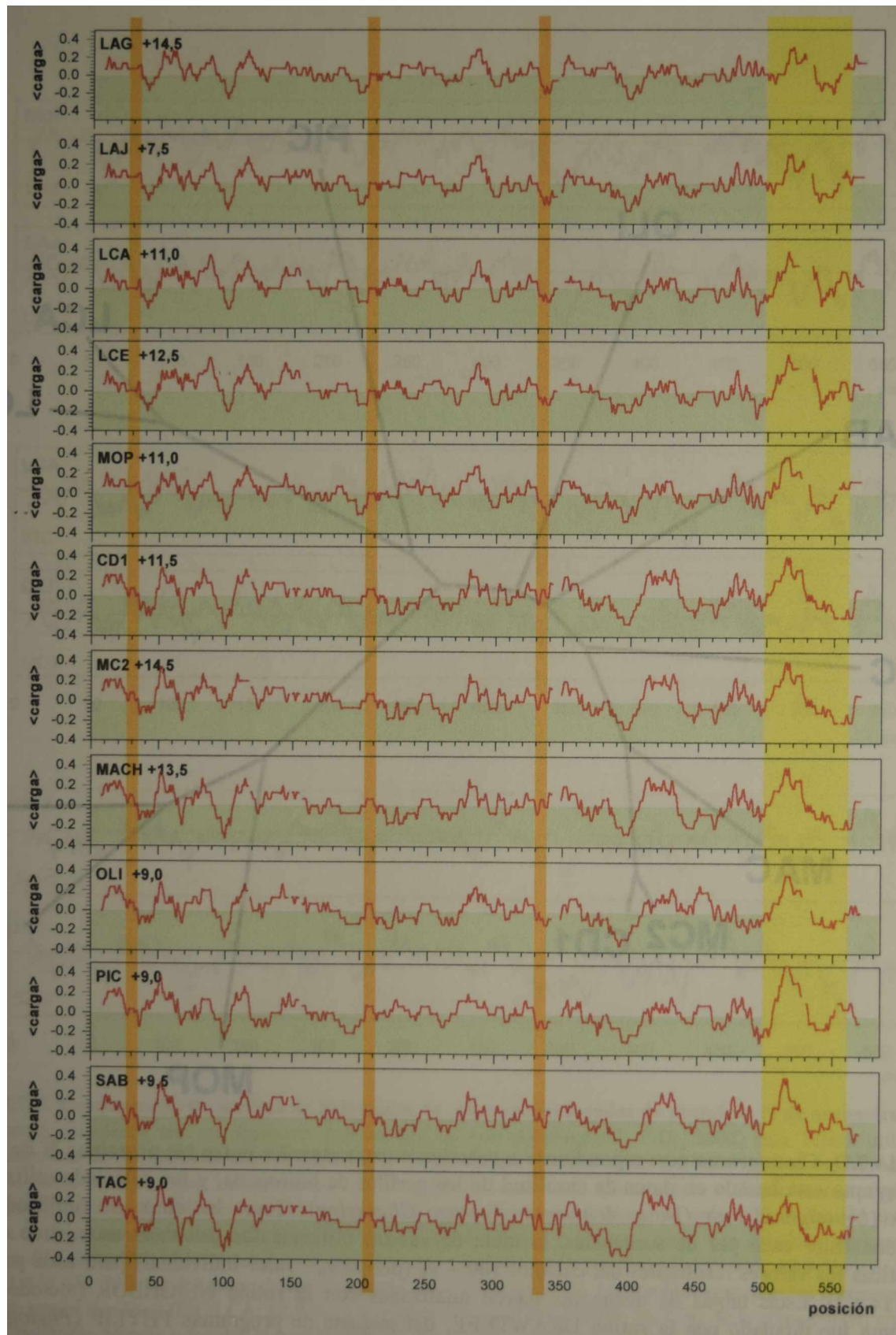


Figura 14.8. Perfiles de carga neta local de las proteínas N de los arnavirus. Los perfiles fueron calculados con un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis), utilizando ventanas sucesivas de 15 residuos. Con un sombreado anaranjado se señalan los nonapéptidos que constituyen las señales de dimerización. Con un sombreado amarillo claro se indica la región correspondiente al zinc *finger* (ver Apéndice 6, Figura Ap6.4). Al lado de los nombres, se indica la carga neta global a pH 7,0. Para el cálculo, se utilizó +1,0 para K y R, +0,5 para H y -1 para D y E. La mitad inferior de cada gráfico, sombreada con verde claro, corresponde a las regiones con carga neta negativa. LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE- LCM-WE; MOP: Mopeia; CD1: Junín-Candid #1; MC2: Junín-MC2; MACH: Machupo; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe.

