



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

***CORRELACIONES ENTRE LAS PROPIEDADES
FISCOQUÍMICAS DE LA SECUENCIA DE AMINOÁCIDOS DE
LA PROTEÍNA DE ENVOLTURA Gp41 DEL VIRUS DEL SIDA
VIH-1 Y SU TROPISMO INFECCIOSO***

TESIS

QUE PARA OBTENER EL GRADO DE:

**MAESTRA EN CIENCIAS
(FÍSICA MÉDICA)**

P R E S E N T A:

EVANGELINA FIGUEROA MEDINA

DIRECTOR DE TESIS:

DR. CARLOS VILLARREAL LUJÁN

COMITÉ TUTORAL:

DRA. LEONOR HUERTA HERNÁNDEZ

DR. GERMINAL COCHO GIL

DRA. MERCEDES RODRÍGUEZ VILLAFUERTE

POSGRADO EN CIENCIAS FÍSICAS

MÈXICO, D. F., A 18 DE JUNIO DE 2007.



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Dedico esta tesis a la memoria de mi madre, Juana Medina Morales. Por la fortaleza que me heredó y su recuerdo que llevo siempre en el corazón. A mi Padre, José Lino Figueroa Barajas, por haberme heredado la habilidad de sumergirme en el mundo de la ciencia. Aún recuerdo aquellos juegos de razonamiento que me impulsaron hacia el gusto por las matemáticas. Con todo el amor de su hija Evangelina.

A mi hermano Edgar, quien me enseñó a leer y escribir.

A mis queridísimas y entrañables hermanas:

Liliana, con quien terminé de crecer gracias a su apoyo incondicional.

Alicia, de quien aprendí con su ejemplo a disfrutar de la vida a pesar de todas sus complicaciones.

Laura, con quien aún a través de la distancia siento su apoyo, cariño y comprensión.

Rosi, mi también amiga y compañera de tertulias, con quien he vivido experiencias inolvidables desde siempre.

A mis amigos:

Polo, mi hermano y amigo incondicional, a quien admiro su lucha y perseverancia. Por nuestro compañerismo y amistad de años.

Mónica, mi hermana de la vida. Gracias por el apoyo y los sabios consejos. Tu amistad es invaluable.

Agradezco profundamente la dirección de mi asesor principal, el Dr. Carlos Villarreal Luján, por su apoyo académico y moral y por todos los conocimientos que me transmitió durante las clases y la realización de esta tesis. A la Dra. Leonor Huerta Hernández, que tuvo la paciencia de enseñarme y guiarme en la parte fundamental de este trabajo: la cuestión biológica. Al Dr. Germinal Cocho Gil, de quien surgieron las geniales ideas que fomentaron esta investigación. A los sinodales, que se tomaron el tiempo de revisar esta tesis y hacerme las correcciones necesarias.

Gracias a todos y cada uno de los profesores han contribuido en mi formación académica. A mis compañeros de la Maestría: Paola, Carla, Karina, Flor y Lenin, ejemplos de dedicación y entereza.

“El físico debe aceptar que el universo físico está gobernado por algún sistema de leyes que puede ser comprendido, incluso aun cuando él no pueda abrigar la esperanza de ser capaz de comprenderlo”

Max Planck

Resumen	9
Introducción.....	11
Capítulo I	13
Nociones de Inmunología y enfermedad asociada al VIH.....	13
<i>I.1 Conceptos básicos de Inmunología.....</i>	<i>13</i>
I.1.1 Constitución del Sistema Inmune	13
I.1.2 Algunas células del Sistema Inmune	14
Monocitos y Macrófagos.....	14
Linfocitos T.....	15
Linfocitos B.....	16
Células Dendríticas (CDs).....	16
I.1.3 Órganos del Sistema Inmune	17
Médula Ósea	18
Timo	18
Órganos Secundarios.....	18
Bazo	19
Ganglios Linfáticos	19
I.1.4 Tráfico Linfocítico.....	19
<i>I.2 Patogenia asociada al VIH</i>	<i>20</i>
<i>I.3 Aspectos terapéuticos.....</i>	<i>22</i>
Capítulo II.....	23
Estructura viral, mecanismos de infección y tropismo infeccioso	23
<i>II.1 Estructura del VIH.....</i>	<i>23</i>
<i>II.2 Proceso de infección.....</i>	<i>24</i>
Capítulo III	29
Variabilidad genética del VIH.....	29
<i>III.1 Clasificación por subtipos</i>	<i>29</i>
<i>III.2 Recombinación genética.....</i>	<i>31</i>
Capítulo IV.....	35
Proteínas de envoltura del VIH-1: gp120 y gp41	35
Capítulo V	41
Fuerzas de unión en moléculas biológicas y estructura de los correceptores utilizados por el VIH	41

V.1 <i>Interacciones No Covalentes</i>	41
Interacciones electrostáticas	41
Puentes de Hidrógeno	42
Interacciones hidrófobas	42
Interacciones de van der Waals.....	43
V.2 <i>Interacciones Covalentes</i>	44
V.3 <i>Quimioquinas y sus receptores: regulación de la respuesta inmune</i>	45
V.5 <i>Función biológica y estructural de los correceptores del VIH</i>	46
Capítulo VI	49
Correlaciones entre el ectodominio de Gp41 y el tropismo infeccioso del VIH-1	49
VI.1 <i>Obtención de las secuencias del ectodominio de gp41</i>	49
VI.2 <i>Construcción de secuencias consenso y alineamientos</i>	50
VI.3 <i>Determinación de sitios variables en la secuencia de aminoácidos de gp41</i>	51
Criterio I	52
Criterio II.....	52
VI.4 <i>Comparación estadística del índice hidropático y la carga entre grupos de virus con distinto tropismo</i>	52
VI.5 <i>Resultados</i>	53
VI.5.1 Gp41: ectodominio	53
VI.5.1.1 Variabilidad de Gp41.....	54
VI.5.1.2 Relación entre el uso del correceptor y la carga e hidropatía de los aminoácidos	54
VI.5.2 Subtipo B.....	56
VI.5.3 Gp120: Lazo V3	56
VI.6 <i>Tablas y Figuras</i>	57
Discusión	79
Conclusiones	81
Apéndice A	83
Proteínas: conceptos básicos	83
A.1 <i>Las hélices alfa y láminas beta: patrones comunes de plegamiento</i>	85
A.2 <i>Glicosilación proteica</i>	86
Apéndice B	89
Pruebas estadísticas	89
B.1 <i>La prueba χ^2</i>	89
B.2 <i>La prueba U de Mann-Whitney</i>	90
BIBLIOGRAFÍA	91

Resumen

La interacción de la glicoproteína de envoltura gp120/gp41 del virus de la inmunodeficiencia humana del tipo 1 (VIH-1) con las moléculas correceptoras en la superficie celular, principalmente CCR5 y CXCR4, tienen implicaciones importantes para el mecanismo de la entrada del virus, el reconocimiento celular específico y la patogénesis asociada. El papel de la subunidad transmembranal gp41 en este proceso no está completamente entendido, ni tampoco la relevancia de la variabilidad en su ectodominio. En este trabajo, se analizaron las propiedades fisicoquímicas de los aminoácidos variables del ectodominio de la proteína viral gp41 proveniente de dos tipos de virus que difieren en su capacidad para infectar dos tipos de células: linfocitos T y macrófagos. El análisis se realiza para un conjunto de 259 secuencias de gp41 obtenidas de la base de datos de Los Álamos (*Los Alamos National Laboratory*). Mediante un análisis entrópico se identificaron sustituciones de aminoácidos no conservativas correlacionadas con el tropismo infeccioso en las posiciones de mayor variabilidad. Encontramos que la presencia de residuos hidrofóbicos o no cargados en dichas posiciones está asociada con el reconocimiento del correceptor CXCR4 (presente mayormente en la superficie de linfocitos T CD4+ no activados), lo cual a su vez, está relacionado con un avance acelerado hacia el SIDA. Por otro lado, la presencia de residuos hidrofílicos o cargados negativamente está asociada con el reconocimiento del correceptor CCR5 (presente mayormente en la superficie de macrófagos, células dendríticas y linfocitos T CD4+ no activados).

Introducción

Desde su descubrimiento en 1981 como una enfermedad mortal, el SIDA (Síndrome de Inmunodeficiencia Adquirida) ha cobrado más de 25 millones de muertes alrededor del mundo, siendo el 2005 el año en que se han registrado más contagios: 5 millones de personas, según el último informe de la Organización de las Naciones Unidas (ONU)¹. De esta cifra, sólo al África subsahariana le corresponden 3.2 millones. En el mismo año, 3 millones de personas fallecieron a causa del SIDA, de los cuales, más de medio millón eran niños.

Aunque África sigue a la cabeza en los casos de contagio y muerte, existen 1.8 millones de latinoamericanos con SIDA, de los que 200.000 se contagiaron en el 2005. En México la tasa de mortandad en el 2004 fue de 4.5 por cada 100.000 habitantes. Se estima que 3 de cada mil personas adultas podrían ser portadoras del VIH en nuestro país. De acuerdo a las estimaciones más recientes del Centro Nacional para la Prevención y Control del SIDA (CENSIDA), en México existen 182 mil personas adultas infectadas por el VIH.

El virus de inmunodeficiencia humana (VIH), causante del SIDA, ha constituido desde su emergencia un enorme desafío a la investigación biomédica. Hasta la fecha, sigue sin entenderse del todo la dinámica que se lleva a cabo en el organismo del huésped para que el virus pueda perturbar el funcionamiento del sistema inmunológico después de haber permanecido latente durante varios años, al grado que éste termina siendo deficiente en su respuesta natural hacia cualquier agente externo. A su vez, tampoco los mecanismos de infección a nivel molecular se han comprendido en su totalidad.

El VIH utiliza una amplia gama de recursos para interferir y evadir la respuesta inmune normal y expresa una alta frecuencia de mutación (~1 por ciclo de replicación), lo que le permite cambiar repetidamente sus determinantes antigénicos. Su entrada en las células blanco requiere de la fusión entre las membranas viral y celular, un proceso mediado por dos proteínas que constituyen la envoltura del virus, en conjunción con receptores celulares. La infección se inicia por la unión de una de las proteínas de envoltura viral, gp120 y gp41, a un receptor de la superficie celular llamado CD4 y a un correceptor transmembranal, CCR5 y CXCR4; dicho evento induce una serie de cambios conformacionales en las proteínas de envoltura, lo que permite iniciar la fusión entre membranas. La comprensión de los pasos biológicos hacia la fusión virus-célula es de suma importancia para elucidar los mecanismos de la entrada viral y el desarrollo de estrategias inmunológicas y farmacológicas hacia la prevención de la infección por el VIH.

En este trabajo se estudia el proceso que rige la infección por el VIH y sus mecanismos, particularmente, aquellos factores fisicoquímicos que determinan el reconocimiento de éste con las diferentes células blanco de nuestro sistema inmunitario. Con este propósito se realiza un análisis estadístico minucioso sobre una de sus dos proteínas de envoltura, llamada gp41, la cual según nuestra hipótesis fundamental, parece jugar un papel bastante importante en el reconocimiento del correceptor celular y por ende, en el tropismo infeccioso hacia diferentes tipos de células.

Capítulo I

Nociones de Inmunología y enfermedad asociada al VIH^{1,2}

1.1 Conceptos básicos de Inmunología

El principio esencial del funcionamiento del sistema inmune es distinguir entre lo propio y lo extraño, y atacar a esto último con la finalidad de intentar eliminarlo del organismo. Las sustancias foráneas al organismo se denominan antígenas.

Nuestro organismo, al igual que el de todos los seres multicelulares, posee una serie de barreras que lo separan del medio externo, y que previenen la entrada de agentes extraños. Éstas constituyen la primera línea de defensa contra la invasión por microorganismos y sustancias foráneas, cumpliendo un papel muy importante, dado que reducen la labor del sistema inmune. La piel es una eficiente barrera físico-mecánica y a la vez química, su integridad impide el paso de los microorganismos del ambiente exterior hacia los tejidos internos. Otras barreras mecánicas están ejemplificadas por la acción lavadora de las lágrimas, la saliva, las secreciones mucosas y la orina (además de sus acciones químicas). Las secreciones ácidas del estómago constituyen una barrera química para los microorganismos que ingresan diariamente al tracto digestivo junto con los alimentos. La mayor parte de nuestras superficies mucosas (conjuntivas, tracto digestivo, respiratorio y urogenital), se encuentran siempre colonizadas por una serie de microorganismos que son considerados como nuestra flora normal en dichos sitios. Estos agentes se establecen gradualmente en sus nichos naturales, a partir del momento en que nacemos, y conforman una barrera biológica para otros microorganismos, que para poder infectarnos a través de las mucosas, deben competir con ellos por el espacio, sitios de adhesión y nutrientes limitados.

Los materiales extraños que superan las barreras, encontrarán inicialmente al sistema inmune innato y éste debe iniciar la acción con su amplia gama de mecanismos de defensa, sin embargo, así como nuestros sistemas de defensa han evolucionado durante millones de años y se han perfeccionado para protegernos de los agentes infecciosos, estos también han evolucionado paralelamente para adquirir numerosos mecanismos que buscan evadir las defensas.

1.1.1 Constitución del Sistema Inmune

Al igual que otros sistemas de nuestro organismo, el sistema inmune está formado por células. Sin embargo, una propiedad que lo distingue de otros, es la capacidad de sus células para movilizarse y circular en forma libre, además de poder agruparse en tejidos y órganos. Su

desplazamiento es fundamental para la eficiencia de sus funciones defensivas, en analogía con un sistema de patrullaje. Además de actuar directamente contra una variedad de antígenos, dichas células pueden liberar una gran variedad de sustancias al medio circundante y a los fluidos corporales, para coordinar las respuestas frente a los antígenos, así como para combatirlos con sus diversos mecanismos efectores.

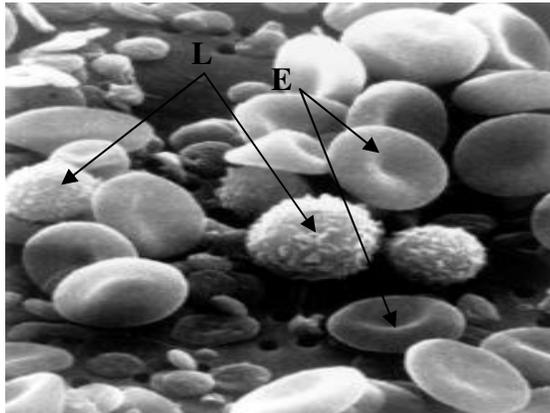


Figura 0. Imagen del sistema circulatorio de un ser humano. Se pueden apreciar eritrocitos (E) y leucocitos (L), también llamados glóbulos rojos y blancos respectivamente. Fuente: es.wikipedia.org/wiki/Sistema_inmunol%C3%B3gico

Los glóbulos blancos en la sangre, o leucocitos, constituyen las células principales del sistema inmune. Normalmente se encuentran en concentraciones de 5000-10000 leucocitos por microlitro (1mm^3) de sangre circulante en el adulto. Sin embargo, esta cifra puede aumentar notablemente cuando se presentan infecciones, u otros estímulos que promueven la producción de células sanguíneas.

Podemos dividir a los leucocitos sanguíneos en dos grupos, con base en su morfología microscópica. Los mononucleares, denominados así por su núcleo único, incluyen a los monocitos y los linfocitos. Por otra parte están los granulocitos, así llamados por su citoplasma de aspecto granular.

I.1.2 Algunas células del Sistema Inmune

Monocitos y Macrófagos

Al igual que todas las células sanguíneas, los monocitos y macrófagos se forman en la médula ósea. Los monocitos se encuentran en el sistema circulatorio y tienen un diámetro de $10\text{-}18\mu\text{m}$, un núcleo no segmentado con cromatina laxa y un citoplasma de granularidad tenue. Estas células se distribuyen rápidamente por los tejidos, en donde se diferencian en macrófagos. El sistema mononuclear fagocítico está compuesto por macrófagos que tapizan gran parte de la microvasculatura en distintos órganos. Los macrófagos son una población celular heterogénea, que puede adquirir características particulares según los tejidos en donde se encuentre y su denominación varía según la localización anatómica.

Las funciones principales de los macrófagos incluyen la fagocitosis (capacidad de ingerir o internalizar partículas y microorganismos, (Fig. 1) junto con la subsiguiente acción

microbicida, así como el procesamiento y presentación de antígenos a los linfocitos T.

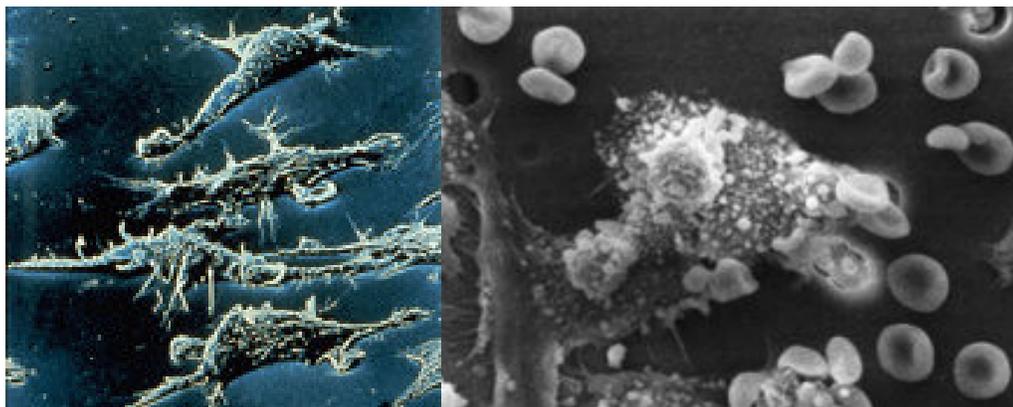


Figura 1. Izquierda: los macrófagos son células que pueden tomar formas muy variadas pudiendo deslizarse en hendiduras muy estrechas y penetrar así en los tejidos más variados. Pueden desarrollar pseudópodos que les permiten desplazarse, capturar, ingerir y destruir los agentes patógenos y otros cuerpos extraños. Una célula dendrítica folicular puede tener un aspecto muy parecido a un macrófago. Derecha: los macrófagos han identificado una célula cancerosa (la grande). Fusionándose con la célula cancerosa, los macrófagos (las células blancas de menor tamaño) inyectarán toxinas que la matarán. Fuente: www.aids-info.ch/s_te/aas-s-imm.htm,

Linfocitos T

Los linfocitos (tanto los T como los B) son los únicos leucocitos capaces de actuar de modo específico en respuesta a los antígenos. Un adulto promedio posee aproximadamente 2×10^{12} linfocitos en todo su organismo, lo cual representa cerca del 2% de su peso corporal. Normalmente, un 20-40% de los leucocitos sanguíneos son linfocitos, y de estos, alrededor del 70-80% corresponden a linfocitos T.

Los linfocitos T se pueden considerar células centrales de nuestro sistema inmune, ya que de ellas depende, en gran medida, el reconocimiento específico de antígenos, la activación de las distintas ramas de la respuesta inmune, y sobretodo, su regulación. Estas células poseen receptores proteicos en su superficie, que les permiten reconocer específicamente a los antígenos (Fig. 2).