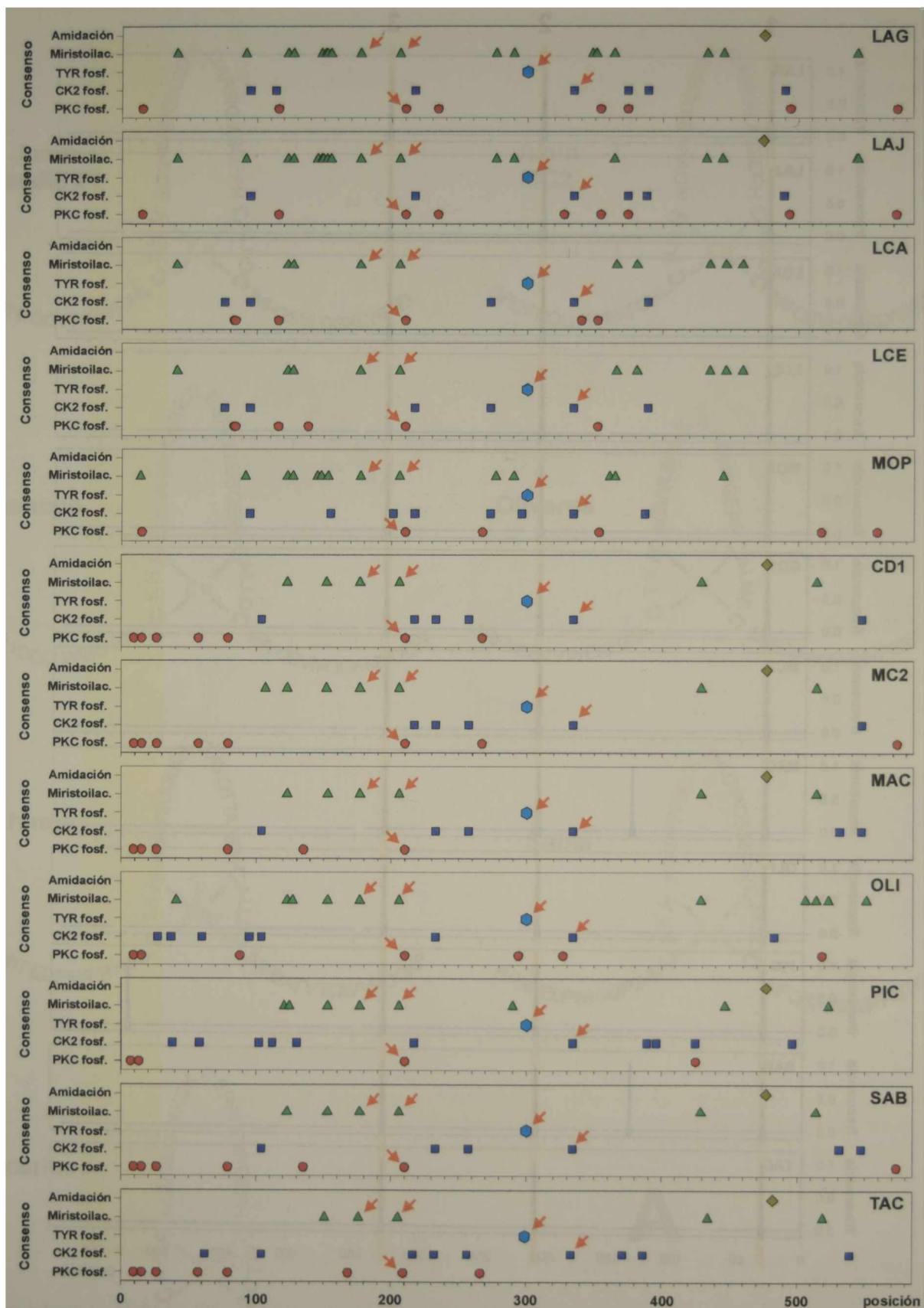


Figura 14.9.A. Sitios consenso de fosforilación y amidación en las proteínas N de arnavirus. Los diferentes sitios consenso (indicados en el eje Y) se representan con distintos símbolos rellenos. Sólo se han representado los sitios consenso presentes en, al menos, 6 de los 12 arnavirus. Con círculos rojos se representan los sitios de fosforilación por proteinquinasa C (PKC). Con cuadrados azules se representan los sitios de fosforilación por caseinquinasa II (CK2). Con hexágonos celestes se representan los sitios de fosforilación por tirosinquinasa. Con triángulos verdes se representan los sitios de miristoilación. Con rombos amarillo oscuro se representan los sitios de amidación. Las flechas color naranja señalan los sitios conservados en tipo y posición en todas las proteínas N de los arnavirus. En este gráfico no se han representado los sitios de dimerización (ver Figura 14.7.A). LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE- LCM-WE; MOP: Mopeia; CD1: Junín-Candid #1; MC2: Junín-MC2; MAC: Machupo; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe.

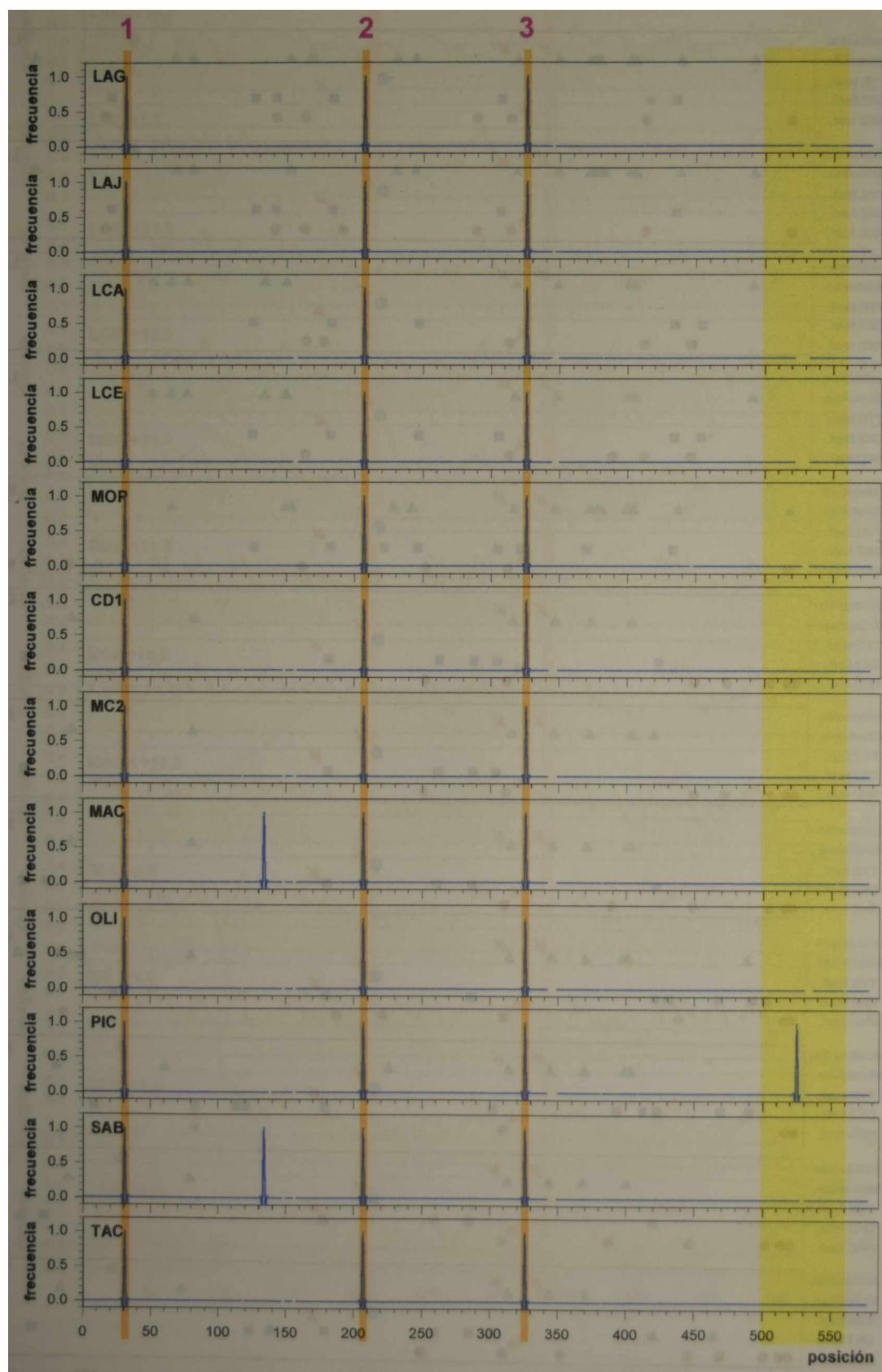


Figura 14.9.B. Frecuencia de sitios potenciales de dimerización en las proteínas N de los arenavirus. Los sitios indicados corresponden a todos los nonapéptidos que coinciden con un consenso general definido en esta tesis: [CIKMNQSTV]-[ADKQRST]-[AIGLSTV]-[AHIDGNQST]-[ILV]-[FILKMQTV]-[GKNRSTY]-[DHKNQRT]-[ANSTV]. Con un sombreado anaranjado se indican los nonapéptidos identificados con el PROSITE, para algunos arenavirus (1: OLI; 2: OLI; 3: LAG, MOP, CD1, MC2, MAC y TAC), y extrapolados para el resto (ver Figura 14.7.A). Aún cuando este consenso es más amplio que el definido en el PROSITE, sólo se identifica un nonapéptido adicional en 3 de los arenavirus. Con un sombreado amarillo claro se indica la región correspondiente al potencial zinc finger (ver Apéndice 6, Figura Ap6.4). LAG: Lassa-Nigeria; LAJ: Lassa-Josiah; LCA: LCM-Armstrong; LCE- LCM-WE; MOP: Mopeia; CD1: Junín-Candid #1; MC2: Junín-MC2; MAC: Machupo; OLI: Oliveros; PIC: Pichindé; SAB: Sabiá; TAC: Tacaribe. Los cálculos fueron realizados con un algoritmo desarrollado *ad hoc* (Ghiringhelli, P.D., 2002, esta Tesis).