Durante su desarrollo, los precursores de los linfocitos T formados en la médula ósea, viajan al timo para completar allí un proceso de maduración y de selección. De ello se deriva su nombre de linfocitos “T” (timo). Este proceso de selección y diferenciación celular en el timo, culmina con la generación de distintas subpoblaciones de linfocitos T, que se especializan en funciones diferentes. Un grupo de linfocitos, al cual distinguimos por la presencia de una proteína de membrana denominada CD4, se conoce como la subpoblación de linfocitos T cooperadores (Th, del inglés *helper*). Este grupo de células Th se encarga, entre otras funciones, de la activación y regulación de otras células a través de la secreción de una serie de moléculas de comunicación intercelular, denominadas colectivamente citocinas. Por otro lado, la subpoblación de los linfocitos T citotóxicos (Tc), reconocible por la presencia de otra proteína de superficie denominada CD8, tiene como función principal la destrucción o lisis de células que han sido alteradas o infectadas por patógenos intracelulares, por ejemplo los virus.

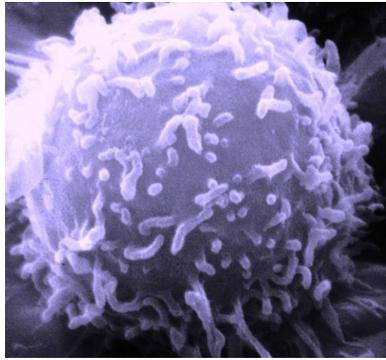


Figura 2. Imagen por microscopía electrónica de un linfocito T humano.

Fuente:

http://commons.wikimedia.org/wiki/Image:SEM_Lymphocyte.jpg

Linfocitos B

Alrededor del 15-20% de los linfocitos sanguíneos corresponden a esta población celular. Los linfocitos B se forman y maduran en la médula ósea, y durante este proceso adquieren la capacidad de expresar receptores en su superficie que les permiten el reconocimiento específico del antígeno. En el caso de los linfocitos B, estos receptores son los anticuerpos (inmunoglobulinas) anclados en su membrana celular. Cuando los linfocitos B se activan, por distintos estímulos, se diferencian gradualmente en células plasmáticas, cuya función es secretar anticuerpos a los fluidos del organismo (Fig. 3). A pesar de la reconocida longevidad de los linfocitos (de meses o incluso años), las células plasmáticas son terminales, con una vida media de apenas unos días.

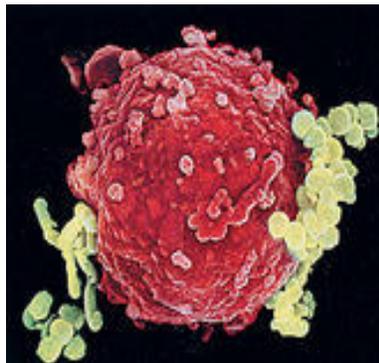


Figura 3. Un linfocito B formando anticuerpos contra bacterias, es decir, racimos de pequeños cuerpos esféricos.

Fuente:

<http://icarito.aconcagua1.copesa.cl/icarito/2002/853/pag2.htm>

Células Dendríticas (CDs)

Las células dendríticas son identificadas como células accesorias de la respuesta inmune. Se las reconoce como las células presentadoras de antígenos por excelencia (CPA). Los diversos tipos de CDs tienen un origen común en la médula ósea diferenciándose luego bajo la influencia de variados estímulos y distribuyéndose en órganos linfáticos y no linfáticos.

Desde los tejidos periféricos migran a los ganglios linfáticos donde presentan el antígeno a los linfocitos T. Dependiendo del micro ambiente expresan diversos marcadores de superficie siendo capaces de secretar citocinas. Éstas son proteínas que regulan la función de las células que las producen u otros tipos de células, son los agentes responsables de la comunicación intercelular, inducen la activación de receptores específicos de membrana, funciones de proliferación y diferenciación celular, quimiotaxis, crecimiento y modulación de la secreción de inmunoglobulinas. Como CPAs cumplen un importante papel en la patogenia de enfermedades autoinmunes y virales destacándose su participación en la infección por el VIH. Se las encuentra en el infiltrado de numerosos cánceres humanos donde actuando como CPA podrían inducir una respuesta inmune antitumoral².

I.1.3 Órganos del Sistema Inmune

Los órganos que componen al sistema inmune se pueden agrupar en dos categorías según sus características funcionales. Los órganos denominados primarios son los responsables de la producción y la diferenciación de los linfocitos T y B. En ellos, los precursores de los linfocitos logran alcanzar su estado maduro e inmunológicamente competente para reconocer y responder ante los antígenos. Por otra parte, los órganos secundarios o periféricos, son aquellos sitios anatómicos que son colonizados por los linfocitos maduros (provenientes de los órganos primarios), en donde se inician y se llevan a cabo, en gran medida, las respuestas inmunes.

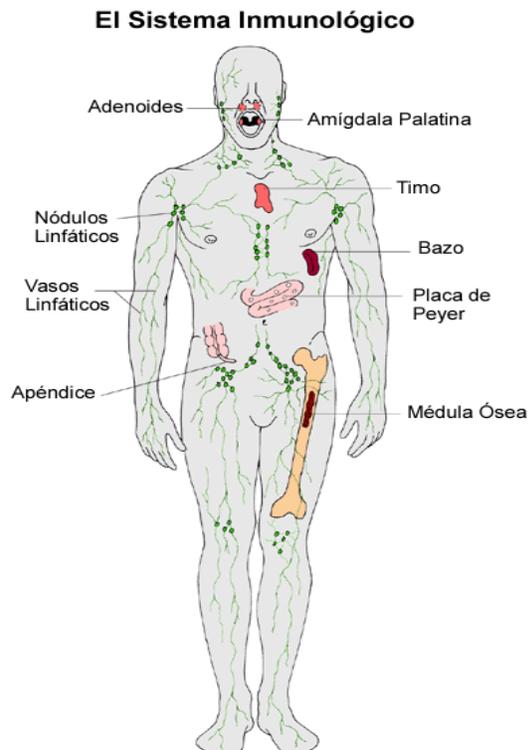


Figura 4. El sistema inmunológico y los órganos que lo componen.

Fuente:

<http://healthcare.utah.edu/healthinfo/spanish/infectious/immune.htm>

Los dos órganos primarios del sistema inmune son la médula ósea y el timo. Los órganos secundarios incluyen fundamentalmente el bazo, los ganglios linfáticos, los nódulos linfáticos y acúmulos linfoides de las mucosas, las placas de Peyer y las amígdalas, entre otras estructuras (Fig. 4).

Médula Ósea

Este tejido se encuentra en el interior de los huesos, y contiene las células precursoras de todos los tipos celulares sanguíneos. El proceso de formación de las células sanguíneas se denomina *hematopoyesis*. En la médula ocurren la proliferación y la diferenciación gradual de las células precursoras pluripotenciales en las distintas líneas especializadas.

Cuando los distintos tipos de leucocitos abandonan la médula ósea para salir a la circulación, ya han alcanzado su estado maduro, con una sola excepción: los linfocitos T. Estos tienen la particularidad de ser liberados como precursores inmaduros que migran hacia el timo, en donde completan su proceso de diferenciación celular.

Timo

Este órgano primario es sin duda central para el sistema inmune, en especial durante las primeras etapas de la vida, dado que es el sitio anatómico en donde ocurre la maduración y diferenciación de los distintos tipos de linfocitos T. Los precursores de los linfocitos T, producidos por la médula ósea, ingresan a este órgano guiados por factores quimioatrayentes como las quimiocinas e inician un proceso de diferenciación que depende de su estrecha interacción con el tejido tímico, el cual les provee de una serie de sustancias indispensables para dichos procesos (factores tímicos). La estructura celular especializada del timo es crucial para la selección y “educación” de dichos precursores, en cuanto a establecer el principio de la distinción entre lo propio y lo no-propio, en un complejo proceso del cual aún se desconocen los detalles. Una evidencia de la selección celular que se desarrolla dentro del timo, es el hecho de que menos de un 10% de los precursores que ingresan al órgano sobreviven y salen como linfocitos T maduros: el 90% restante o más de las células en desarrollo muere durante el proceso.

El timo es un órgano que juega un papel inmunológico primordial en las primeras etapas de la vida, y que gradualmente involuciona y degenera con la edad. Especialmente después de alcanzarse la pubertad, el tejido tímico funcional es reemplazado aceleradamente por tejido graso y tejido conectivo. La ausencia de formación del timo durante el desarrollo embrionario tiene consecuencias mortales para el individuo, que queda a merced de los agentes infecciosos. Sin embargo, la eliminación quirúrgica de este órgano en la vida adulta no parece traer ninguna consecuencia grave.

Órganos Secundarios

Como ya se describió anteriormente, estos órganos incluyen el bazo, los ganglios linfáticos, los nódulos y acúmulos submucosos, las amígdalas, las placas de Peyer y el apéndice, entre otras estructuras anatómicas. Todos estos sitios son colonizados por linfocitos maduros y funcionales (inmunocompetentes), y se ubican estratégicamente en el organismo para defender los posibles sitios de entrada de agentes infecciosos, tales como las superficies

mucosas (respiratoria, oral, intestinal, genitourinaria, conjuntival), así como las vías de circulación sanguínea y linfática. Todos estos órganos son sitios importantes de captación de antígenos, microorganismos, etc. En ellos, los linfocitos T y B, en asociación con otras células del sistema inmune, se activan, proliferan y generan las distintas armas de la respuesta específica, tales como los anticuerpos y los linfocitos T citotóxicos.

Bazo

En el bazo se capturan eficientemente los antígenos que han alcanzado el torrente sanguíneo. La compleja red de pequeños vasos capilares del bazo (sinusoides esplénicos) actúa a modo de un filtro, en el cual las células fagocíticas atrapan a las sustancias extrañas, para iniciar la respuesta inmune específica en las áreas ricas en linfocitos que rodean a las arteriolas, denominadas “pulpa blanca” del bazo. En estas zonas se encuentran aglomeraciones de linfocitos T y B, junto con células del grupo de los macrófagos, en una intensa actividad proliferativa. Estas aglomeraciones, denominadas centros germinales, poseen una organización característica, en la cual la porción central es rica en linfocitos T, mientras la periferia posee principalmente linfocitos B, que se diferencian gradualmente en células plasmáticas, liberando grandes cantidades de anticuerpos a la circulación.

Ganglios Linfáticos

Constituyen un sistema de filtración distribuido a lo largo del conjunto de pequeños vasos que recogen la linfa desde los tejidos, llevándola de nuevo hasta la circulación sanguínea. Para realizar esto, la linfa debe desplazarse por los vasos linfáticos a través de un conjunto de ganglios, dispuestos en series o cadenas, colocados en diversas regiones del organismo. Los ganglios son estructuras encapsuladas pequeñas, de apenas unos cuantos milímetros de grosor, que poseen entradas para los vasos linfáticos aferentes que drenan un determinado tejido, y una o más salidas para los vasos linfáticos eferentes, por los cuales la linfa sale y continúa su recorrido, hasta desembocar finalmente en la circulación sanguínea, a través del conducto torácico.

Cuando ocurre una fuerte estimulación de linfocitos en los ganglios que drenan una determinada región tisular, por ejemplo en respuesta a una infección localizada u otro tipo de lesión, los ganglios se inflaman y aumentan de volumen, pudiendo palparse mediante un examen físico.

I.1.4 Tráfico Linfocítico

Los linfocitos recirculan constantemente entre el sistema sanguíneo y el sistema linfático, lo cual aumenta su probabilidad de encontrar agentes extraños al organismo. En los ganglios linfáticos, los linfocitos pasan de la circulación sanguínea a la linfática, atravesando el endotelio especializado de las vénulas. Al ser vertidos en el sistema de vasos linfáticos, los linfocitos vuelven a la circulación sanguínea a través del conducto torácico, el cual desemboca en la vena subclavia izquierda. La regulación de este tráfico está mediado por quimiocinas y receptores.

El sistema linfático se origina a partir de pequeñísimos vasos ciegos que nacen en los tejidos, y que se agrupan gradualmente en vasos de mayor calibre, que llevan a la linfa y a los

linfocitos a través de los ganglios, para desembocar finalmente en la circulación sanguínea. El mayor vaso linfático del organismo, el conducto torácico, posee un diámetro de apenas unos milímetros. Dado que no existe un órgano de bombeo (como el corazón) en el sistema linfático, el movimiento de la linfa ocurre gracias a conjuntos de válvulas distribuidas a lo largo de los vasos, que actúan unidireccionalmente cuando éstos son presionados por el movimiento muscular, la respiración, etc. Por ejemplo, la linfa viaja más lentamente al dormir o descansar, y más rápidamente durante la actividad física, dadas las diferencias de movimiento muscular en dichas condiciones².

1.2 Patogenia asociada al VIH

La enfermedad causada por el VIH comienza con una infección aguda y evoluciona a una infección progresiva y crónica de los tejidos linfáticos periféricos. El curso de la enfermedad puede seguirse determinando por la cantidad de virus presentes en el plasma del paciente y mediante el recuento de las células T CD4⁺ en la sangre. Los estadios de la infección (Fig. 5) se pueden dividir en cinco etapas principalmente, cada una está asociada a un rango de cuentas de linfocitos T que llevan la molécula CD4 en su superficie (linfocitos CD4+) por cada milímetro cúbico en la sangre del paciente:

- 1 Síndrome agudo (~1000-500)
- 2 Asintomático (~750-200)
- 3 Sintomático temprano (~500-100)
- 4 Sintomático tardío (~200-50)
- 5 Enfermedad avanzada (~50-0)

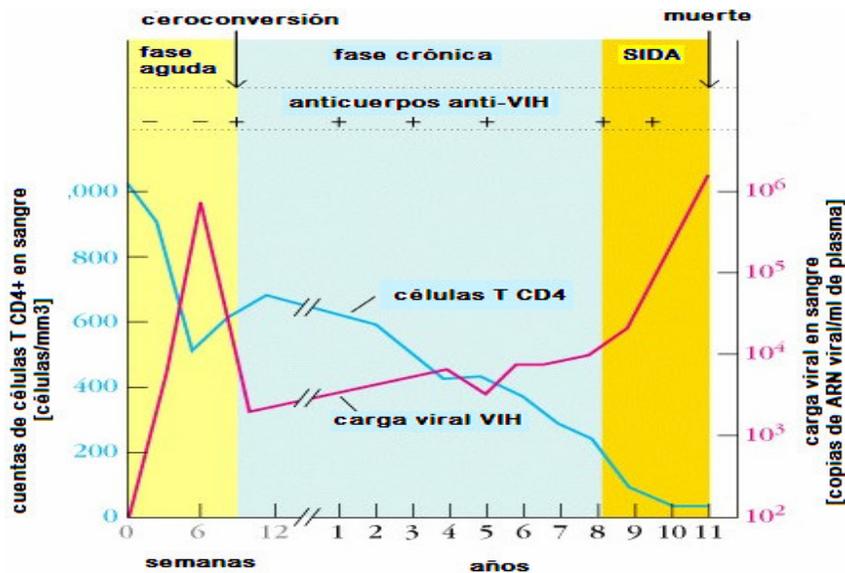


Figura 5. Gráfica de la evolución de la infección por el VIH-1 a lo largo del tiempo: se muestran dos escalas distintas: las primeras 12 semanas y los 11 años después de la infección. Se distinguen la fase aguda, la fase crónica y la etapa del SIDA en amarillo claro, azul y amarillo intenso, respectivamente. Del lado izquierdo y con la línea azul está representado el conteo de células TCD4+ en la sangre; del lado derecho y con la línea morada, la carga viral en la sangre. Fuente: [2].

La etapa final de la infección por el VIH es el SIDA (Síndrome de Inmunodeficiencia Adquirida), caracterizada por una inmunodeficiencia severa que da lugar a múltiples infecciones oportunistas por bacterias, hongos e incluso otros virus.

La infección primaria se produce cuando los viriones presentes en la sangre, el semen u otros líquidos corporales de un individuo entran en las células de otro a través de receptores celulares. Dependiendo del sitio de la exposición inicial al virus, las primeras células infectadas pueden ser, entre otras, las células T CD4⁺ y los monocitos en la sangre o las células T CD4⁺ y los macrófagos en los tejidos mucosos. Es probable que las células dendríticas desempeñen un papel en la diseminación inicial del VIH a los tejidos linfoides debido a que la función normal de estas células consiste en atrapar los antígenos en los epitelios y a continuación migrar a los ganglios linfáticos. Una vez en los tejidos linfáticos, las células dendríticas pueden transferir el VIH a las células T CD4⁺ mediante contacto célula-célula. Días después de la primera exposición al virus, es posible detectar una notable replicación viral en los ganglios linfáticos. Esta replicación ocasiona viremia, durante la cual existen cifras muy elevadas de partículas del VIH en la sangre del paciente. La viremia permite al virus diseminarse por todo el organismo e infectar a las células T colaboradoras, macrófagos y células dendríticas en los tejidos linfoides periféricos. A medida que se disemina la infección, el sistema inmunitario adaptativo pone en marcha respuestas inmunitarias que controlan parcialmente tanto la propia infección como la producción de virus. Este control se refleja en el descenso de la viremia a niveles bajos o indetectables a las 12 semanas de la exposición primaria.

Tras la infección aguda inicial se desarrolla una segunda fase de la enfermedad durante la cual los ganglios linfáticos y el bazo se convierten en sitios de replicación continua del VIH y de destrucción tisular. Durante este período de la enfermedad, el sistema inmunitario permanece competente y es capaz de controlar la mayoría de las infecciones debidas a otros microorganismos, existiendo escasas o nulas manifestaciones clínicas de la infección por el VIH. Por consiguiente, esta fase de la enfermedad a menudo recibe el nombre de período de latencia clínica. Durante esta fase sólo se producen niveles bajos de virus. La mayoría de las células T de la sangre periférica no albergan el virus. No obstante, en los tejidos linfáticos la destrucción de las células T CD4⁺ en los tejidos linfáticos progresa ininterrumpidamente y el número de células T CD4⁺ en la sangre disminuye constantemente.

Durante la progresión crónica de la enfermedad por el VIH, las respuestas inmunitarias del paciente a otras infecciones pueden estimular la producción de VIH y acelerar la destrucción de los tejidos linfoides. La transcripción de los genes del VIH puede potenciarse por medio de estímulos que activan las células T, como los antígenos y diversas citocinas que son especialmente eficaces estimulando la producción de VIH. Por ello, a medida que el sistema inmunitario intenta erradicar otros microorganismos, acelera su propia destrucción.

La enfermedad causada por el VIH progresa hacia la fase final y casi siempre mortal, denominada SIDA, en la que la destrucción del tejido linfático periférico es casi total y el recuento de células T CD4⁺ en la sangre desciende por debajo de 200 células/mm³. La viremia puede ascender de forma espectacular a medida que la replicación viral en otros reservorios se acelera sin control. Los pacientes con SIDA sufren combinaciones de infecciones oportunistas, neoplasias (como el sarcoma de Kaposi, tumor más frecuente en pacientes con SIDA), caquexia (síndrome de desgaste), insuficiencia renal, neuropatía, etc.