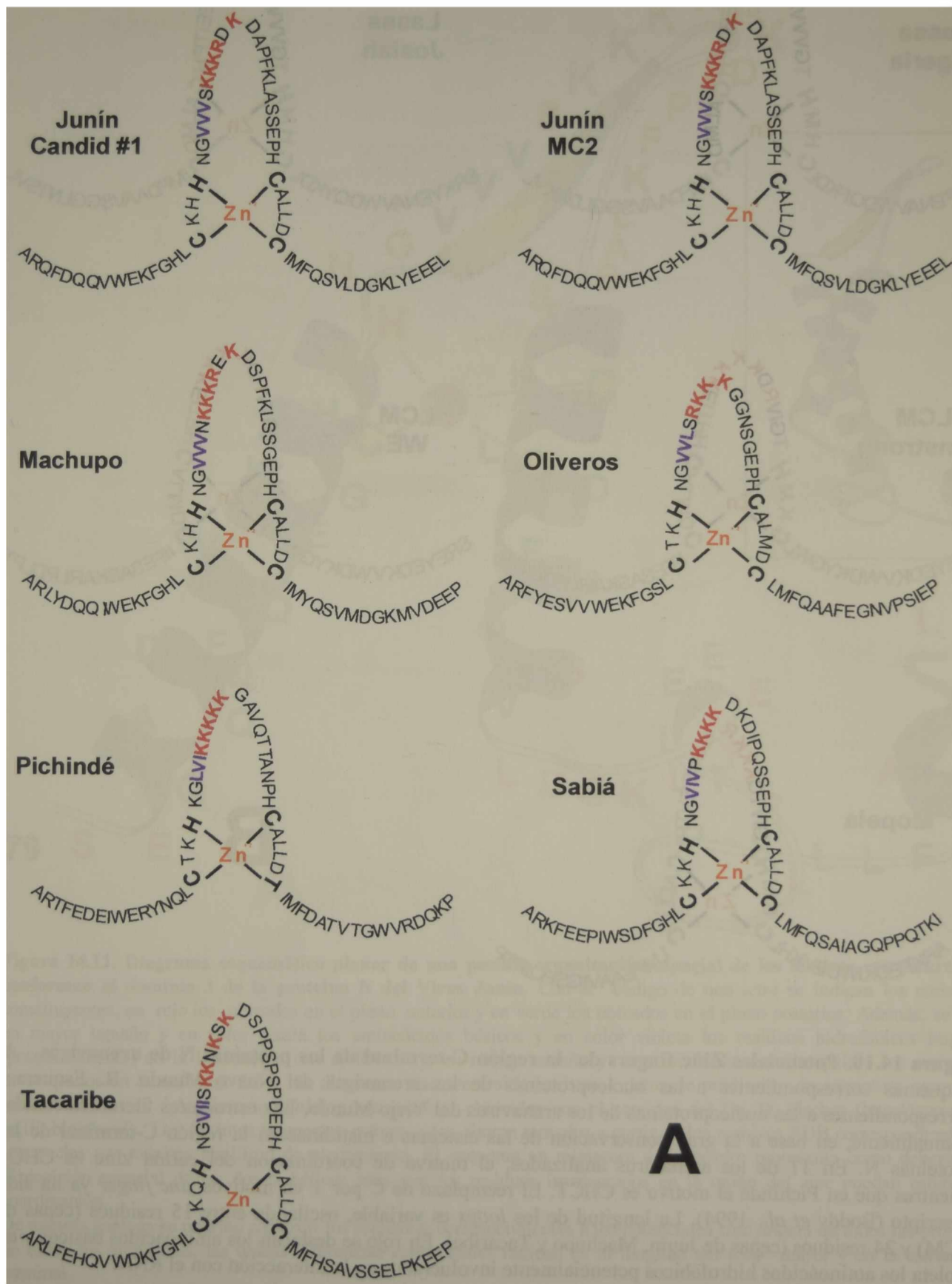


Figura 14.10. Potenciales Zinc fingers de la región C-terminal de las proteínas N de arenavirus. (Ver leyenda en la página siguiente).