A nivel de investigación, la infección por el VIH plantea problemas únicos. Puesto que el virus se replica en células del sistema inmune, la activación inmunológica necesaria para el

control del virus favorece al mismo tiempo la replicación de éste. Actualmente existe una creciente tendencia a otorgar mayor importancia a la fusión celular inducida por el VIH, y a los fenómenos de activación inmunológica en la patogénesis del SIDA.

I.3 Aspectos terapéuticos

La necesidad de un tratamiento efectivo para individuos infectados con VIH ha motivado la búsqueda exhaustiva de blancos moleculares potencialmente útiles para una intervención farmacológica. Entre los procesos virales que no afectan la funcionalidad celular se encuentran: la interacción virus-receptor celular, la transcripción inversa del ARN viral, la integración del provirus en el ADN celular, el procesamiento de sus productos proteicos y la regulación trascricional por los productos Tat y Rev virales.

Debido al conocimiento incompleto de la patogenia del VIH, ha sido difícil predecir la efectividad relativa de inhibidores dirigidos a algunos de estos blancos. Sin embargo, se ha comprobado la predicción inicial, referida a que la inhibición de la replicación viral *in vivo*³ resultaría en una disminución de la carga viral en el paciente. La continua replicación viral y la rápida eliminación de las partículas virales que caracteriza la infección por VIH determinan que incluso una modesta inhibición de la replicación viral reduzca la carga viral y aumente las células CD4⁺, lo que se refleja en un beneficio clínico. Sin embargo la alta tasa de replicación del VIH (10¹¹ partículas por día) y la gran frecuencia de error de la transcriptasa inversa dan como resultado la rápida emergencia de variantes resistentes a los antivirales. Este problema se ha observado con todos los compuestos probados clínicamente hasta la fecha, aunque se ha reducido significativamente gracias a una combinación de inhibidores de transcriptasa reversa y de proteasa viral⁴. Esta estrategia terapéutica reduce considerablemente la carga viral en la mayoría de los pacientes.

1 Richard A. Goldsby, Thomas J. Kindt, Barbara A. Osborne, Janis Kuby; Immunology, fifth edition 2003, McGrawHill.

2 Abul K. Abbas, Andrea H. Lichtman, Immunología Celular y Molecular, quinta edición 2000, Elsevier España.

3 John M. Coffin, HIV Population Dynamics in Vivo: Implications for Genetic Variation, pathogenesis and Therapy, Science 267 1995.

4 E. Crivellato, A. Vacca and D. Ribatti, Trends in Immunology 25, Setting the stage: an anatomist's view of the immune system, 210-217 (2004).

Capítulo II

Estructura viral, mecanismos de infección y tropismo infeccioso

II.1 Estructura del VIH

Una partícula VIH infecciosa está formada por dos cadenas idénticas de ARN empaquetadas dentro de un centro de proteínas virales, y rodeadas de una envoltura formada por una bicapa fosfolipídica derivada de la membrana celular del huésped, pero que incluye proteínas de membrana codificadas por el virus (Fig. 6).

El genoma del VIH contiene los genes *gag*, *pol* y *env* flanqueados por terminaciones largas repetidas (LTR) donde se ubican los elementos reguladores que dirigen la expresión viral (Fig. 7). El gene *gag* (antígenos grupo específico) codifica para productos que forman el esqueleto del virión; el gene *pol* (polimerasa) para las enzimas esenciales de la replicación: transcriptasa reversa, integrasa y una proteasa que procesa los precursores *gag-pol*; el gene *env* (envoltura, Fig. 7) codifica para las glicoproteínas de superficie (SU) **gp120** y transmembranal (TM) **gp41** (Fig. 8).

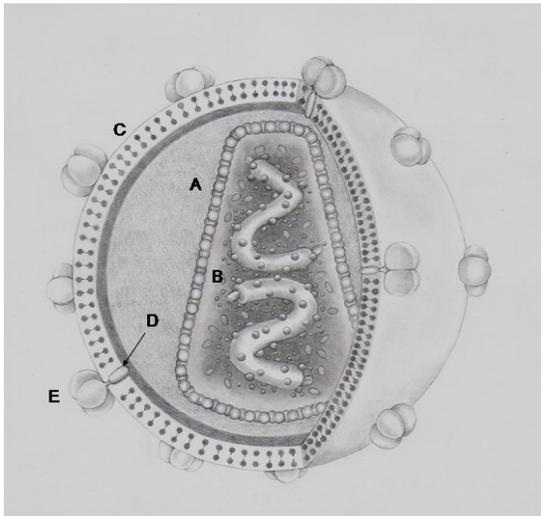


Figura 6. El VIH está constituido por un núcleo con forma de cono truncado en el interior de una envoltura esférica (A). El núcleo está formado por proteínas y en su interior se encuentran dos cadenas de ácido ribonucleico (ARN) que codifican la información genética del virus (B). Por otro lado, la membrana exterior está formada por una bicapa lipídica (C) cuya función es aislar al núcleo viral del medio acuoso exterior. Adosados a la membrana aparecen empotrados, de manera bastante simétrica, receptores proteínicos con una alta afinidad por las moléculas CD4 expresadas en la superficie de los linfocitos T y los macrófagos. El cuerpo del receptor que está enclavado en la membrana viral, y está constituido por la proteína gp41 (D), mientras que la cabeza que sobresale de ella, está cubierta por la proteína gp120 (E), la cual es utilizada por el virus para anclarse en aquellas células por las que tiene afinidad, infectarlas y utilizarlas en su proceso reproductivo.

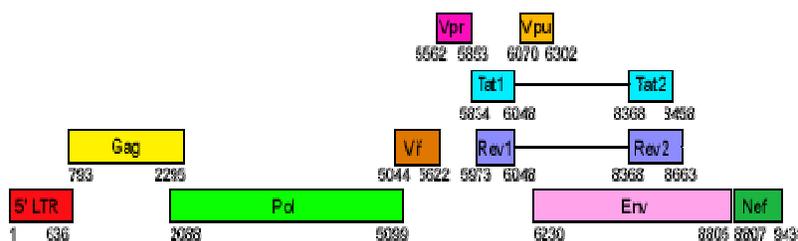


Figura 7. Organización genómica del VIH. Los números debajo de cada rectángulo indican el inicio y el término del gen según la numeración de ácidos nucleicos. Fuente: Los Alamos National Laboratory <http://www-medlib.med.utah.edu/WebPath/TUTORIAL/AIDS.html>.

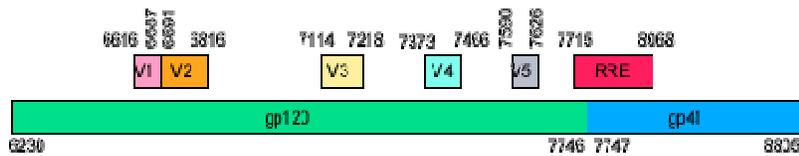


Figura 8. Gene *Env*. Los números indican el inicio y el término de cada proteína a la que codifica y las regiones variables (VI-V5) según la numeración de ácidos nucleicos. Fuente: Los Alamos National Laboratory <http://www-medlib.med.utah.edu/WebPath/TUTORIAL/AIDS.html>.

Las glicoproteínas gp41 y gp120 son los productos virales más superficiales y más antigénicos de la partícula viral, los cuales participan en la unión a los receptores celulares y en la penetración del virus en la célula.

II.2 Proceso de infección

La infección por el VIH se inicia cuando una de las glicoproteínas que componen la envoltura del VIH (*Env*) de una partícula viral se une en la superficie celular a la molécula CD4 y a un correceptor miembro de la familia de receptores naturales de quimiocinas. El complejo *Env* se compone de una subunidad transmembranal, gp41, y una subunidad externa asociada no covalentemente, gp120 (Fig. 8). Estas subunidades se generan mediante escisión proteolítica² de un precursor, gp160.

El complejo *Env* se expresa en forma de estructura trimérica (tres pares) de gp120/gp41. Este complejo media un proceso de múltiples pasos de fusión de la envoltura del virión con la membrana de la célula blanco. Debido a que existen diversas vertientes en dicho proceso, se ahondará más sobre éste en el capítulo IV.

El primer paso en este proceso es la unión de las subunidades gp120 a las moléculas CD4, lo que induce un cambio de configuración que promueve la unión secundaria de las subunidades gp120 al receptor de quimiocinas¹. La unión a este correceptor induce un cambio de configuración en gp41 que expone una región hidrofóbica, denominada péptido de fusión, que se inserta en la membrana celular y promueve la fusión con la membrana viral² (Fig. 9).

Las partículas de VIH, una vez liberadas de una célula infectada pueden unirse e infectar a otra célula. Otra posibilidad es que el virus sea transferido directamente (sin ser liberado) entre dos células. Dado que las proteínas gp120 y gp41 se expresan en la membrana plasmática de las células infectadas, éstas pueden mediar la fusión célula-célula con una célula CD4- no infectada que exprese el correceptor, y entonces los genomas VIH pueden transferirse directamente entre las células fusionadas.

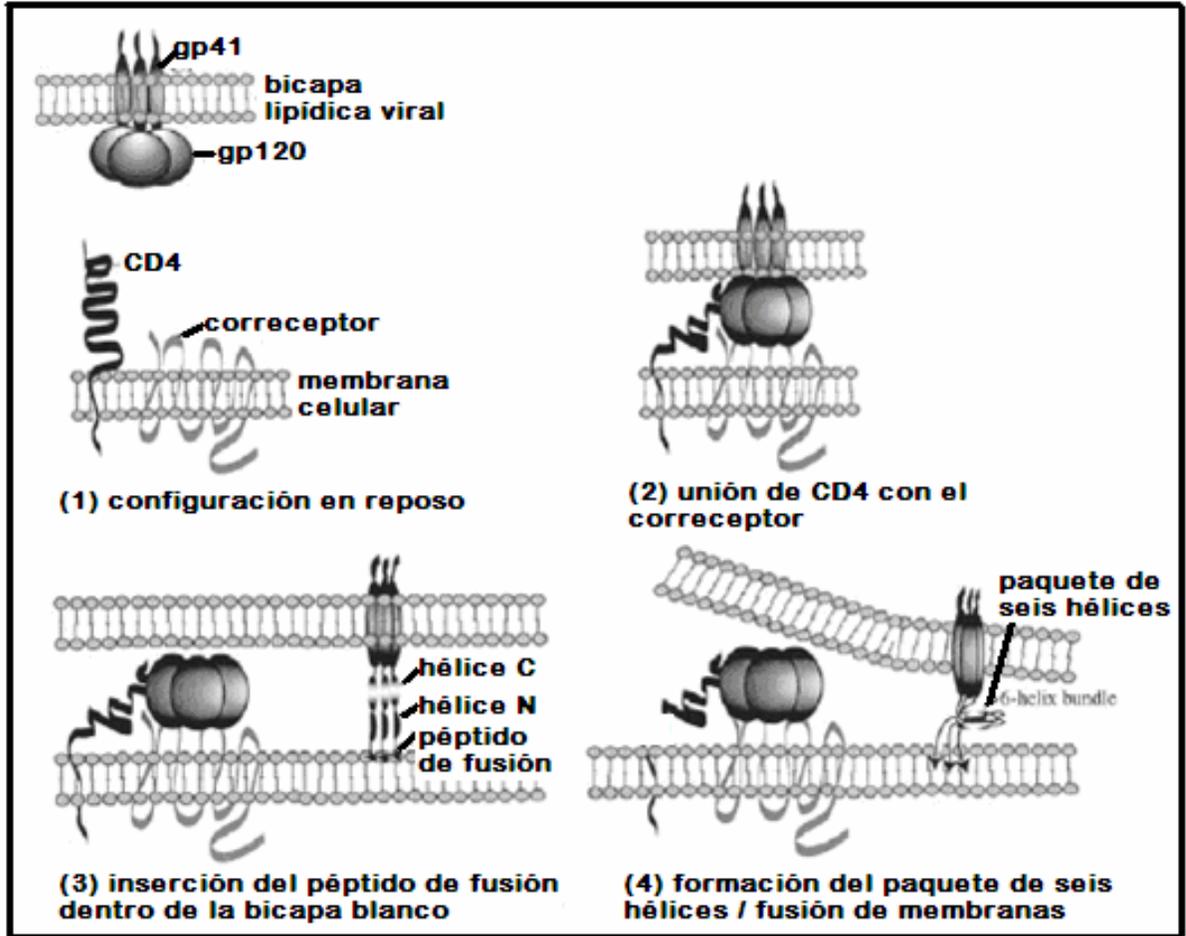


Figura 9. Proceso que permite la fusión entre membranas del VIH y la célula huésped.

La identificación de las moléculas CD4 y de los receptores de quimiocinas (CXCR4 y CCR5) como los correceptores del VIH, se basó en diversos enfoques experimentales y observaciones clínicas^{3,4,5,6}. Se ha demostrado que más de siete receptores de quimiocinas actúan como correceptores para la entrada del VIH en las células, y otras proteínas de la familia de receptores de siete hélices transmembranales acopladas a proteínas G. No obstante, son dos los principales correceptores de los que el virus se vale para infectar: **CCR5** (CC chemokine receptor type 5) y **CXCR4** (CXC chemokine receptor type 4). Esta forma surgió de la clasificación del virus según su afinidad por cada uno, recibiendo el nombre de virus **R5** y **X4** respectivamente, y **R5X4** al que tiene capacidad por reconocer ambos correceptores.

II.3 Clasificación fenotípica del VIH y su tropismo infeccioso

Tanto los receptores de quimiocinas en la célula como las proteínas de envoltura en el VIH-1 juegan un papel principal en la infección, sirviendo como base para la clasificación fenotípica del virus y su tropismo celular.

La noción de que se requiere un correceptor para la entrada del VIH-1 surgió al descubrir que la expresión de CD4 no es suficiente para explicar el tropismo del VIH-1 para diferentes células blanco *in vitro*⁷. Una de las razones para llegar a esta conclusión fue la existencia de distintos tropismos en diferentes aislados VIH-1 para varios tipos de células T

CD4⁺ humanas *in vitro*. Todas las cepas de VIH-1 infectan y se replican en linfocitos T CD4⁺ primarios activados; la distinción del tropismo emerge cuando se examinaron otras células blanco. Algunos aislados mostraban una infectividad eficiente para líneas celulares T CD4⁺ continuas, pero poca infectividad para macrófagos primarios; este fenotipo fue observado originalmente con aislados que habían sido adaptados en el laboratorio para replicarse en líneas celulares T y después con algunos casos clínicos⁸. Tales virus se designaron LCT-trópicos (Línea Celular T-trópico) y generalmente son inductores de sincicios (un sincicio es un tejido plurinucleado sin membranas celulares internas de separación entre los núcleos). Por otro lado, otras cepas mostraron una preferencia opuesta, infectando macrófagos primarios de una manera mucho más eficiente que a las líneas celulares T continuas; a éste tipo se le llamó M-trópico (Macrófago-trópico) o no inductor de sincicios. A los aislados que se replicaban eficientemente en ambos tipos celulares se les denominó dual-trópicos.

Basándose en características de replicación *in vitro* y efectos citopáticos que determinan su fenotipo biológico, las cepas del VIH-1 se pueden clasificar como inductoras (SI=Sincitium Inducing) o no inductoras de sincicios (NSI=Non-Sincitium Inducing).

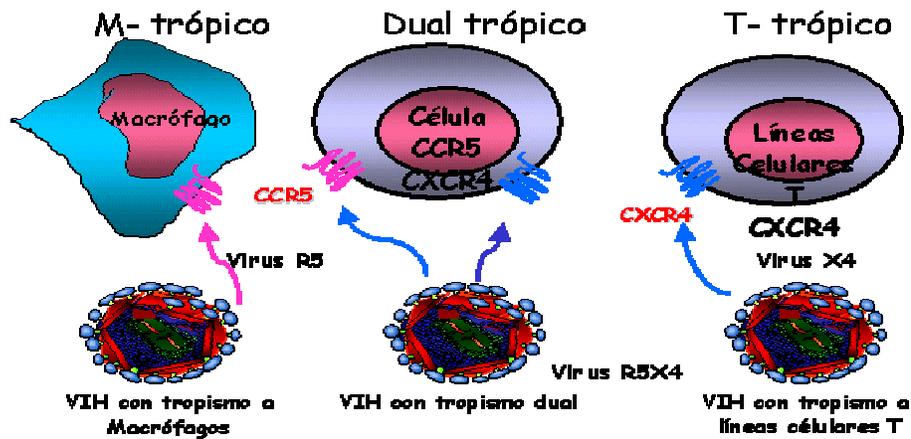


Figura 10. Esquema que representa la clasificación del VIH-1 de acuerdo al fenotipo y el uso del correceptor.

Una diferencia característica entre los dos fenotipos es que las líneas celulares T CD4⁺, incluyendo las MT-2, se pueden infectar fácilmente con el virus SI, pero no con el NSI⁹. Los correceptores principales utilizados por los virus NSI y SI para la infección son los receptores de quimiocinas CCR5 y CXCR4, respectivamente^{10,11}.

Se ha observado que en aproximadamente el 50% de los sujetos infectados por el VIH, existe una evolución desde la detección del virus que utiliza CCR5 (virus R5) en las primeras fases de la infección, al virus que se une a CXCR4 (virus X4) en las fases tardías de la enfermedad^{12,13}. Este cambio está relacionado con un cambio de los virus aislados: en las primeras fases del proceso se aíslan virus con tropismo por los macrófagos, mientras que en las fases tardías se aíslan virus con tropismo por líneas celulares T (Fig. 10).

Existe una clara tendencia hacia la transmisión del virus R5 en la mayoría de los pacientes infectados por el VIH-1 aún cuando en individuos infectados por vía sanguínea debería dominar X4, pues en sangre hay una mayor expresión de CXCR4 debido a patrones preferenciales de expresión de correceptores y sus respectivos ligandos. Por otro lado se ha probado que los virus R5 son los que se establecen en mucosas después de la transmisión

sexual. Sin embargo, análisis epidemiológicos sugieren que la ruta de infección no representa una influencia mayor en la tasa de progresión a la enfermedad, debido a que en individuos infectados por vía sanguínea debería dominar el virus X4 y observarse una rápida progresión al SIDA y a la muerte (como los usuarios de drogas intravenosas, por ejemplo). Lo anterior argumenta la presencia de una barrera que funciona como un bloqueo físico insuperable contra la transmisión del virus X4 pues si éste puede transmitirse con éxito, entonces surge la pregunta del ¿por qué el virus X4 no domina en el “pool” de virus que se amplifica durante la infección primaria del virus, en donde R5 lleva la delantera?¹⁴ Después de todo, (*in vitro*) X4 tiene más células blanco en los cultivos de células linfoides y hay muchas más células T CXCR4⁺⁺ CD4⁺⁺ que T CCR5⁺⁺ CD4⁺⁺ en la sangre. La hipótesis más probable sugiere que los virus R5 (*in vivo*) tienen una ventaja selectiva sobre las cepas X4 debido a su tropismo preferencial por las células dendríticas durante la formación del conjugado célula dendrítica-célula T. Es decir, no es la expresión de CCR5 en la célula dendrítica lo que determina la transmisión preferencial del virus R5, sino la ventaja proveída a la replicación del virus R5 por el conjugado dendrítica-T dentro del tejido linfático, funcionando así como un caballo de Troya para el VIH, sirviendo como vehículo de transporte para poder llegar a las células que expresan CCR5¹⁵. Con lo anterior, vemos que R5 simplemente deja fuera de combate a su contraparte X4 en la carrera por replicarse *in vivo*, una vez que se transmitieron con éxito.