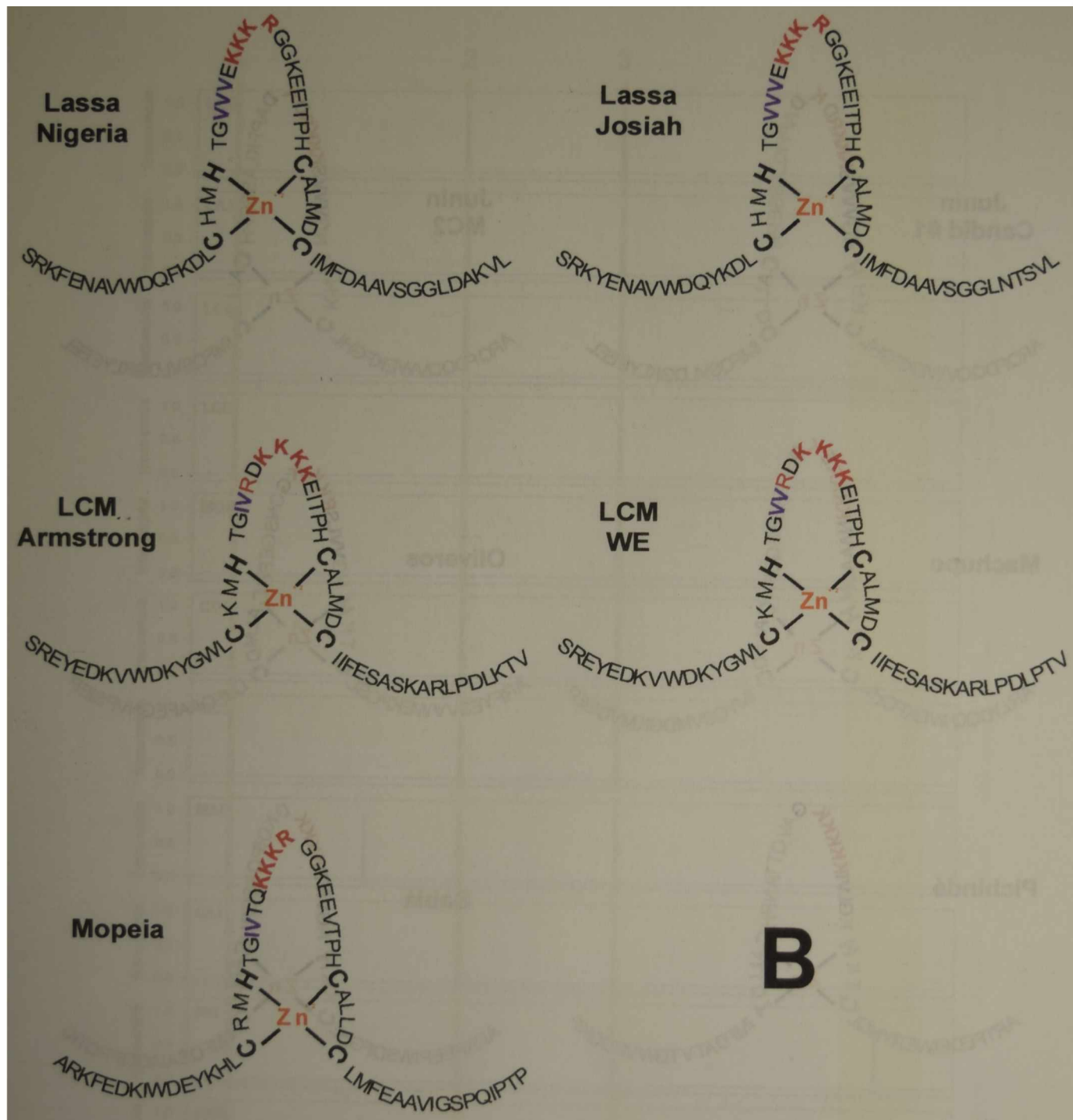


Figura 14.10. Potenciales Zinc fingers de la región C-terminal de las proteínas N de arnavirus. A. Esquemas correspondientes a las nucleoproteínas de los arnavirus del Nuevo Mundo. **B.** Esquemas correspondientes a las nucleoproteínas de los arnavirus del Viejo Mundo. Las estructuras fueron calculadas manualmente, en base a la gran conservación de las cisteínas e histidinas en la región C-terminal de las proteínas N. En 11 de los arnavirus analizados, el motivo de coordinación del catión zinc es CHC_2 , mientras que en Pichindé el motivo es CHCT. El reemplazo de C por T en motivos *zinc finger* ya ha sido descrito (Boddy *et al.*, 1994). La longitud de los *loops* es variable, oscilando entre 15 residuos (cepas de LCM) y 24 residuos (cepas de Junín, Machupo y Tacaribe). En rojo se destacan los aminoácidos básicos y en violeta los aminoácidos hidrofóbicos potencialmente involucrados en la interacción con el RNA.

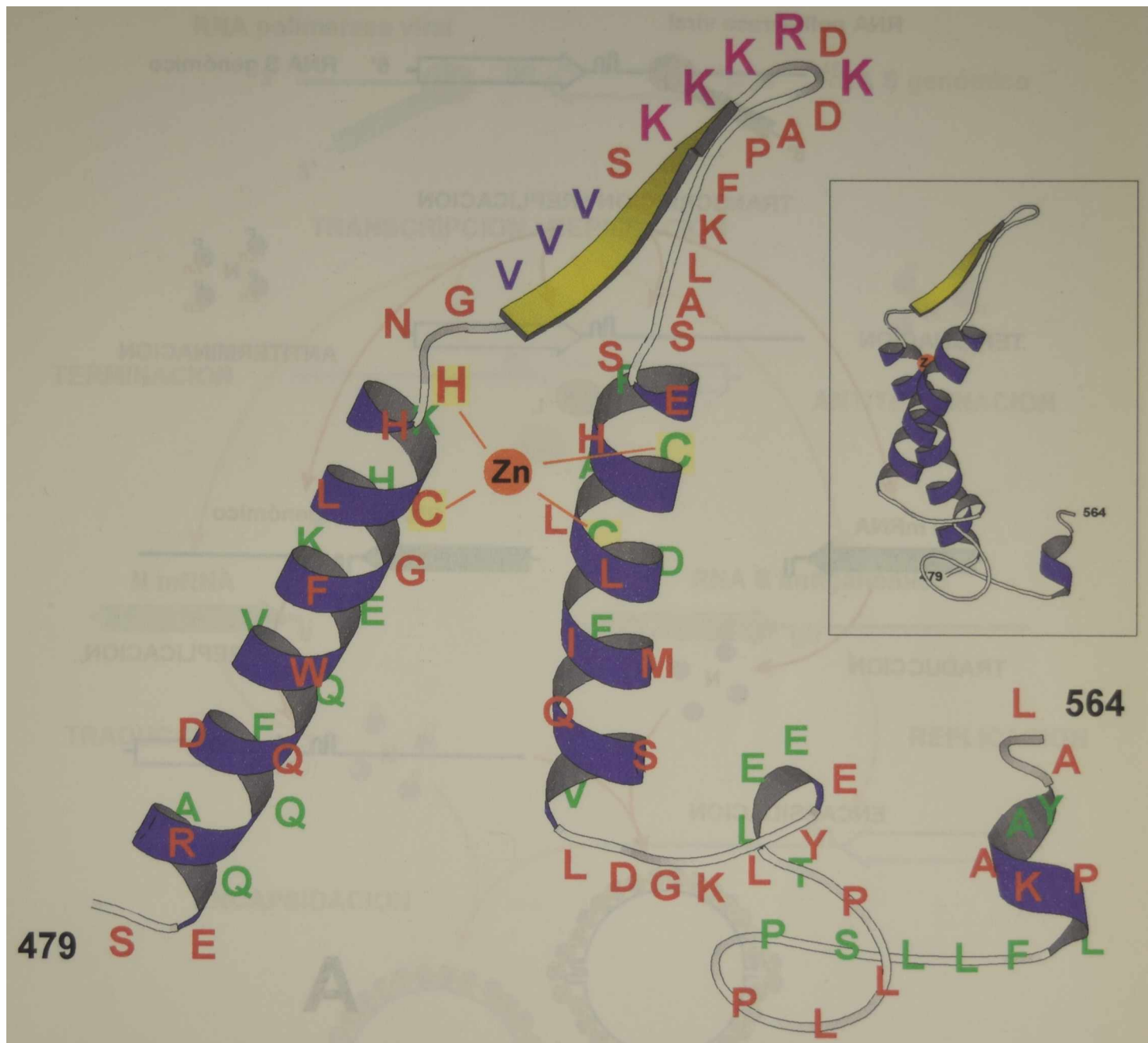


Figura 14.11. Diagrama esquemático planar de una posible organización espacial de los motivos estructurales que conforman el dominio 3 de la proteína N del Virus Junín. Con el código de una letra se indican los aminoácidos constituyentes, en rojo los ubicados en el plano anterior y en verde los ubicados en el plano posterior. Además, se indican en mayor tamaño y en color fucsia los aminoácidos básicos y en color violeta los residuos hidrofóbicos implicados (potencialmente) en la interacción con el RNA; y, en mayor tamaño y con un fondo amarillo pálido los aminoácidos involucrados en la unión al Zn^{+2} . El diagrama fue construido en base a datos de predicción de estructura secundaria y reconocimiento de plegamiento (*Folding recognition*) obtenidos con el servidor del Dr. D. Fischer (<http://www.cs.bgu.ac.il/~bioinbgu/>). Los distintos segmentos estructurales fueron tomados a partir de los archivos PDB correspondientes a las secuencias con máxima similitud de plegamiento. El esquema en miniatura del recuadro representa como debería ser la disposición espacial de las dos α hélices para que los residuos involucrados en la unión del zinc puedan establecer la coordinación.

De manera análoga se podrían construir los esquemas correspondientes a los potenciales *zinc fingers* de todas las proteínas N de todos los arenavirus; las únicas diferencias estructurales radican en la longitud del *loop* mayor y en la estructura C-terminal.