Es bien conocido que los virus R5 son menos citopáticos que los X4 en timocitos en cultivo. Además, niños cuyos virus X4 pudieron ser aislados tienen un daño mayor en la función del timo que aquellos infectados con virus R5. Una cuestión interesante es que el virus de la inmunodeficiencia simia (VIS) nunca usa el correceptor CXCR4 para infectar, todas las cepas utilizan CCR5 y algún otro receptor acoplado a proteínas G¹⁶. Esto es interesante, ya que el síndrome causado por el VIS es mucho menos agresivo que el causado por el VIH.

1 Joanne York and Jack H., Role of Hydrophobic Residues in the Central Ectodomain of gp41 in Maintaining the Association between Human Immunodeficiency Virus Type 1 Envelope Glycoprotein Subunits gp120 and gp41. *J Virology* 2004, vol. 78 no. 9: 4921-4926.

2 W. R. Gallaher, J. M. Ball, R. F. Garry, A. M. Martin-Amedee and R. C. Montelaro, A General Model for the Surface Glycoproteins of HIV and Other Retroviruses, *AIDS Research and human retroviruses* vol. 11, num. 2, 1995.

3 Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, Nagashima KA, Cayanan C, Maddon PJ, Koup RA, Moore JP, Paxton WA. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* 1996, 381: 667-673.

4 Deng HK, Unutmaz D, KewalRamani VN & Littman DR. "Expression cloning of new receptors used by simian and human immunodeficiency viruses". *Nature* 388, 296-300 (1997).

5 Lapham CK, Ouyang J, Chandrasekhar B, Nguyen NY, Dimitrov DS, Golding H. Evidence for cell-surface association between fusin and the CD4-gp120 complex in human cell lines. *Science*. 1996 Oct 25; 274 (5287):602-605.

6 Wu L, LaRosa G, Kassam N, Gordon CJ, Heath H, Ruffing N, Chen H, Humblias J, Samson M, Parmentier M, Moore JP, Mackay CR. 1997. Interaction of chemokine receptor CCR5 with its ligands: multiple domains for HIV-1 gp120 binding and a single domain for chemokine binding. *J Exp Med*. 186 (8): 1373-1381.

7 EA. Berger, HIV entry and tropism: the chemokine receptor connection, *AIDS* 11 (Suppl. A): S3-S16.

8 T. Dragic, An Overview of the determinants of CCR5 and CXCR4 co-receptor function. *J. General Virol.* (2001), 82, 1807-1814.

9 Simmons G, Wilkinson D, Reeves JD, Dittmar MT, Beddows S, Weber J, Carnegie G, Desselberger U, Gray PW, Weiss RA, Clapham PR. Primary, syncytium-inducing human immunodeficiency virus type 1

isolates are dual-tropic and most can use either Lestr or CCR5 as coreceptors for virus entry. *J Virol.* 1996 70 (12): 8355-8360.

10 A. Berger, M. Murphy and M. Farber, Chemokine receptors as HIV-1 coreceptors: Roles in viral Entry, Tropism and Disease, *Annu. Rev. Immunol.* 1999 17, 657-700.

11 Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di Marzio P, Marmon S, Sutton RE, Hill CM, Davis CB, Peiper SC, Schall TJ, Littman DR, Landau NR. Identification of a major co-receptor for primary isolates of HIV-1. *Nature* 1996, 381: 661-666.

12 Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR. 1997. Change in coreceptor use correlates with disease progression in HIV-1 infected individuals. *J Exp Med.* 1997; 185 (4):621-628.

13 Scarlatti G, Tresoldi E, Bjorndal A, Fredriksson R, Colognesi C, Deng HK, Malnati MS, Plebani A, Siccardi AG, Littman DR, Fenyo EM, Lusso P. In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression. *Nat Med.* 1997; 3(11):1259-1265.

14 J. P. Moore, S. G. Kitchen, P. Pugach and J. A. Zack, The CCR5 and CXCR4 Coreceptors- Central to Understanding the Transmission and Pathogenesis of Human Immunodeficiency Virus Type 1 Infection, *AIDS Res. Hum. Retroviruses* 2004 20, 111-126.

15 Cleo G. Anastassopoulou and Leondios G. Kostrikis, Viral Correlatos of HIV-1 Disease, *Current HIV Research* 2005, 3, 113-132.

16 A. Jacobs, J. Sen, L. Rong and M. Caffrey, Alanine Scanning Mutants of the HIV gp41 Loop, *The Journal of Biological Chemistry* 2005, Vol. 280, No. 29, 27284-27288.

Capítulo III

Variabilidad genética del VIH

El VIH es un miembro de la familia de los lentivirus de los retrovirus animales. Los lentivirus son capaces de producir en las células infecciones latentes a largo plazo y efectos citopáticos a corto plazo, y todos ellos producen enfermedades lentamente progresivas y mortales, como la degeneración del sistema nervioso central. Los retrovirus son aquellos virus que contienen su material genético en forma de ARN. Se han identificado dos variantes íntimamente relacionadas de VIH, denominadas VIH-1 y VIH-2. El primero es el más virulento, siendo la causa más frecuente de SIDA y responsable de la pandemia mundial, a diferencia del segundo, que se diferencia en su estructura genómica y en su antigenicidad, caracterizándose por causar un síndrome similar aunque menos patagénico además de permanecer restringido geográficamente en el África Occidental.

III.1 Clasificación por subtipos

Por otro lado, en base a diferencias en distintas regiones del genoma del VIH-1, éste se clasifica en subtipos los cuales dependen un tanto de las razas así como de la distribución geográfica (Fig. 11). Los subtipos más conocidos del VIH-1 son A, B, C, D, E, F, G, H y O. La diferencia principal de los subtipos es su composición genética con divergencias del orden del 30-40% en el gen de envoltura (*env*) y de un 12-15% en el gene de la coraza (*gag*)^{1,2}.

En la Figura 11 se muestra una representación esquemática de la filogenia de los subtipos del VIH y el VIS (Virus de la Inmunodeficiencia en Simios). Los genomas del VIH-1 y VIH-2 tienen una similitud de entre 40 y 50%, mientras que el VIH-2 presenta una homología del 75% con el VIS. Estos subtipos del VIH-1 varían con la distribución geográfica y están distribuidos de un modo diferente por todo el mundo, si bien es cierto que uno de los subtipos suele ser el preponderante en un área determinada, en África, donde se piensa que la infección es más antigua, coexisten prácticamente todos los subtipos con predominio del A y el C, mientras que los D, E, F, G, H y O tienen un predominio moderado (Fig. 12).

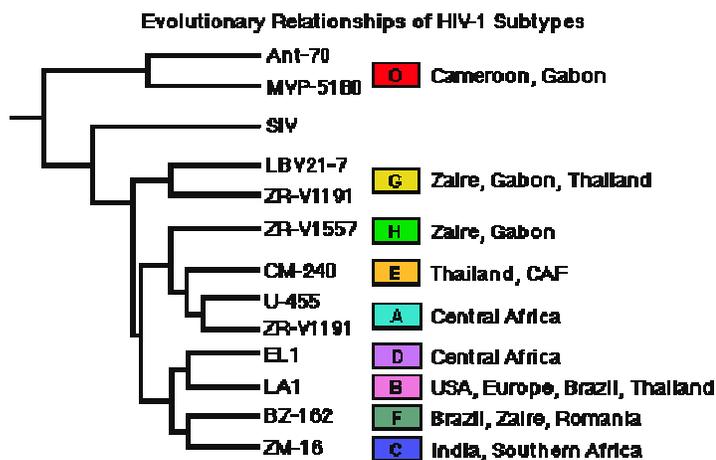


Figura 11. Esquema sobre la relación filogenética que guardan los distintos subtipos del VIH-1 y el VIS. Fuente: <http://www-Medlib.med.utah.edu/WebPath/TUTORIAL/AIDS/AIDS009.html>

En Europa y el continente americano predomina el subtipo B; sin embargo, mientras que el predominio de los subtipos A, C, D, E y O es muy bajo en América del Norte, o el subtipo F en América del Sur, en Europa Occidental tienen un predominio moderado y sólo el subtipo O lo tiene muy bajo. En el sudeste de Asia el subtipo predominante es el E y en el sur de este continente existe preponderancia del subtipo C, mientras que la presencia del B sólo es moderada. Se han descrito subtipos F en Brasil y Rumania, G y H en Rusia y África Central, I en Chipre u O en Camerún pero prevalecen muy poco.

La diversidad genética de los subtipos del VIH está dada por dos acontecimientos muy importantes que son mutación y recombinación (ver siguiente sección). Ambas se manifiestan a lo largo de un genoma viral de 10,000 pares de bases. La enorme diversidad genética del VIH-1 puede ser atribuida a una extensión sin precedentes en la replicación viral, y a la natural tendencia al error que existe en el proceso de la transcripción inversa³. Esta variabilidad es notable, no sólo entre individuos infectados (variabilidad inter-huésped) sino también dentro de cada individuo infectado (variación intra-huésped) tanto con el paso del tiempo como entre distintos compartimentos. Los tipos celulares o tejidos con un flujo de virus restringido entre ellos se llaman compartimentos virológicos y se pueden diferenciar de los reservorios virológicos (que implican una supresión relativa en la replicación viral entre tipos celulares o tejidos), usando un criterio filogenético. La variación extensiva del VIH, con la que provee las cuasi-especies del virus con la capacidad de ajustarse a diversas presiones de selección como la inmunológica o farmacológica, y tienen su origen principalmente de la infidelidad inherente de la enzima de la transcriptasa inversa en conjunción con la falta de mecanismos de prueba, la gran movilidad de viriones y una tendencia hacia la recombinación.

Estas agrupaciones filogenéticas de las cepas de VIH-1, subtipos y formas recombinantes son importantes porque sirven como registro histórico de las epidemias de VIH-1 habidas, resaltando además las diferencias en la distribución geográfica y las áreas donde las distintas variantes son endémicas.

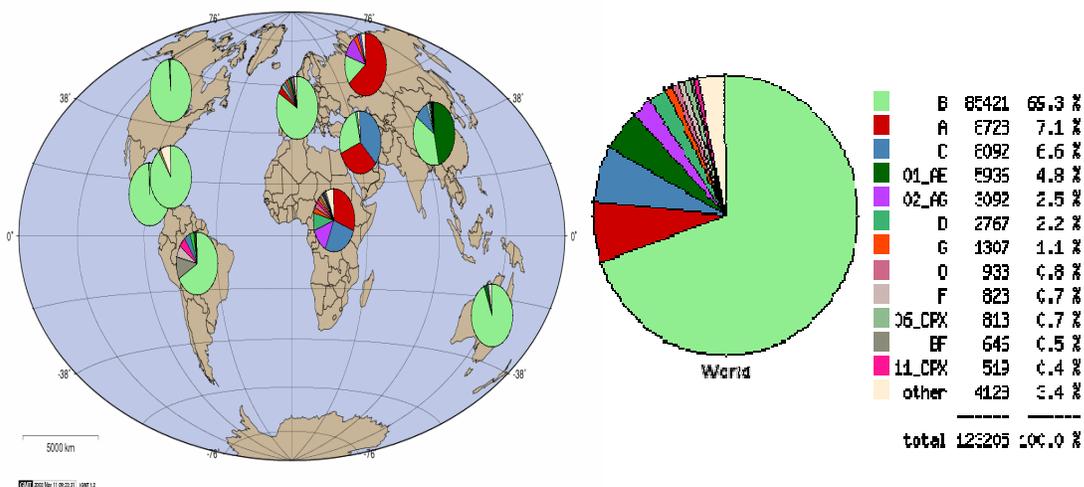


Figura 12. Distribución de subtipos de las secuencias VIH-1 colectadas en el mundo. Fuente: Los Alamos Database http://www.hiv.lanl.gov/components/hiv-db/new_geography/geography.comp

III.2 Recombinación genética

La recombinación desempeña un papel fundamental a la hora de generar variabilidad genética en el VIH-1, permitiendo al virus escapar de vacunas y/o medicamentos.

Para que la recombinación genere nueva variabilidad, una célula ha de ser infectada por dos provirus diferentes, ya sea simultánea (infección dual o múltiple) o secuencialmente (superinfección), de manera que se produzca el encapsulado de un transcrito de ARN de cada provirus en un mismo virión heterocigoto. Debido a su forma de operar, transitando entre distintas hebras, la enzima transcriptasa inversa (que copia el ARN viral en DNA celular) favorece una elevada frecuencia de recombinación. Después de la infección de una nueva célula, la transcriptasa inversa generará una nueva molécula de ADN retroviral que será recombinante de los dos genomas parentales. En individuos que hayan sido infectados por virus genéticamente diferentes se pueden generar así combinaciones genómicas originales y muy diferenciadas.

Por tanto, para que se genere una nueva partícula vírica recombinante portadora de nuevas combinaciones genéticas deben ocurrir dos procesos diferentes, en primer lugar una célula debe ser co-infectada por al menos dos partículas víricas y en segundo lugar, durante el proceso de replicación del virus la enzima transcriptasa inversa debe copiar ADN procedente de los distintos virus en una misma secuencia, formando un mosaico genético. Este último proceso sucede de manera frecuente debido al modo en que ocurre la replicación.

En los últimos años se ha intensificado el interés por descubrir nuevas formas recombinantes del VIH-1 (CRF, de sus siglas en inglés *Circulating Recombinants Forms*), aunque muchas de las nuevas CRFs no se han publicado. La Figura 14 muestra un árbol filogenético reconstruido, inferido del gene *env* basado en alineamientos de todas las secuencias no recombinantes⁴, donde se puede observar la divergencia entre subtipos; a diferencia de la Figura 11, ésta trata sólo de la diversidad del gene *env*, no del genoma completo.

En el año 2000 se realizó un análisis⁵ a partir de un monitoreo sobre la dinámica de la dispersión global del VIH-1, y dio como resultado que la proporción más grande de todas las infecciones por VIH-1 en todo el mundo y en ese año fue debido al subtipo C (47.2%), mientras que el subtipo A/CRF02_AG constituyó la segunda causa de todas las infecciones (27%), seguido del subtipo B (12.3%). En el mismo análisis se confirmó que el papel de los CRF es importante en la pandemia y su impacto en ésta va incrementando con el tiempo. Los efectos de la recombinación y superinfección con múltiples cepas o subtipos del VIH y VIS se han revisado en detalle⁶. Sin embargo, la evolución de una segunda generación, un recombinante interCRF entre CRF07_BC y CRF08_BC, fue descubierta meses antes en la provincia Yunnan de China⁷, justificando de ese modo la caracterización del área del sudeste de Asia que incluye el oeste de Yunnan y Burma como un “crisol de culturas” de recombinación para el VIH-1⁸.

Es importante notar que tales hallazgos pueden reflejar una ventaja selectiva de nuevas formas recombinantes, alguna de las cuales exhiben estructuras de un virión pseudo tipo con las secciones externas de sus glicoproteínas de envoltura intercambiadas con cepas de distintos linajes sobre cepas paternas más simples como ya se observó antes⁹, colocando así un reto adicional para los investigadores en el desarrollo de vacunas y terapias efectivas.

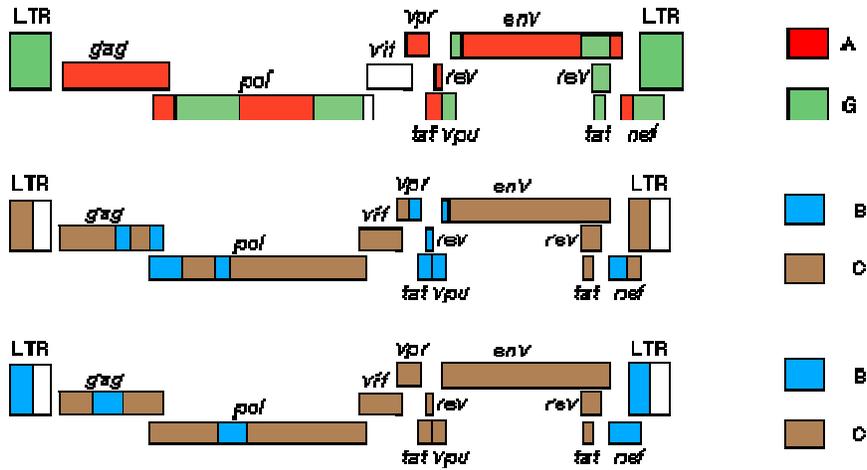
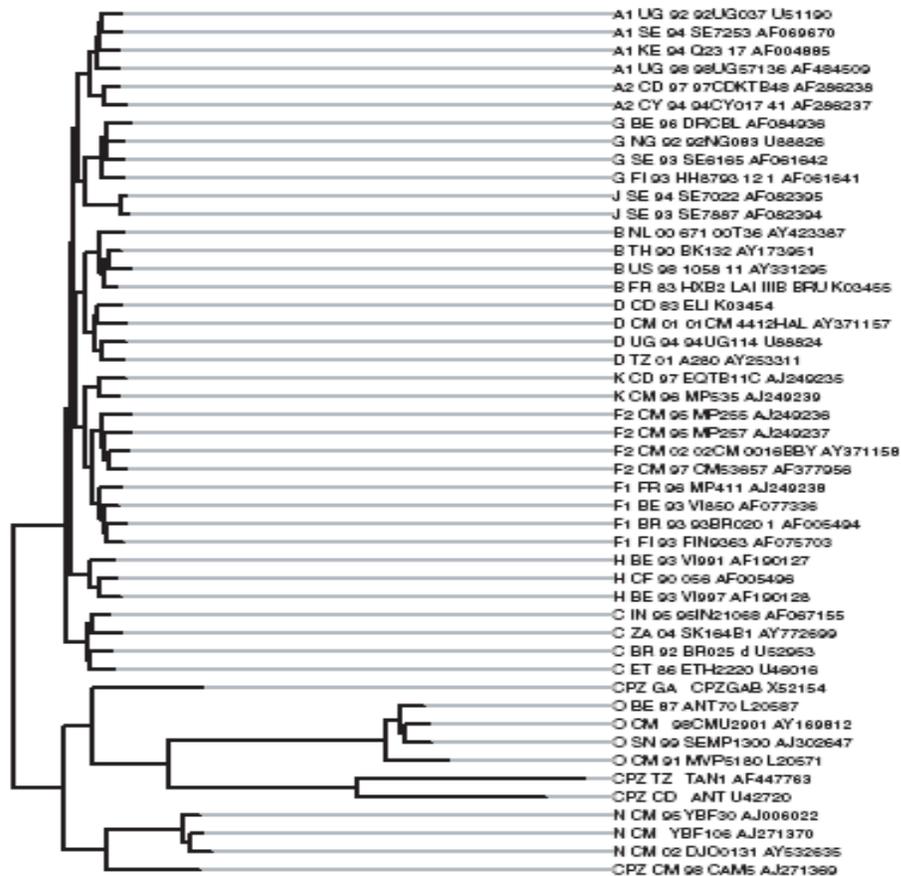


Figura 13. Formas recombinantes recirculantes (CRFs). Paneles superior, medio e inferior: CRF01_AG, CRF07_BC y CRF08_BC respectivamente. Cada color ilustra las regiones del genoma del VIH-1 que pertenecen a cada subtipo (indicado a la derecha). Nótese que incluso dentro de cada gene existen recombinaciones. Las regiones en blanco indican aquellas de las que no se conoce su procedencia. Fuente: <http://www.hiv.lanl.gov/content/hiv-db/CRFs/CRFs.html>



env

Figura 14. Árbol filogenético del gene *env* de secuencias de referencia no recombinantes del VIH-1. Fuente: Los Alamos National Laboratory, <http://hiv.lanl.gov/>.

Por otro lado, en diversos estudios experimentales se ha encontrado una capacidad similar de replicación del VIH-1 dentro de las células del linaje de los monocitos entre diferentes subtipos genéticos. En estas células las clonas NSI se replican más eficientemente que las SI. La emergencia del genotipo SI está asociada con un decline acelerado de células T CD4+ y una progresión más rápida hacia el SIDA en sujetos infectados por el VIH-1 hospedando no sólo el subtipo B sino también otros subtipos genéticos tales como el E e incluso el recombinante CRF02_AG, que prevalece en el centro y oeste de África, como se describió recientemente. En niños infectados por el VIH-1, supuestamente con el subtipo B, también se ha asociado el cambio en el uso del correceptor CCR5 en CXCR4 con la progresión acelerada a la enfermedad; más aún, ya se ha mostrado una asociación significativa entre la disminución de los niveles de CD4+ y la severidad de la enfermedad, aunque es interesante observar que el cambio en el uso del correceptor apareció meses o incluso años después del comienzo del deterioro inmunológico en pacientes pediátricos infectados con subtipos A, D o CRF01_AE, sugiriendo así que los virus X4 pueden emerger como una posible consecuencia de la deficiencia inmune.

1 Burke, D. S. and F. E. McCutchan. 1997. Global distribution of human immunodeficiency virus type 1 clades. *AIDS: Biology, Diagnostics, Treatment and Prevention*. Ed. Lippincott-Raven Publishers, New York. 119 – 126.

2 Crandall K.A., D.A. Vasco, D. Posada and H. Imamichi. *Advances in understanding the evolution of HIV*. *AIDS* 1999. 13 (suppl): S39 – S47.

3 Roberts, J. D., K. Bedenek, and T. A. Kunkel. 1988. The accuracy of reverse transcription from HIV-1. *Science*. 242:1171 - 1173.

4 T. Leitner, B. Korber, M. Daniels, C. Calef, F. Foley, Subtypes & CRF's, Los Alamos National Laboratory, Los Alamos <http://hiv.lanl.gov/>

5 Osmanov S, Pattou C, Walter N, Schwardlander B and Esparza J. *J Acquir Immune Defic Syndr* 2002. 29: 184-190.

6 Blackard JT, Renjifo B, Fawzi W, Hertzmark E, Msamanga G, Mwakagile D, Hunter D, Spiegelman D, Sharghi N, Kagoma C and Essex M. *Virology* 2001. 287: 261-265.

7 Yang R, Kusagawa S, Zhang C, Xia X, Ben K and Takebe Y. *J Virol* 2003.77: 685-695.

8 Takebe Y, Motomura K, Tatsumi M, Lwin HH, Zaw M and Kusagawa S. *Aids* 2003. 17: 2077-2087.

9 Delgado E, Thomson MM, Villahermosa ML, Sierra M, Ocampo A, Miralles C, Rodríguez-Pérez R, Diz-Aren J, Ojea-de Castro R, Losada E, Cuevas MT, Vazquez-d Parga E, Carmona R, Perez-Álvarez L, Medrano L, Cuevas L, Tabeada JA and Najera R. *J Acquir Immune Defic Syndr* 2002. 29:536-543.

Capítulo IV

Proteínas de envoltura del VIH-1: gp120 y gp41

El complejo de glicoproteínas de la envoltura del VIH (Env) consiste de dos subunidades asociadas no covalentemente, gp120 y gp41, formando una especie de espiga trimérica (en realidad cada espiga está formada por tres subunidades idénticas o monómeros de gp41 y tres gp120, Fig. 15) que sobresale de la membrana viral. Los viriones del VIH-1 contienen alrededor de 14 espigas por partícula. Aún no existe evidencia de un espaciamiento periódico de sus espigas pero la distancia más pequeña de separación que se ha encontrado entre éstas es de $\sim 15\text{nm}$ ¹. Además se ha encontrado una tendencia a formar cúmulos de espigas de al menos siete miembros separadas por distancias menores a 30nm y que un aumento en el número de espigas por virión está correlacionado con un incremento en la infectividad, por lo que la eficiencia en la fusión podría depender de la densidad local de espigas.

Gp41 y gp120 son sintetizadas inicialmente como un precursor único (gp160), el cual es procesado por una proteasa celular. La gp120 es la proteína más externa y contiene los sitios específicos necesarios para la unión del virus a las células blanco. La gp41 es transmembranal y contiene las estructuras requeridas para la oligomerización del complejo, su fijación a la membrana viral y los dominios requeridos para la fusión de las membranas viral y celular.

El mecanismo molecular de la fusión virus-célula y la estructura del complejo gp120/gp41 han sido objeto de una investigación intensa^{2,3}, de la cual se ha derivado un modelo general de fusión, en el que una estructura altamente organizada del complejo gp120/gp41 atraviesa por una secuencia de complejos cambios conformacionales (Fig. 16) y a su vez tolera una alta variabilidad en su secuencia de aminoácidos. El correceptor se une a una región constante entre la base del rizo V1 y V2 de gp120 conocida como el “bridging sheet”⁴ (Fig. 18).

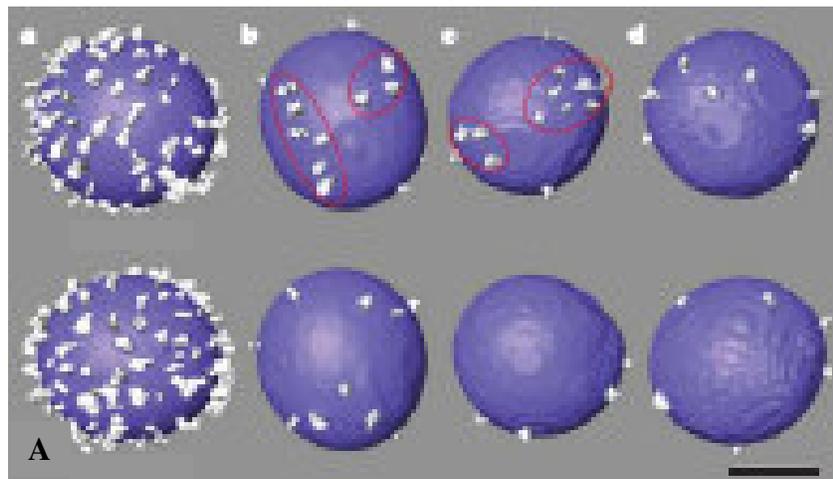


Figura 15A. Espigas del VIS (abajo) y VIH (arriba), se pueden apreciar algunos cúmulos delimitados con rojo, los cuales no se observan en el virus de inmunodeficiencia de los simios. Longitud barra negra = 50nm. Fuente: [31].

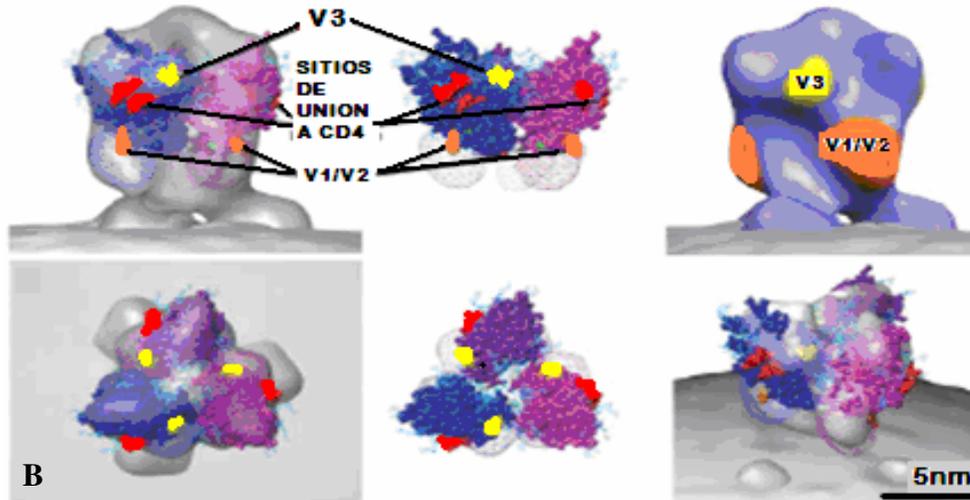


Figura 15B. Morfología en 3D y dimensiones de las espigas, donde se observan las regiones sobresalen en la asociación de las dos proteínas gp120/Gp41, como las regiones supervariables supervariables expuestas (V3 en amarillo y V1/V2 en naranja) de gp120 y la localización de los sitios de unión a CD4 (rojo). Gp41 queda cubierta por gp120 y sólo se logra observar la región transmembranal como una especie de pies sobre la membrana viral. Fuente: [31].

La interacción correceptor-Env está fuertemente influenciada por el rizo adyacente V3 constituido por una cadena de 36 aminoácidos. La estructura del complejo monomérico gp120 con CD4 indica que la unión a esta molécula induce el que gp120 exponga el sitio de unión para el correceptor, una región muy conservada situada en su centro⁵.

La interacción gp120-CD4 induce un cambio conformacional en Gp41, la cual se transforma en una estructura extendida transitoria e inestable, en donde la formación de un trímero enrollado sobre su parte N-terminal ayuda a que el extremo amino terminal (péptido de fusión) se exponga, favoreciendo su inserción dentro de la membrana blanco^{6,7}. Finalmente, el paquete de tres C-hélices enrolladas antiparalelas de gp41 forma una estructura compacta compuesta de tres hélices N y tres hélices C (paquete de seis hélices o también llamado “hairpin”)⁸. Este drástico cambio conformacional parece proveer la energía libre para la fusión de las membranas celular y viral.

Aunque los modelos estructurales tratan de definir una región crucial para la interacción de *Env* con la molécula correceptora, estudios funcionales indican que la función del correceptor es bastante compleja^{9,10}. La interacción del “bridging sheet” con el correceptor está fuertemente influenciada por el lazo adyacente V3¹¹. En este caso la presencia de residuos cargados positivamente (R o K) en la posición 11 y 25 de V3 (ver Fig. 17 y 18) en el capítulo) está asociada con el fenotipo inductor de sincicios (X4), mientras que las secuencias que no presentan esas sustituciones están asociadas a las cepas no inductoras de sincicios (R5). A esta regla se le conoce como 11KR/25K¹². Sin embargo, tales condiciones no son siempre suficientes para conferir un fenotipo particular. Además, estudios realizados con anticuerpos monoclonales anti-CCR5 han demostrado que un dominio en el lazo extracelular 2 (ECL2) de CCR5, críticamente importante para la función del correceptor durante la infección, no participa en la unión con gp120¹³. Otros estudios indican se requiere de una estructura compleja que involucra todas las regiones extracelulares del correceptor para la interacción

con *Env*¹⁴. Tales observaciones sugieren que los correceptores están involucrados en otros pasos del proceso de fusión entre membranas, además de la función de unirse a gp120. Con estos antecedentes, resulta razonable suponer que sustituciones localizadas en otras regiones hipervariables de *Env* puedan alterar el mecanismo de la determinación del fenotipo/correceptor, ya sea por medio de la misma gp120 o en el ectodominio de Gp41 (Gp41-e).

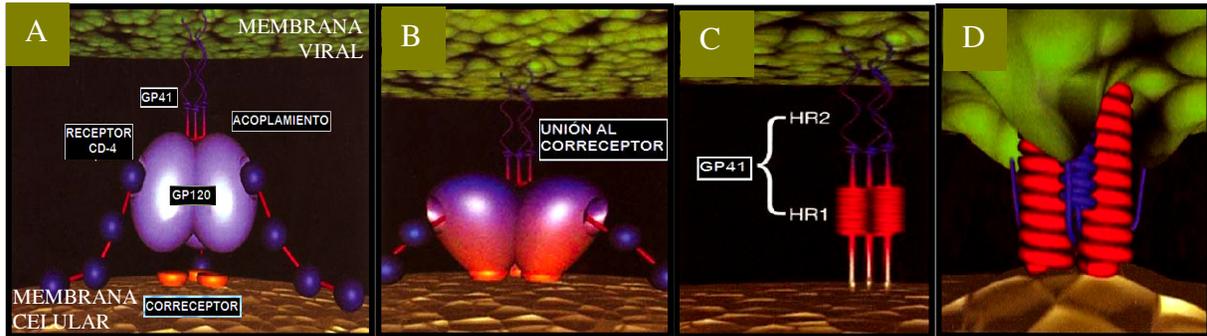


Figura 16. Esquema del VIH con sus proteínas de infección: Gp41 atravesando la membrana viral, asociada a gp120 y la membrana de una célula blanco que expresa la molécula CD4 y el correceptor transmembranal determinante para que el virus infecte. En cada recuadro se ilustran los pasos más probables que rigen el proceso de fusión. A) Acoplamiento entre *Env* y la molécula CD4, abajo se aprecia el correceptor en color naranja en la membrana celular B) Reconocimiento de gp120 con el correceptor donde se intenta ilustrar el cambio de conformación en ésta proteína durante este paso. C) Liberación de gp41 y su péptido de fusión (HR1: parte amino terminal en rojo), la parte morada indica la parte carboxiterminal (HR2). D) Acercamiento en la formación del paquete de seis hélices (“o hairpin”), donde las N (rojas) quedan cubiertas por las hélices C (moradas) que se encuentran más expuestas. En las Figuras 18, 19 y 20 se encuentran los esquemas tridimensionales de ambas glicoproteínas de envoltura del virus y del receptor CD4.

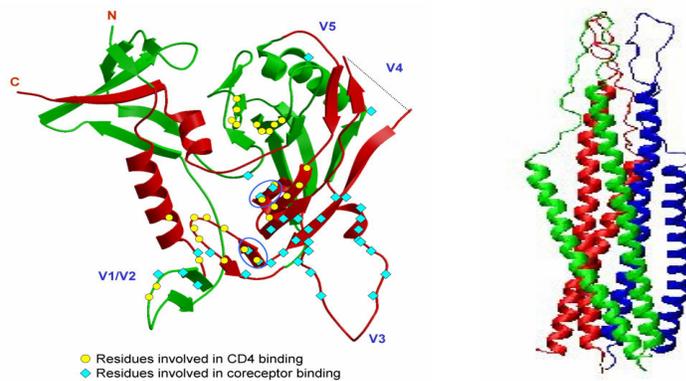


Figura 17. Modelos estructurales de cinta para las proteínas gp120 y gp41 respectivamente. Izquierda: existe una localización de los sitios de unión de esta proteína a CD4 (puntos amarillos) y al correceptor (puntos azules). Derecha: representación tridimensional del trímero gp41. Los colores son para diferenciar en la localización espacial entre un monómero y otro idéntico, nótese el antiparalelismo entre hélices, existe una cierta inclinación.

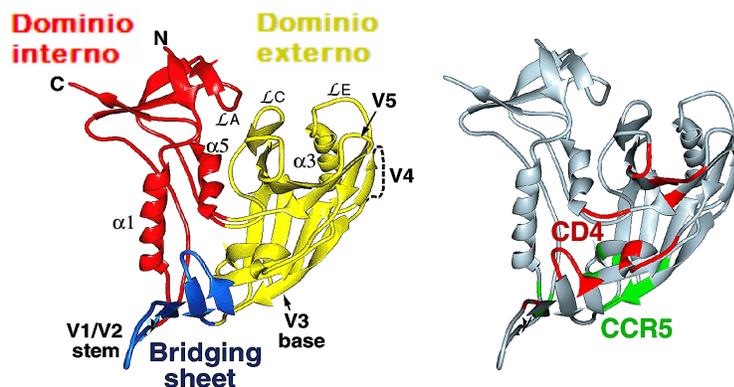


Figura 18. Izquierda: Modelo estructural de cinta de gp120. En rojo se muestra el dominio interno (la parte que queda hacia dentro en la conformación del trímero) y en amarillo el externo, donde se encuentra el lazo V3 en la parte de abajo, V4 y V5 arriba. En azul el llamado “bridging sheet” y las regiones V1/V2. Derecha: mismo modelo pero ahora se resaltan las regiones que interactúan con el receptor CD4 y el correceptor CCR5. Nótese que ésta última coincide con la localización del lazo V3.

Interacciones entre las hélices N y C de Gp41

Péptidos sintéticos correspondientes a la secuencia de las hélices C y N (péptidos C4 y N36) se auto asocian en solución, formando una estructura compacta como la mostrada en la Figura 20.

El paquete consta de tres hélices C34 orientadas en antiparalela con tres hélices de N36¹⁶. Éstas interactúan con las N36 principalmente a través de residuos hidrofóbicos en tres ranuras sobre la superficie del trímero central. Comparaciones de secuencias de Gp41 del VIH contra secuencias del VIS muestran que los residuos que se alinean en esas ranuras son muy conservados entre los dos virus, mientras que los residuos en N36 que se encuentran a costados de las C34 son muy divergentes entre el VIH y el VIS¹⁷. En el diagrama de la Figura 19 los residuos en C34 se muestran como elipses para indicar la inclinación oblicua de las hélices C34 relativa a las N36 y para enfatizar que C34 no es parte de la “rosca” o “carrete” que se forma al centro. Los residuos en las posiciones *e* y *g* de las hélices N36 caen en la parte externa del carrete central y dispuestas dentro de un espacio interhelicoidal triangular entre dos hélices N37 y una hélice C34 como apoyo. En general, los residuos en las posiciones *a* y *d* del paquete C34 están enfrente de los residuos en las posiciones *e* y *g* de las N36, aunque puede haber contacto con otras posiciones.

Comparaciones entre Gp41 de VIH y VIS, no existen cambios no-conservativos en *e* y *g* de N36 y sólo dos ocurren en las posiciones *a* y *d* de C34. Por el contrario, ocurren 8 de 9 cambios no-conservativos en la hélice N36 en las posiciones expuestas *f*, *b* y *c*, mientras que 13 de 15 cambios no-conservativos ocurrieron en la hélice C34 exceptuando *a* y *d*.

La secuencia de la región N de Gp41 está entre las más conservadas dentro de toda la envoltura del VIH, lo cual es probable que resulte de una presión selectiva en las posiciones *e* y *g* con el fin de retener las interacciones con el péptido C34, así como en las posiciones *a* y *d* para mantener interacciones triméricas en el carrete central.

Cada uno de los surcos en la superficie del trímero N36 tiene una profundidad particular. Esta cavidad es relativamente grande (largo $\sim 16\text{\AA}$, ancho $\sim 7\text{\AA}$ y profundidad 5-6 \AA) y acomoda tres residuos hidrofóbicos de la hélice C34: Ile-635, Trp-631 y Trp-628.