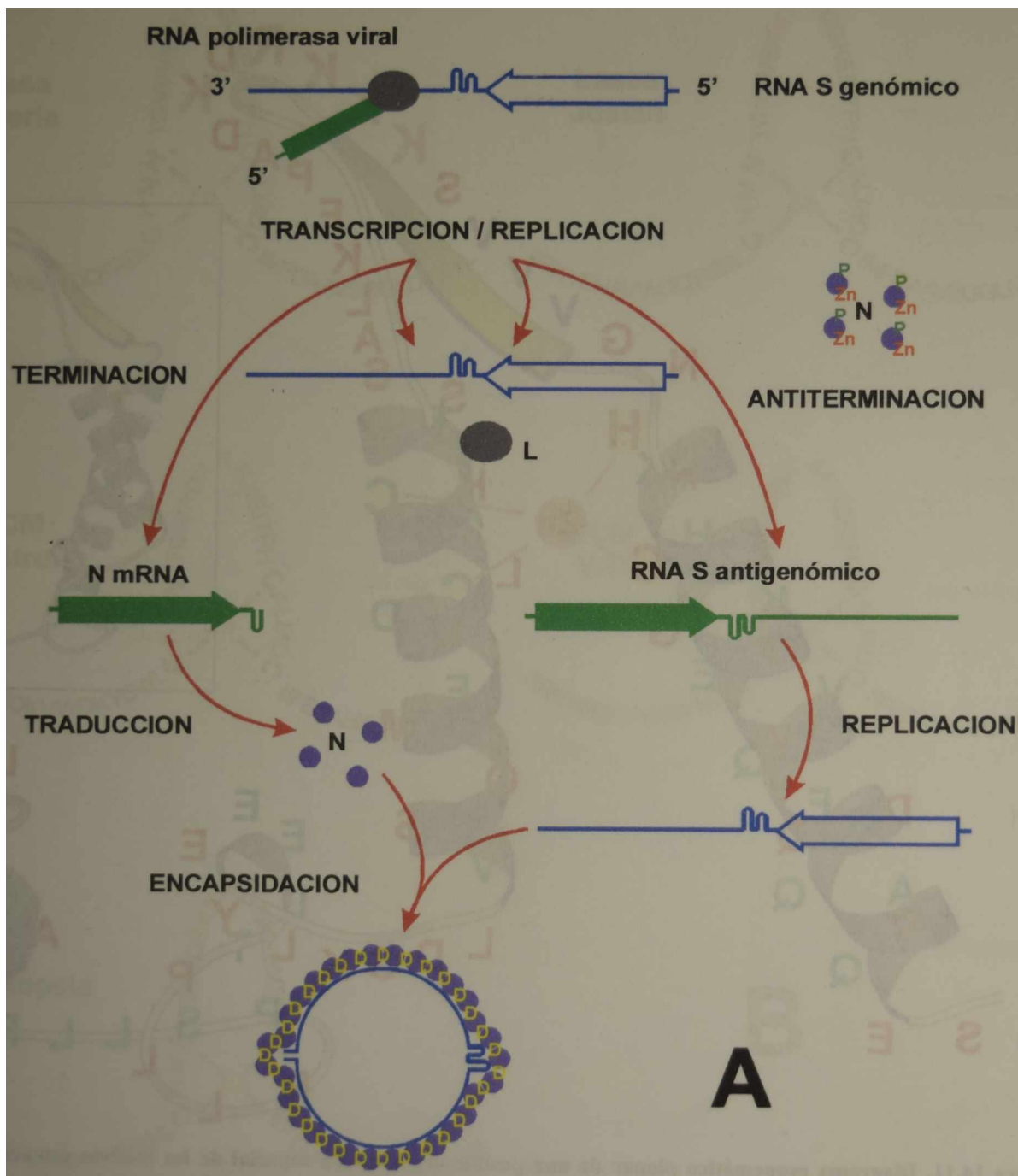


Figura 14.12. Modelos alternativos propuestos para la transición transcripción/replicación en los arenavirus. A. Primer modelo. En este modelo se postula que la coordinación de zinc es un prerequisite para la actividad antiterminadora de la proteína N, la cual se activa por fosforilación en, al menos, alguno de los sitios conservados (PKC, CK2 o tirosinquinasa). Mientras que las moléculas de proteína N no fosforiladas son utilizadas para la encapsidación. La estabilidad de las nucleocápsides se aseguraría mediante interacción proteína-proteína (dominios de dimerización) y mediante interacción electrostática proteína-RNA (regiones de carga neta local positiva). Las P en verde oscuro representan los fosfatos; el zinc se indica con el símbolo correspondiente en naranja y los dominios de dimerización se indican con una letra D en amarillo. Las esferas lila claro representan las moléculas de proteína N. (Ver Segundo modelo en la página siguiente)

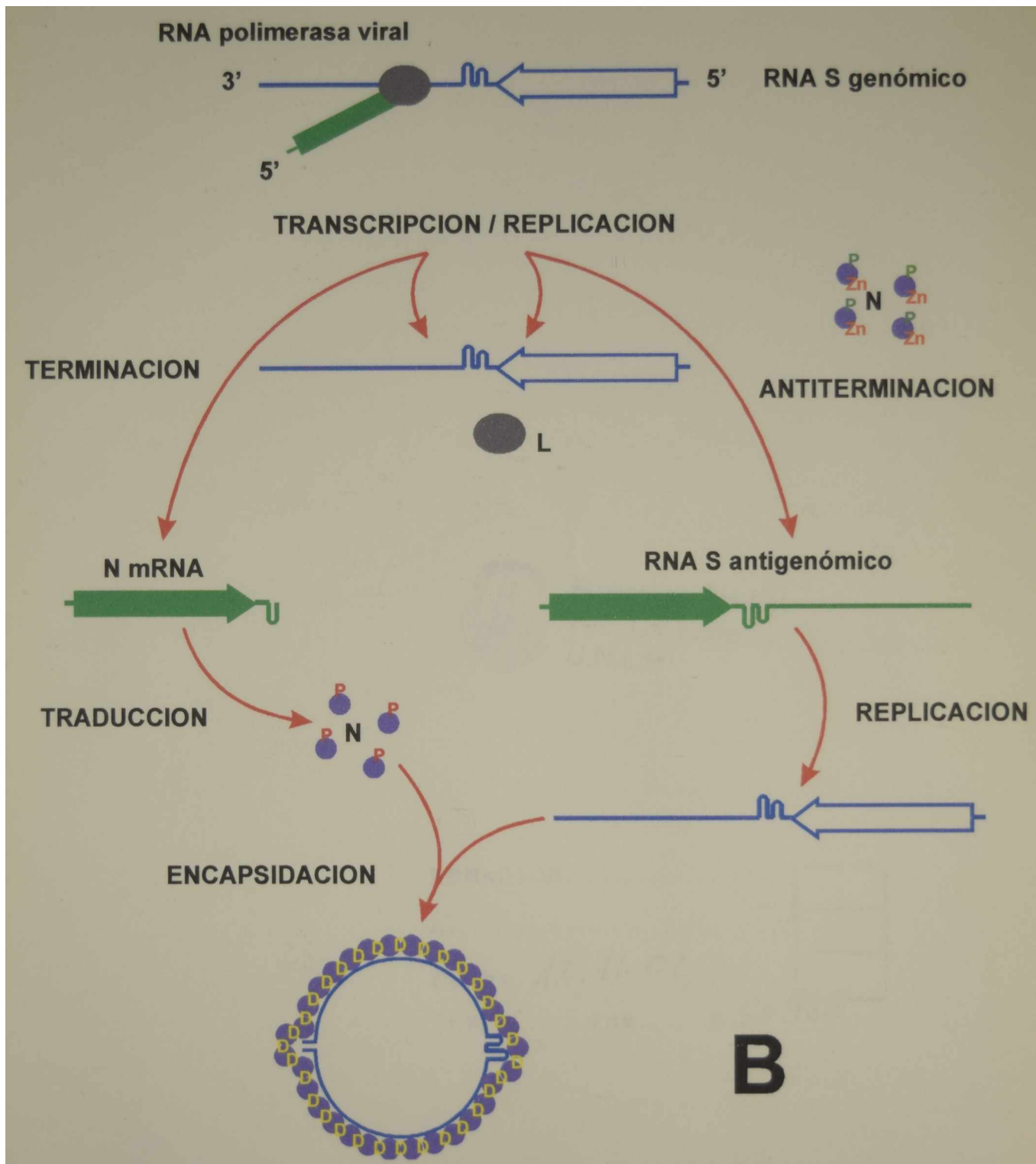


Figura 14.12. Modelos alternativos propuestos para la transición transcripción/replicación en los arenavirus. B. Segundo modelo. En este modelo se postula que la coordinación de zinc y la fosforilación en, al menos, uno de los sitios conservados (por ej.: el de tirosinquinasa), son prerequisites para la actividad antiterminadora de la proteína N. Mientras que las moléculas de proteína N fosforiladas en un sitio diferente (por ej.: PKC o CK2) son utilizadas para la encapsidación. La estabilidad de las nucleocápsides se aseguraría mediante interacción proteína-proteína (dominios de dimerización) y mediante interacción electrostática proteína-RNA (regiones de carga neta local positiva). Las letras P representan los fosfatos, en verde oscuro representan la fosforilación en un tipo de sitio y en rojo en un sitio diferente; el zinc se indica con el símbolo correspondiente en naranja y los dominios de dimerización se indican con una letra D en amarillo. Las esferas lila claro representan las moléculas de proteína N.