Estudios anteriores encontraron que mutaciones en sitios que estabilizan esta estructura inhiben la fusión de membranas. Esto sugiere que las interacciones entre N36 y C34 son críticas para que se lleve a cabo la fusión y por lo tanto la infección. Por lo que se considera que el paquete mostrado en la Figura 20 representa el estado fusogénico de Gp41.

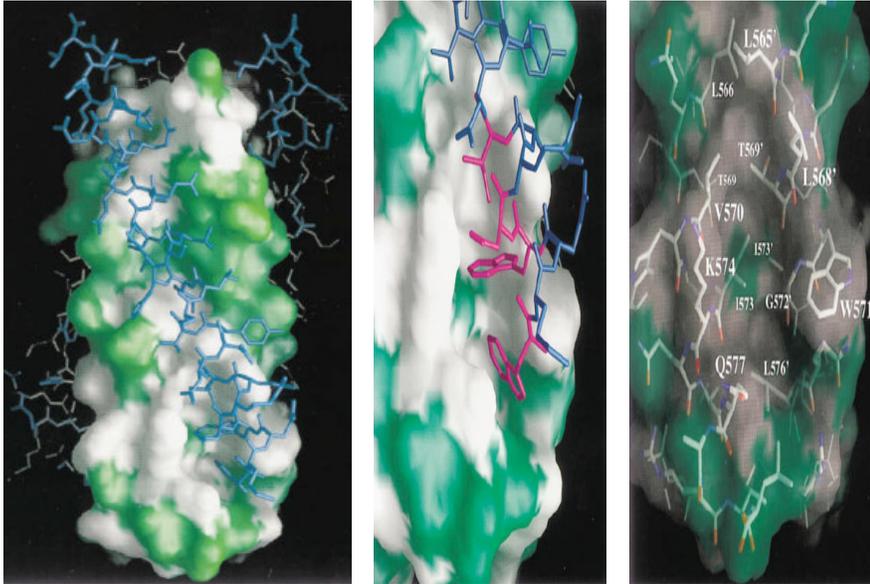


Figura 20. Izquierda: Los péptidos C34, representados con el esqueleto azul, se encuentran frente a la representación en superficie del trímero N36. Los residuos en la superficie del trímero N36 que son idénticos entre el VIH y el VIS están coloreados en blanco y los no idénticos en verde. Nótese que la hélice C34 se empaqueta en un surco conservado (blanco) sobre la superficie del trímero mientras que las regiones no conservadas (verdes) de N36 se encuentran la mayoría expuestas. Centro: residuos que encajan dentro del paquete C34 en magenta (Ile-635, Asp-632, Trp-631 y Trp-628). Derecha: Los residuos que forman la cavidad en el carrete N36 se encuentran etiquetados. Esta orientación está rotada $\sim 30^\circ$ hacia la izquierda con respecto a la imagen central para facilitar la visualización de todas las cadenas laterales. Fuente: [45].

1 Ping Zhu, Jun Liu, Julian Bess Jr, Elena Chertova, Jeffrey D. Lifson, Henry Grisé, Gilad A. Ofek, Kenneth A. Taylor & Kenneth H. Roux. Distribution and three-dimensional structure of AIDS virus envelope spikes, *Nature* 2006 vol 441.

2 J. P. Moore, B. A. Jameson, R. A. Weiss and Q. J. Sattentau, "The HIV-cell fusion reaction" in *Viral Fusion Mechanism*, edited by J. Bentz, CRC Press, Boca Raton, 1993, pp. 233-289.

3 Alam SM, Paleos CA, Liao HX, Searce R, Robinson J, Haynes BF. An inducible HIV type 1 gp41 HR-2 peptide-binding site on HIV type 1 envelope gp120. *AIDS Res Hum Retroviruses* 2004; 20(8):836-845.

4 D. S. Dimitrov, X. Xiao, D.J. Chabot, C.C. Broder, HIV coreceptors, *J. Membrane Biol.* 1998, 16675-16690.

5 Rizzuto CD, Wyatt R, Hernandez-Ramos N, Sun Y, Kwong PD, Hendrickson WA, Sodroski J. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding, *Science* 1998 Jun 19; 280(5371):1949-53.

6 Kliger Y, Gallo SA, Peisajovich SG, Munoz-Barroso I, Avkin S, Blumenthal R, Shai Y. Mode of action of an antiviral peptide from HIV-1. Inhibition at a post-lipid mixing stage. *J Biol Chem.* 2001 Jan 12; 276(2): 1391-7.

7 Koshiba T, Chan DC. The prefusogenic intermediate of HIV-1 gp41 contains exposed C-peptide regions. *J Biol Chem* 2003 Feb 28;278(9):7573-7579.

8 Chang D-K, Cheng S-F, and Trivedi VD. Biophysical characterization of the structure of the amino-terminal region of gp41 of HIV-1. Implications on viral fusion mechanism. *J. Biol. Chem.* 1999 274(9): 5299-5309.

9 Picard L, Simmons G, Power CA, Meyer A, Weiss RA, Clapham PR. Multiple extracellular domains of CCR-5 contribute to human immunodeficiency virus type 1 entry and fusion. *J Virol.* 1997. 71(7):5003-11.

10 Lee B, Sharron M, Blanpain C, Doranz BJ, Vakili J, Setoh P, Berg E, Liu G, Guy HR, Durell SR, Parmentier M, Chang CN, Price K, Tsang M, Doms RW. Epitope mapping of CCR5 reveals multiple

conformational states and distinct but overlapping structures involved in chemokine and coreceptor function. *J Biol Chem.* 1999 274(14):9617-26.

11 J.J. de Jong, A. de Ronde, W. Keulen, M. Tersmette and J. Goudsmit, *J. Virol.* 1992 66, 6777-6780.

12 G. Hoffmann, F. Seiller-Moiseiwitsch, J. Ahn, M. Walker and R. Swanstrom, Variability in the Human Immunodeficiency Virus Type 1 gp120 Env Protein Linked to Phenotype-Associated Changes in the V3 Loop, *J. Virol.* 2002 76, 3852-3864.

13 Olson WC, Rabut GE, Nagashima KA, Tran DN, Anselma DJ, Monard SP, Segal JP, Thompson DA, Kajumo F, Guo Y, Moore JP, Maddon PJ, Dragic T. Differential inhibition of human immunodeficiency virus type 1 fusion, gp120 binding, and CC-chemokine activity by monoclonal antibodies to CCR5. *J Virol.* 1999 73(5):4145-55.

14 Doranz, B., Orsini, M., Turner, J., Hoffman, T., Berson, J., Hoxie, J., Peiper, S., Brass, L. & Doms, R., Identification of CXCR4 domains that support co-receptor and chemokine receptor functions, *J. Vir.* 1999 73, 2757-2761.

15 R. W. Sanders, B. Korber, Min Lu, B. Berkhout and J. P. Moore, Mutational Analyses and Natural Variability of the gp41 Ectodomain, LANL Database, <http://hiv-web.lanl.gov/content/hiv-db/REVIEWS/reviews.html>

16 R. Sutoke and C. Chan, The Fusion Activity of HIV-1 gp41 Depends on Interhelical Interactions, *J. Biol. Chem.* 2005 280, 19852-19857.

17 David C. Chan, Deborah Fass, James M. Berger, and Peter S. Kim. Core structure of Gp41 from the HIV envelope glycoprotein. *Cell* 1997 89, 263-273.

Capítulo V

Fuerzas de unión en moléculas biológicas y estructura de los correceptores utilizados por el VIH

La mayoría de las fuerzas que proveen estabilidad y especificidad a la estructura tridimensional de macromoléculas son interacciones no covalentes. Estas interacciones individualmente débiles pero colectivamente fuertes, incluyen puentes de hidrógeno, interacciones iónicas entre grupos cargados, interacciones de van der Waals e interacciones hidrofóbicas entre grupos no polares. Estas interacciones débiles son temporales; individualmente forman y rompen en fracciones pequeñas de un segundo. La naturaleza transitoria de interacciones covalentes da a las macromoléculas una flexibilidad que es crítica para su función. Más aún, el gran número de interacciones no covalentes en una macromolécula sola, hace improbable que en algún momento dado todas las interacciones se rompan; así las estructuras macromoleculares son estables en el tiempo.

En este capítulo se dará una breve reseña sobre interacciones covalentes y no covalentes.

Tabla 1. Enlaces químicos covalentes y no covalentes.

Tipo de enlace	Longitud [nm]	Fuerza [kcal/mol]	
		En el vacío	En el Agua
Covalente	0.15	90	90
No Covalente	Iónico	80	3
	Hidrógeno	4	1
	Fuerzas de van der Waals (por átomo)	0.1	0.1

V.1 Interacciones No Covalentes

Las interacciones moleculares no covalentes son más débiles que los enlaces covalentes porque no son el resultado de compartir pares de electrones entre átomos. A continuación describiremos brevemente aquellas de mayor relevancia en sistemas biológicos.

Interacciones electrostáticas

También llamadas enlaces iónicos, resultan de la atracción o repulsión electrostática entre dos grupos ionizados de carga opuesta, tales como un grupo carboxilo ($-\text{COO}^-$) y un grupo amino ($-\text{NH}_3^+$). En disolución acuosa, estos enlaces son relativamente débiles comparadas con su magnitud en el vacío (Tabla 1), pero en ambientes hidrofóbicos, tales como las zonas protegidas del agua en el centro de las moléculas de proteína o determinadas zonas de contacto proteína-proteína, son mucho más fuertes. La expresión matemática para esta fuerza es

$$F = \frac{qQ}{4\pi\epsilon d^2}$$

en donde q y Q son las magnitudes de carga eléctrica de las moléculas y d es la distancia entre ellas.

Puentes de Hidrógeno

Los enlaces de hidrógeno determinan la estructura secundaria de las proteínas y el apareamiento de bases complementarias en los ácidos nucleicos. Resultan de la atracción electrostática entre un átomo electronegativo, como el N o el O, y un átomo de H que se encuentra unido covalentemente a un segundo átomo electronegativo:



La formación de puentes de hidrógeno reversibles entre grupos hidrófilos como $\cdot\text{OH}$, $\cdot\text{NH}_2$ y COOH depende mucho del acercamiento íntimo de las dos moléculas que les poseen.

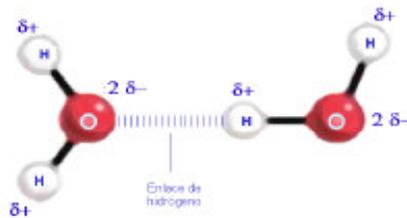


Figura 21. Enlace de Hidrógeno entre dos moléculas de agua debido a la atracción electrostática del átomo de Oxígeno de una de ellas con el de Hidrógeno de la otra.

Aunque los enlaces de H son mucho más débiles que los enlaces covalentes, resultan ser del mismo orden que los enlaces iónicos en un medio acuoso. La exclusión de moléculas de agua entre las cadenas laterales reactivas aumenta considerablemente la energía de la unión a través de la marcada reducción de la constante dieléctrica.

Interacciones hidrófobas

Del mismo modo en que las gotitas de aceite en agua se congregan para formar una sola gota grande, así los grupos hidrófobos no polares como las cadenas laterales de la valina, leucina y fenilalanina tienden a asociarse en un medio acuoso. La fuerza propulsora de esta unión hidrófoba se origina en el hecho de que el agua en contacto con moléculas hidrófobas, con las cuales no puede formar puentes de H, se asociará con otras moléculas de agua pero la cantidad de configuraciones que permiten la formación de enlaces de hidrógeno no será tan grande como la que hay cuando se encuentran completamente rodeadas por otras moléculas de agua, es decir que la entropía es menor. Cuanto mayor sea el área de contacto entre el agua y las superficies hidrófobas, más baja será la entropía del sistema y mayor el estado energético. Por consiguiente, si los grupos hidrófobos de dos proteínas se acercan de modo que se excluyan las moléculas de agua entre ellos, la superficie neta en contacto con el agua se reduce y las proteínas adoptan un estado energético menor que cuando están separadas, o en otras palabras, hay una fuerza de atracción entre ellas.

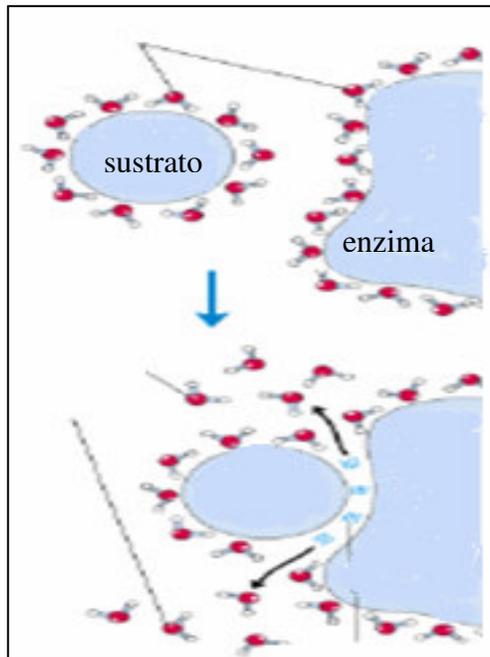


Figura 22. Arriba, las moléculas de agua están ordenadas al interactuar con el sustrato y la enzima. Abajo, el agua se encuentra desordenada, desplazada por la interacción enzima-sustrato, interacción estabilizada por enlaces de hidrógeno e interacciones hidrofóbicas.

Interacciones de van der Waals

Conforman el tipo más débil de fuerza intermolecular que puede darse en la naturaleza y es una de las interacciones no covalentes que estabilizan la estructura de las proteínas. Dependen de la interacción entre las nubes de electrones externas de los átomos. Son fuerzas de origen eléctrico que pueden tener lugar entre dipolos inducidos fluctuantes o entre dipolos permanentes.

Las fuerzas entre dipolos fluctuantes se producen entre moléculas no polares como es el caso del O_2 , N_2 y Cl_2 . Las nubes electrónicas de las moléculas o de los átomos sufren oscilaciones y producen dipolos con una orientación determinada, pero de vida muy breve ya que un instante después el dipolo tiene una orientación distinta. Estos dipolos inducen otra deformación en las moléculas contiguas y originan fuerzas de atracción entre ellas.

Las fuerzas entre dipolos permanentes se producen cuando las moléculas son polares, como el dióxido de azufre (SO_2) o el monóxido de carbono (CO), y existe atracción entre el dipolo de una molécula y el de otra molécula contigua. Estas fuerzas siguen siendo débiles pero son más intensas que las fuerzas de dispersión y, aunque éstas siguen existiendo, predominan las de atracción dipolo-dipolo.

La expresión matemática para la fuerza de atracción entre dos moléculas se puede escribir así

$$F \propto \frac{1}{d^7}$$

y como consecuencia ésta aumenta con rapidez a medida que las moléculas que interactúan se acercan más y más.

Por otro lado existe una interacción derivada del principio de exclusión de Pauli, de naturaleza repulsiva. Esta fuerza es responsable de que tanto los átomos como las moléculas no se colapsen. Su forma matemática es la siguiente

$$F \propto \frac{1}{d^{13}}$$

Este tipo de dependencia subraya una característica esencial común a los cuatro tipos de fuerzas y es que es necesario que las moléculas estén muy próximas para que adquieran una magnitud significativa, en contrapartida con las interacciones electrostáticas que son de largo alcance. Esto es esencial, por ejemplo, para el reconocimiento del antígeno con el anticuerpo. Las formas complementarias de la nube de electrones en el sitio de combinación del anticuerpo y el determinante superficial del antígeno permiten que las dos moléculas encajen perfectamente una en la otra, de manera que la distancia intermolecular disminuye mucho y las fuerzas de interacción proteica inespecíficas aumentan considerablemente; cuanto mayores son las áreas del antígeno y el anticuerpo que participan en la unión, mayor es la fuerza de atracción, en particular si hay oposición de cargas opuestas y agrupaciones hidrófobas.

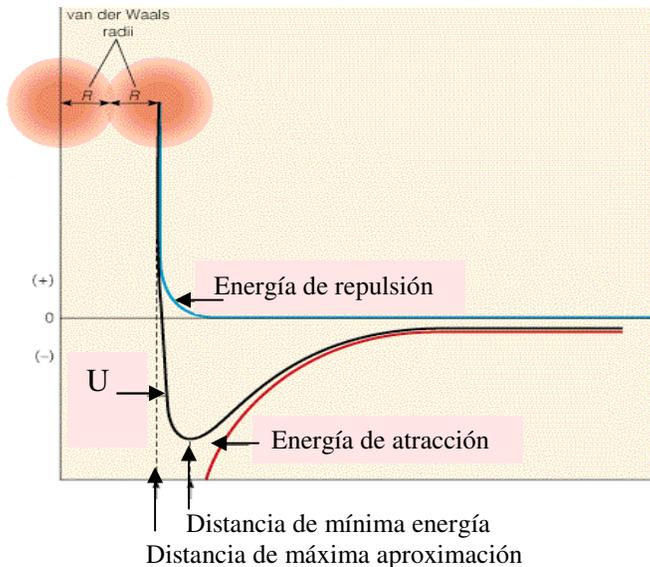


Figura 23. Dependencia de la energía de interacción (atractiva: negativa o repulsiva: positiva) con la distancia de separación entre dos átomos. U es la energía total. Cada átomo tiene un "radio de van der Waals". Dos átomos se atraen máximamente a una distancia igual a la suma de sus radios (mínimo en la curva).

V.2 Interacciones Covalentes

En las soluciones acuosas, los enlaces covalentes son de unas 10 a 100 veces más fuertes que el resto de fuerzas de atracción entre átomos y sus conexiones permiten definir los

límites entre moléculas distintas. Sin embargo la mayor parte de la biología depende de las interacciones específicas entre distintas moléculas. Esta unión entre moléculas se facilita por un conjunto de atracciones no covalentes que individualmente son bastante débiles, pero cuyas energías de enlace pueden sumarse de manera cooperativa dando lugar a una fuerza efectiva entre dos moléculas independientes. En la Tabla 1, se compara la fuerza de diferentes tipos de interacciones con un enlace covalente típico, en presencia y ausencia de agua a causa de su importancia fundamental en todos los sistemas biológicos.

Por definición una estructura polar (en el sentido eléctrico) es aquella que presenta una densidad de carga positiva concentrada en uno de sus extremos y una densidad de carga negativa concentrada en el otro extremo, siendo su carga neta igual a cero. Por lo tanto los enlaces covalentes en los que los electrones se comparten de una manera no equitativa se conocen como *enlaces polares covalentes*. Por ejemplo, los enlaces covalentes entre un oxígeno y un hidrógeno, -O-H, o entre un nitrógeno y un hidrógeno -N-H, son polares, mientras que en el enlace entre un carbono y un hidrógeno, -C-H, los electrones son atraídos de manera más equitativa por ambos átomos, por lo que el enlace es relativamente no polar.

Los enlaces covalentes polares son muy importantes en biología, ya que originan dipolos permanentes que permiten a las moléculas interactuar por medio de fuerzas electrostáticas. Cualquier molécula grande con muchos grupos polares tendrá una configuración de cargas parcialmente positivas y negativas en su superficie. Cuando encuentre una segunda molécula con una serie complementaria de cargas, ambas se atraerán a través de interacciones dipolares permanentes.

V.3 Quimiocinas y sus receptores: regulación de la respuesta inmune ¹

En un organismo, procesos como la morfogénesis embrionaria, la reparación de heridas, la regeneración tisular o la propia respuesta inmune entre otras, dependen del movimiento organizado de las células. El proceso de migración celular es de hecho un proceso altamente coordinado en el que participan multitud de moléculas.

En el caso del sistema inmunológico, la migración de células es importante no sólo en los procesos de extravasación que se producen durante la respuesta inflamatoria, sino que también representa un punto clave en la homeostasis del propio sistema. La migración de leucocitos desde el torrente circulatorio hasta un tejido determinado es un proceso muy regulado en el que participan de forma orquestada muchas proteínas.

Las quimiocinas son una familia de proteínas de bajo peso molecular (8-14KDa) con un alto grado de homología, que originalmente fueron identificadas por su capacidad quimioatrayente, actuando sobre distintas poblaciones leucocitarias. Estructuralmente, las quimiocinas son proteínas muy relacionadas entre sí. Presentan cuatro residuos de cisteína altamente conservados y en función de la disposición de dos de ellos, los más cercanos al extremo amino terminal, se puede establecer una primera clasificación en cuatro familias. Así, clásicamente se definen las quimiocinas CC, donde los dos residuos de cisteína están adyacentes y las quimiocinas CXC que presentan un residuo aminoacídico entre las dos

primeras cisteínas. Las dos familias restantes son XC y CX3C.

Al primer grupo, pertenecen las quimiocinas que juegan un papel relacionado con la propia organización estructural del sistema inmune, es decir, con los procesos de homeostasis; entre ellas están SDF-1 α , TECK o SLC, que son expresadas constitutivamente en tejidos específicos. En el segundo grupo se incluyen aquellas quimiocinas relacionadas con procesos inflamatorios, cuya liberación es inducida por diferentes estímulos como por ejemplo citocinas, y que están implicadas en la infiltración de distintos tipos celulares como macrófagos, linfocitos, neutrófilos, etc. que ocurre durante la inflamación. En este grupo se incluyen por ejemplo IL-8, IP-10, MIG, MIPs o RANTES ⁴⁸.

Las quimiocinas ejercen su función uniéndose a receptores específicos que se expresan en la membrana de otras células. Estos receptores pertenecen a la familia de receptores de siete dominios transmembranales acoplados a proteínas G, GPCRs (del inglés “G Protein-Coupled Receptor”), y están formados por una única cadena polipeptídica de unos 350 aminoácidos. Al atravesar siete veces la membrana plasmática, tanto el extremo amino terminal como tres de los bucles que unen los dominios transmembranales quedan expuestos hacia el exterior de la célula, participando en la unión del ligando, mientras que el extremo carboxi terminal y otros tres bucles se orientan hacia el interior de la célula y participan principalmente en la transmisión de la señal intracelular.

Los receptores de quimiocinas presentan como característica importante la presencia de la secuencia DRYLAIV en el segundo bucle intracelular, que es crítica para su función. Es interesante notar que en el caso del receptor llamado CCR2, la mutación de la tirosina del dominio DRY por una fenilalanina resulta en un receptor no funcional.

V.5 Función biológica y estructural de los correceptores del VIH

Los virus han evolucionado para reconocer una variedad de moléculas superficiales en la célula y utilizarlas para poder insertar sus genomas dentro de la misma. El término *receptor del virus* se utiliza comúnmente para aquellas moléculas superficiales que originalmente no fueron diseñadas para servir como receptor del virus y en la mayoría de los casos no transmiten ninguna señal hacia el interior de las células después de la unión sino que son los mismos componentes del virus los que se transmiten.

En la evolución del VIH, sus actualmente denominados *correceptores* pudieron haber servido inicialmente como receptores pues se podría pensar que el receptor “primario” CD4 sea una adaptación más reciente. Lo anterior está soportado por diversos hallazgos donde se ha observado que CXCR4 uno de los principales correceptores del VIH puede mediar la infección de células que no expresan CD4² por el VIH-2. Por otro lado, aislados del virus de la inmunodeficiencia de los virus (VIS) puede también infectar células CD4⁻ utilizando solamente el otro principal correceptor CCR5 como una molécula receptora, indicando así una posible evolución en el sentido correceptor-receptor. Sin embargo debido a que toda la importancia clínica de los aislados del VIH son críticamente dependientes de CD4 y en la mayoría de los casos el virus utiliza CXCR4, CCR5 u otras moléculas además de CD4 para obtener la entrada dentro de las células, CD4 permanece designado como receptor primario para el VIH.

Además de proveer nuevas perspectivas en aspectos fundamentales de la transmisión y patogénesis del VIH-1, el papel de los correceptores tanto biológica como estructuralmente

constituye un nuevo camino para el desarrollo de estrategias preventivas y terapéuticas novedosas para combatir la epidemia del SIDA. Por lo anterior, esta sección está dedicada a ahondar un poco más sobre la caracterización de CXCR4 y CCR5 para tener un mejor entendimiento de la interacción de éstos con las proteínas de envoltura del VIH-1: gp120 y gp41.

Como ya se dijo antes, los receptores naturales de quimiocinas CXCR4 y CCR5 son los más importantes en el proceso de infección por el VIH³. Casi todos los receptores son selectivos para una clase de quimiocinas.

La molécula CXCR4 contiene dos sitios de glicosilación potencial (ir al apéndice A para una explicación más amplia sobre este concepto), uno en la parte amino terminal y otro en el segundo lazo extracelular (Fig. 24A), mientras que CCR5 tiene sólo uno en el tercer lazo extracelular (Fig. 24B). La parte C-terminal de ambas moléculas es rica en Serinas (muy conservadas) y residuos de treonina. Los residuos altamente conservados de cisteína, que se cree forman parte de puentes disulfuro entre el primer (ECL1) y segundo (ECL2) lazo extracelular, la parte N-terminal y el tercer (ECL3) lazo extracelular, respectivamente, pueden conferir en ellos una forma de barril, llevando los dominios extracelulares a una mayor cercanía entre ellos (Figura 24C).

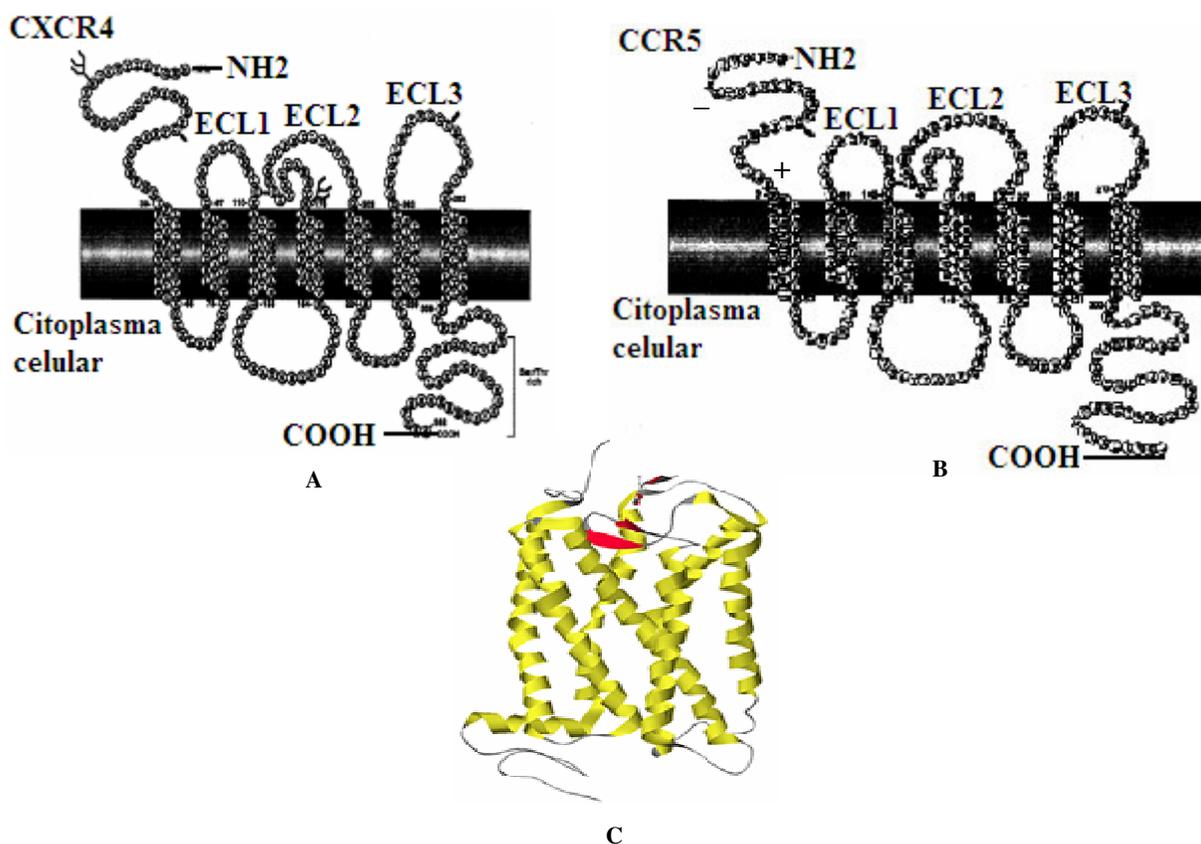


Figura 24. (A) y (B) Representaciones bidimensionales de los receptores CXCR4 y CCR5 respectivamente. Como se observa, no parece haber una diferencia estructural significativa. Cada círculo representa un aminoácido. La parte amino terminal se encuentra extracelular semejándose en forma a la letra “S”. (C) Modelo de cinta para el receptor CXCR4 que muestra las siete hélices que atraviesan la membrana celular (en (A) y (B) están atravesando la parte oscura que simula la membrana celular) para formar 3 rizos en la superficie y una especie de cola (arriba) quedando expuestos y libres para interactuar ya sea con sus ligandos naturales (quimiocinas) o con cualquier molécula que tenga afinidad por él, como es el caso de las proteínas de envoltura del VIH. Lo mismo aplica para CCR5. Fuente: [34].

En lo que respecta a CCR5, su segmento amino terminal contiene 30 residuos aminoácidos, entre los cuales dos son cargados positivamente y tres negativamente, más aún, todas las cargas negativas se encuentran dentro de la región de aminoácidos del 1 al 19 (hacia la punta), mientras que los cargados positivamente están cercanos al dominio transmembranal.

El hecho de que hasta el año 2003 ningún receptor de quimiocinas había sido cristalizado y a que no se habían determinado sus estructuras 3D experimentalmente, no tenía un modelo 3D confiable de CCR5 usando el método de modelación por homología. Sin embargo, aunque la rodopsina tiene una homología de secuencia baja con CCR5 (~20%), ésta pertenece a la familia de las GPCR's y su estructura cristalina despliega una característica conformacional explícita del enrollado de siete hélices transmembranales, además la identidad en la secuencia transmembranal entre CCR5 y la rodopsina es de alrededor del 30%, lo que ayudaba a generar modelos donde la mayoría de los átomos de carbono de las hélices estén en las posiciones correctas con un error de 3.5\AA ⁴. Sin embargo, en el 2003 Shuqun Liu et. al.⁵ realizaron una caracterización funcional y estructural de CCR5 en conjunción con gp120 y CD4 por medio de estudios de modelado molecular, donde se investigaron los puentes de hidrógeno e interacciones de van der Waals entre CCR5 y gp120. Obtuvieron tres modelos instantáneos de CCR5-gp120-CD4, revelando que interacciones iniciales de CCR5 con gp120 están involucradas en la región amino terminal negativamente cargada de CCR5 y la región del "bridging sheet" de gp120. Más aún, fuertes interacciones ocurrieron entre ECL2 de CCR5 y la base del lazo V3 de gp120, las cuales pueden inducir los cambios conformacionales en gp120 y permitir la entrada del VIH a la célula blanco.

Por otro lado, poniendo especial atención en la flexibilidad de la conformación de CCR5 y en interacciones electrostáticas, encontraron que los segmentos extracelulares de este receptor formaban un dominio globular empaquetado con complejas interacciones entre ellos durante la mayor parte del tiempo de simulación, sin embargo la parte amino terminal salía de este dominio algunas veces⁶. También resultó que el puente disulfuro entre el ECL3 y la parte amino terminal es esencial en el control de la orientación específica de esta región y en mantener la integridad en la conformación de todo el dominio extracelular. Por otro lado, encontraron que el lazo ECL1 permanece relativamente rígido mientras que el ECL2 y la parte amino terminal son más flexibles.

1 O. M. Pello, J. M. Rodríguez-Frade, L. Martínez-Muñoz, M. Mellado. Modulación del tráfico leucocitario: Papel de las quimiocinas y de los opioides. *Inmunología* 2006 25 (1):39-49.

2 Endres, M.J., Clapham, P.R., Marsh, J., Ahuja, M., Turner, J.D., McKnight, A., Thomas, J.F., Stoebenu-Haggarty, B., Choe, S., Vance, P.J., Wells, T. N., Power, C.A. et al., CD4-independent infection by HIV-2 is mediated by fusin/CXCR4. 1996 87:745-756.

3 A. Koning, P. van Rij and H. Schuitemaker, Biological and Molecular Aspects of HIV-1 Coreceptor Usage, HIV Sequence Database Los Alamos,

4 R. Sahchez, A. Sali, *J. Comput. Phys.* 1999, 151:388-395.

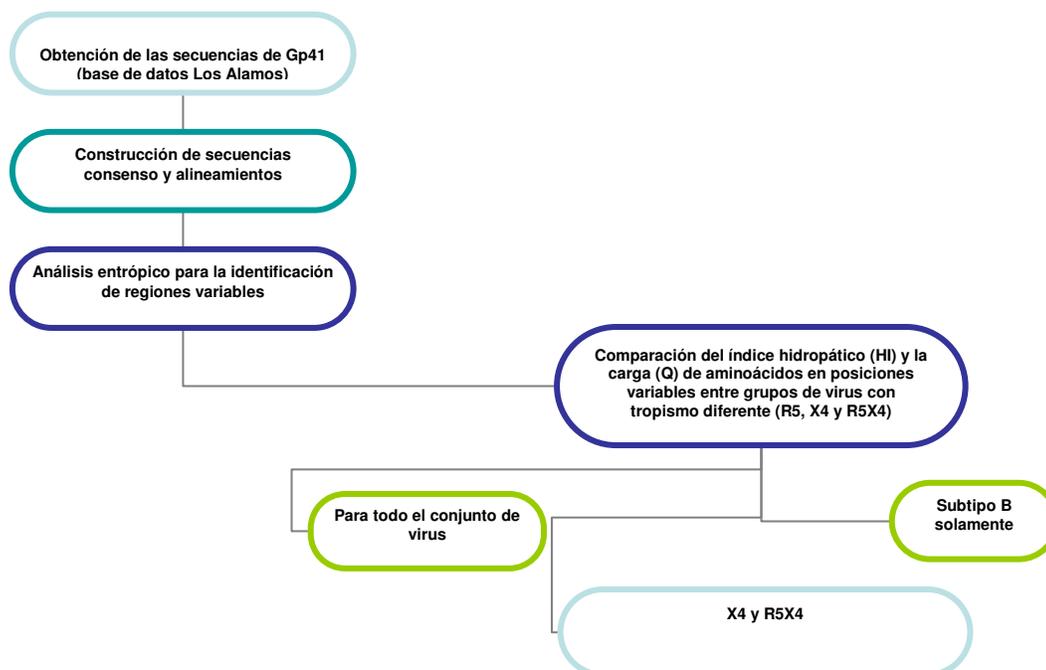
5 Shuqun Liu, Shixiu Fan, Zhirong Sun, Structural and functional characterization of the human CCR5 receptor in complex with HIV gp120 envelope glycoprotein and CD4 receptor by molecular modeling studies, *J. Mol. Model* 2003, 9:329-336.

6 Shuqun Liu, Xiufan Shi, Ciquan Liu, Zhirong Sun, Characterize dynamic conformational space of human CCR5 extracellular domain by molecular modeling and molecular dynamics simulation, *J. Mol. Structure* 2004, 673:133-143.

Capítulo VI

Correlaciones entre el ectodominio de Gp41 y el tropismo infeccioso del VIH-1

Con el objeto de ilustrar la metodología a seguir, se muestra el siguiente diagrama de procedimiento:



VI.1 Obtención de las secuencias del ectodominio de gp41

Con el objeto de buscar correlaciones entre la secuencia de aminoácidos del ectodominio de gp41 y el tropismo viral (es decir, el tipo de correceptor que utiliza el virus en la célula para infectarla, o dicho de otra manera, el tipo de célula a la que prefiere infectar), usamos como “materia prima” todos los virus que se encontraron en la base de datos del Laboratorio Nacional de Los Alamos (Los Alamos National Laboratory) ¹ con tropismo conocido. Se obtuvieron un total de 259 secuencias del ectodominio de gp41 del VIH-1.

La longitud del segmento que se analizó fue de 169 residuos aminoácidos, comprendidos entre las posiciones 516 y 684. La numeración empleada fue la del gen precursor gp160 (1-878) de la secuencia patrón del VIH-1: HXBc2 (Fig. 25).

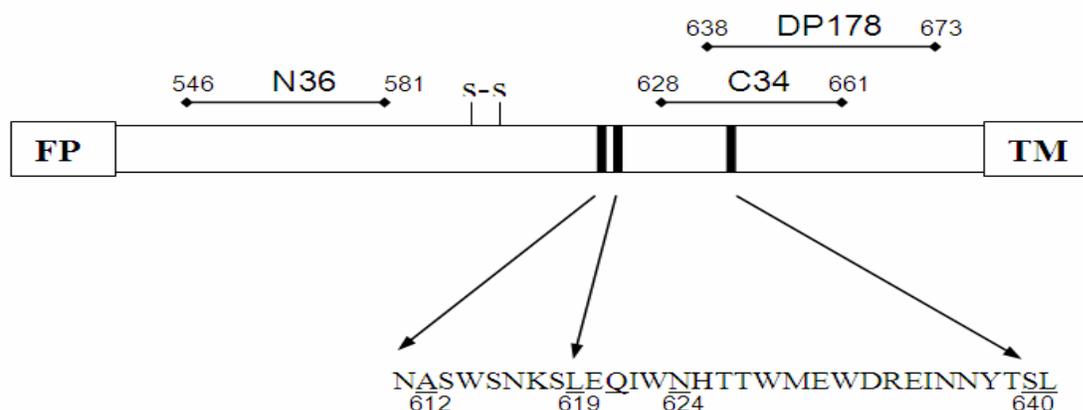


Figura 25. Diagrama de la secuencia del ectodominio de Gp41, mostrando la localización del segmento que va desde el aminoácido 611 hasta el 640 (numeración correspondiente a la cepa de referencia HXBc2), que se encuentran en la hélice N. Las posiciones correspondientes a los péptidos N36 y C34 son las regiones de la hélice N y C respectivamente. También se muestra el péptido de fusión T20 (DP178). Los extremos son del lado izquierdo el péptido de fusión (FP) y en el extremo derecho el dominio transmembranal (TM). S-S indica el puente disulfuro donde interactúan dos cisteínas, localizado justo en la parte central del lazo que separa la hélice N de la C.

VI.2 Construcción de secuencias consenso y alineamientos

Una vez obtenidas las 259 secuencias, en un número considerable de casos el código de identificación del paciente indicaba que existían cepas provenientes del mismo paciente.

Tabla 2. Secuencias de gp41 encontradas hasta octubre de 2005 en la base de datos utilizada. Éstas se refieren a un reconocimiento específico del correceptor.

Tipo de Virus	Número de secuencias
R5	208
X4	34
R5X4	17
Total	259

Para evitar una sobre estimación de secuencias similares de un mismo origen en nuestro análisis, se realizó un consenso² sobre dichas secuencias, dando como resultado una sola que representaría a todas las cepas obtenidas del mismo paciente. Para generar dichas secuencias consenso se utilizó como herramienta el programa de cómputo *BioEdit*, en el cual la frecuencia umbral considerada para la inclusión de un aminoácido en la secuencia consenso fue de 34%, es decir, el aminoácido que se elige como aminoácido consenso en cada posición es aquel que más se repite. Después de este filtrado, la cantidad de secuencias se redujo a 92 del tipo R5, 16 X4 y 12 R5X4, haciendo un total de 120 secuencias.

Por otro lado, es frecuente que ocurran deleciones o inserciones durante el proceso de transcripción de la información genética de cualquier organismo, y el caso del VIH no es la excepción. También pueden existir errores provenientes de la propia secuenciación hecha en el laboratorio donde se aisló el virus. Por ello se procedió a hacer un alineamiento³ de las secuencias en cada grupo, con el fin de situar cada residuo aminoácido en la posición correcta respecto de la numeración de gp160. En la Tabla 11 se pueden apreciar las 120 secuencias que resultaron después de depurar la información.

VI.3 Determinación de sitios variables en la secuencia de aminoácidos de gp41

Se realizó un análisis de la variabilidad de los aminoácidos expresados en cada posición. El objetivo es encontrar aquellas posiciones en el ectodominio de gp41 que puedan contener información relacionada con el tropismo del virus, es decir, con el uso del correceptor, así que buscamos aquellas posiciones que se diferencien entre un grupo y otro, sobre todo entre R5 y X4. Una forma de encontrar tales sitios es considerar aquellos que presentan mutaciones no conservadas entre una cepa y otra, es decir, aquellas posiciones en donde el aminoácido en cuestión es muy cambiante, y que además el cambio involucre un aminoácido con características fisicoquímicas muy diferentes del aminoácido original.

La medida utilizada para medir la variabilidad entre cepas fue la *entropía*. Matemáticamente, la base de la teoría de información fue definida por Claude Shannon⁴:

$$H(l) = -\sum f(b,l) \log_2 f(b,l) \text{ (medida en bits)} \quad (1)$$

donde $H(l)$ representa la entropía, también llamada incertidumbre en la posición l , b representa un residuo aminoácido, y $f(b,l)$ es la frecuencia a la que el residuo b se encuentra en la posición l . El *contenido de información*⁵ en la posición l está definida entonces como una disminución de la incertidumbre en cada posición, o simplemente *entropía*; especialmente en regiones conservadas la entropía debería disminuir.

Con el objeto de determinar la información en cada posición, se deben conocer el número total de residuos posibles y así podemos graficar la entropía como función de la posición en un alineamiento. Ésta dependerá solamente de las frecuencias de caracteres que aparezcan en una columna y será independiente del total de información posible.

En este estudio, utilizamos por conveniencia el logaritmo natural en lugar del logaritmo base 2, ya que los valores en este caso no son bits. Entonces la entropía utilizada en este caso fue

$$S(l) = -\sum f(b,l) \ln f(b,l) \quad (2)$$

la cual proporciona una medida de la incertidumbre en cada posición relativa a otras posiciones. Así, la incertidumbre total máxima se definirá como el número máximo de caracteres diferentes que se encuentren en una columna. Por ejemplo, si están representados los 20 aminoácidos existentes y también el símbolo “-” (representando una delección, p. ej.), entonces la incertidumbre máxima posible asociada a los 21 símbolos resultantes sería

$$S(l) = -21 * \frac{1}{21} \ln\left(\frac{1}{21}\right) = 3.04 \cdot \quad (3)$$

Esto si tuviéramos, digamos, 42 secuencias en el alineamiento y cada caracter estuviera representado dos veces en la columna seleccionada. En la Figura 35 se muestra un ejemplo de cómo se aprecian las secuencias en *BioEdit*, el programa de cómputo utilizado para manipularlas.

Una vez definido el método a utilizar, estableceremos dos criterios para poder

seleccionar aquellas posiciones que nos interesan.

Criterio I

Con el fin de encontrar las posiciones variables del ectodominio de gp41, se graficó la entropía de cada posición (Figura 26 panel superior) calculada según la ecuación 2. Se consideraron como altamente variables aquellas posiciones cuyo valor fuera mayor o igual que 0.9, es decir, $S \geq 0.9$.

Criterio II

También pueden existir posiciones muy variables dentro de un grupo y ser conservadas dentro de otro. Esta diferencia también proporcionaría información acerca de una tendencia a la estabilización de ciertas posiciones favorables al tropismo, por lo tanto calculamos las diferencias de los valores de entropía por cada posición entre dos grupos distintos (Figura 26 panel inferior), en base a la siguiente expresión

$$\Delta S(l) = S_5(l) - S_4(l) = -\sum (f_5(b,l) \ln f_5(b,l) - f_4(b,l) \ln f_4(b,l)) \quad (4)$$

donde $f_5(b,l)$ y $f_4(b,l)$ es la frecuencia con la que aparece el residuo b en la posición l en los grupos R5 y X4 respectivamente. La misma expresión aplica para $f_{5,54}(b,l)$ y $f_{4,54}(b,l)$, al representar la resta de entropía entre el grupo R5 y R5X4, y entre X4 y R5X4 respectivamente. La Figura 25 (panel inferior) muestra las diferencias de entropía entre los grupos mencionados. Las diferencias que se consideraron fueron aquellas que cumplieran con la condición $\Delta S \geq 0.4$.

El conjunto de posiciones seleccionadas en base a los criterios I y II conformó un total de 35 posiciones, dentro de una longitud de 169 aminoácidos, es decir, el 20% del ectodominio de gp41. Dichas posiciones son:

518	535	553	561	578	580	583	588	602	607	612	618
619	620	621	624	625	626	629	636	640	641	643	644
648	655	658	662	665	667	668	674	677	684	687	

(Numeración aminoacídica de acuerdo a gp160 en HXBc2)

Finalmente, basándonos en estos dos criterios pudimos valorar diferencias en la composición de aminoácidos entre los tres grupos de virus (R5 - X4, R5 - R5X4 y X4 - R5X4: Criterio II) y dentro de cada grupo por separado (R5, X4, R5X4: Criterio I).

VI.4 Comparación estadística del índice hidropático y la carga entre grupos de virus con distinto tropismo

Existen 20 aminoácidos diferentes con los cuales se puede conformar una proteína. Cada cual tiene propiedades físico-químicas bien conocidas, como son su conformación estructural, su carga eléctrica y su hidropatía, es decir, la afinidad del aminoácido para interactuar con agua, que está determinada por el índice hidropático (ir al apéndice A para una

explicación más amplia).

En base a estas dos propiedades físico-químicas (carga eléctrica e hidropatía), se realizaron pruebas estadísticas comparativas tanto para conocer qué tan diferentes eran las distribuciones como para medir el grado de importancia de estas diferencias. Esto para cada una de las tres combinaciones de pares: R5 vs. X4, X4 vs. R5X4 y X4 vs. R5X4.

Con el propósito de comprobar que las distribuciones de carga e hidropatía en cada una de estas 35 posiciones eran independientes, se realizó la prueba no paramétrica *U de Mann-Whitney* como un primer acercamiento, determinando estadísticamente la diferencia de las distribuciones entre cada dos grupos. Así mismo se aplicó la prueba χ^2 para tablas de contingencia formadas de acuerdo al tropismo y la magnitud del índice hidropático o la carga. Esta prueba no paramétrica nos informa sobre qué tanto se alejan los valores observados de los esperados, es decir, nos dice el grado de significancia de las diferencias en las distribuciones tanto de carga como de hidropatía. Ver Apéndice B.

Estas dos pruebas se aplicaron a cada uno de los 35 sitios y en cada par de grupos comparados: R5 vs X4, R5 vs R5X4 y X4 vs. R5X4.

A continuación se construyeron histogramas de frecuencia (de carga ó hidropatía, según fue el caso) para aquellas posiciones que tuvieron una *p* significativa, los cuales se pueden apreciar en las Figuras 27 y 28. Además se construyeron histogramas de frecuencia basados en la composición de aminoácidos, los cuales se pueden observar en la Figura 29.

VI.5 Resultados

En una primera aproximación, se analizó estadísticamente la relación entre el uso del correceptor y las propiedades fisicoquímicas de los aminoácidos de todas las secuencias del ectodominio de gp41 con tropismo conocido disponibles en la base de datos *Los Alamos*, primero independientemente del subtipo y posteriormente sólo sobre el subtipo B.

Con el fin de validar la utilidad de nuestro procedimiento para la identificación de sitios relacionados con el tropismo viral, aplicamos el mismo método al análisis al rizo V3 de gp120, del cual se conocen los sitios que están relacionados con el tropismo infeccioso del virus. Estos sitios se utilizan para predecir (con un 71% de certeza) si el virus es X4. Los resultados obtenidos fueron concordantes con los reportados sobre gp120 por varios autores⁴² y a su vez valida los de gp41.

Las tablas y figuras correspondientes a los resultados de este capítulo se encuentran en la sección VI.5.3.

VI.5.1 Gp41: ectodominio

En esta sección se consideraron conjuntamente todas las 120 secuencias obtenidas (Tabla 11). La frecuencia de virus de distintos subtipos en esta muestra se describe en la Tabla 3. La muestra que se utilizó incluye subtipos como el A, B, C, D y otros. Es importante notar que el subtipo B está presente en los tres grupos, aunque en mayor proporción en el X4, mientras que los A sólo están presentes en R5 y R5X4, y los C prácticamente están sólo en R5.

Tabla 3A. En cada casilla se encuentra el porcentaje del subtipo correspondiente dentro de cada grupo. Entre paréntesis se indica el número total de secuencias de cada grupo o subtipo. La última fila indica el porcentaje de cada subtipo tomando en cuenta las 120 secuencias. El grupo OTROS contiene los subtipos F, N, G y O.

Tropismo\Subtipo	A	B	C	D	OTROS
R5 (92)	9% (8)	51% (47)	34% (31)	2% (2)	4% (4)
X4 (16)	0% (0)	75% (12)	6% (1)	19% (3)	0% (0)
R5X4 (12)	8% (1)	67% (8)	8% (1)	8% (1)	8% (1)
TODAS (120)	7% (9)	56% (67)	28% (33)	5% (6)	4% (5)

En la sección VI.5.2 se hizo el mismo estudio pero más selectivo, tomando en cuenta sólo las cepas del subtipo B

VI.5.1.1 Variabilidad de Gp41

En la Tabla 11 se presentan las 120 secuencias resultantes (92 R5, 16 X4 y 12 R5X4) después de la construcción de consensos a partir de las 259 secuencias iniciales.

La variabilidad de las secuencias se presenta en las Figura 25, con los valores de entropía por sitio y diferencia de entropía entre grupos (R5-X4, X4-R5X4 y R5-R5X4). El criterio 1 arrojó un total de 23 posiciones altamente variables. En general las cepas R5 mostraron una variabilidad mayor respecto de las X4 y R5XR4, mientras que entre los grupos X4 y R5X4 no se observó alguna diferencia significativa respecto a su variabilidad.

Los registros más altos de entropía en las cepas R5 están situados en un triplete (posiciones 619-621) y un doblete (posiciones 640-641). Un comportamiento interesante es que tal triplete en R5 se convierte en un doblete en X4 debido a que la entropía en la posición 619 disminuye drásticamente. Un comportamiento similar se observa en el doblete 640-641. Además existe un cúmulo de sitios comprendidos entre las posiciones 662-675 en donde la entropía de las cepas R5 es considerablemente mayor, mientras que en X4 y R5X4 es reducido.

Dado que diferencias muy marcadas de entropía entre las cepas de diferente tropismo (diferente grupo) pueden indicar un papel funcional en el reconocimiento del correceptor, se incorporaron al análisis los sitios que mostraron las diferencias más altas. Así, un total de 35 posiciones satisficieron al menos uno de los dos criterios, mismas que fueron sometidas al análisis estadístico.

VI.5.1.2 Relación entre el uso del correceptor y la carga e hidropatía de los aminoácidos

Los resultados obtenidos en relación al carácter hidropático y la carga eléctrica de los aminoácidos que componen las 35 posiciones seleccionadas se encuentran en las Tablas 4-7. Se muestran los valores p obtenidos para la independencia (prueba U Mann-Whitney) en las distribuciones de HI (índice hidropático) y Q (carga eléctrica), así como los correspondientes para encontrar correlaciones significativas (prueba χ^2).

En lo que respecta al índice hidropático, se obtuvieron valores significativos ($p \leq 0.5$) en las dos pruebas estadísticas aplicadas. Sin embargo, varias de las asociaciones son

artefactos provenientes de distribuciones pico donde la desviación fue muy pequeña o de distribuciones que constaban de un solo valor, provocando así que diferencias muy pequeñas entre dos grupos se magnificaran, dando valores significativos espurios (posiciones 561, 578, 602, 626 y 643). Por otro lado, existen sustituciones conservativas donde el aminoácido en cuestión es cambiante, pero las propiedades fisicoquímicas de los sustitutos son prácticamente las mismas, por lo que diferencias pequeñas en el índice hidropático (menores que 1) con valores p significativos fueron consideradas como no relevantes.

Se encontraron correlaciones relevantes entre HI y el tropismo del virus en 8 posiciones: 553, 612, 619, 621, 640, 641, 662 y 667. La primera está localizada en la hélice N (553), la siguiente en la región del lazo (612), las dos siguientes en la interfase del lazo y la hélice C (619 y 621), las próximas 3 dentro de la hélice (640, 641 y 662) y la última en la región cercana al segmento transmembranal (667). En la Figura 27 se muestran las distribuciones de índices hidropáticos para las posiciones estadísticamente significativas.

En lo que respecta a la carga, se obtuvieron valores significativos ($p \leq 0.055$) mostrando una asociación entre residuos cargados y no cargados con el tropismo infeccioso del virus. Tales correlaciones significativas se encontraron en las posiciones 602, 618, 621, 624, 643 y 662. Así como en el caso del HI, las posiciones realmente relevantes fueron la 621, 624 y 662. Los histogramas de la distribución de carga en tales posiciones se muestran en la Figura 28. En la Figura 32 se muestra de manera esquemática la posición de cada sitio y en la Figura 29 los histogramas de distribución de aminoácidos para todas las posiciones que resultaron ser relevantes según los argumentos arriba mencionados.

Los virus X4 presentan una frecuencia alta de residuos hidrofóbicos ($HI > 0$) en cinco posiciones: A en 612 y 667, L en 619 y 641, G o S en 640. De manera contraria, residuos hidrofílicos ocupan esas mismas posiciones en una fracción considerable de cepas R5. Por otro lado, residuos hidrofílicos (N y E) predominan en las posiciones 553 y 662 de las cepas X4. Por último, las cepas X4 despliegan una frecuencia bastante alta de residuos no cargados en las posiciones 621 y 624 (Q, Y o A en 621 y N en 624), mientras que en las R5 se presenta una alta frecuencia de residuos cargados negativamente. Las cepas R5X4 muestran un comportamiento similar al de las X4 en seis posiciones (553, 619, 640, 641, 662 y 667), al de las R5 en sólo una posición (612) y un patrón intermedio en dos posiciones (621 y 624).

Resumiendo, las sustituciones no conservativas en las secuencias de Gp41 asociadas en nuestro estudio al tropismo del virus pueden clasificarse en cuatro categorías:

1. el residuo es más hidrofóbico en las cepas X4: 612, 619, 640, 641 y 667
2. el residuo es más hidrofílico en las cepas X4: 553 y 662
3. el residuo está cargado negativamente en las cepas R5: 621 y 624
4. el residuo está cargado negativamente en las cepas X4: 662

En la Tabla 10 y Figura 33 se resumen las características más importantes de éstas posiciones según la distribución de frecuencias obtenida así como la ubicación espacial de cada uno de los sitios, mencionando también algunas consideraciones funcionales.

Finalmente, la proximidad entre los sitios 619, 621 y 624, así como entre 640 y 641 sugieren que podría ocurrir una interacción sinérgica durante el reconocimiento del correceptor, ya el grado de correlación como sitios independientes aumenta cuando se consideran como dobles en HI (619/621 y 640/641, $p = 0.001$, y $p = 0.016$, respectivamente) y en Q (621/624, $p = 0.015$), dotando a estas parejas de un vínculo estadístico mayor entre el carácter fisicoquímico del aminoácido y el tropismo.

Es interesante el hecho de que los cúmulos 619-621 y 640-641 son además las regiones más variables del ectodominio de Gp41 en el grupo R5, aunque las posiciones 619, 640 y 641 (junto con la 667) son más conservadas en el grupo X4. La tabla con las pruebas aplicadas y las gráficas correspondientes a las parejas consideradas se encuentran en la Tabla 8 y Figura 30, respectivamente.

VI.5.2 Subtipo B

En la Figura 38 se graficaron las frecuencias de cada subtipo en cada uno de los tres grupos, donde se puede observar la alta frecuencia del subtipo B en los virus X4 y R5X4 en comparación con los R5 (Tabla 3A). Además en la Tabla 12 se tiene reportado el comportamiento fisicoquímico de los aminoácidos en los sitios relevantes de cada subtipo (sin considerar el grupo de tropismo).

Debido a la presencia del subtipo B en los tres grupos y su prevaecía tanto en México como en el continente americano, se realizó el mismo análisis con el objeto de investigar si los resultados obtenidos se seguían cumpliendo aún dentro de este subtipo. Los resultados se muestran en la Tabla 9, donde se observa que las posiciones 612, 619 y 621 resultaron ser significativas en la asociación entre el tropismo y el HI, mientras que la posición 621 lo resultó en Q ($p < .05$), lo cual es consistente con el análisis anterior donde se incluyen todos los subtipos, pues el grupo X4 conserva la tendencia de contar con residuos altamente hidrofóbicos en las posiciones 612 y 619 (A y L, respectivamente) y un residuo sin carga en la posición 624 (N). En comparación con las cepas X4, las R5 mostraron un mayor contenido de residuos cargados (K, D y E) en la posición 621. La distribución correspondiente de aminoácidos en dichas posiciones así como de carga e hidropatía se presentan en la Figura 31, mientras que en la Tabla 9 se muestran los resultados para las parejas (posiciones espacialmente cercanas) significativas.

Por otro lado, la combinación en el HI de las posiciones 619 y 621 resultó ser significativa de manera marginal en la prueba χ^2 ($p < 0.05$).

VI.5.3 Gp120: Lazo V3

Consideramos todas las cepas reportadas de Gp120 con tropismo conocido en la base de datos de *Los Alamos*. Éstas fueron 746 del tipo R5, 166 X4 y 73 R5X4. La longitud de la secuencia fue de 35 aminoácidos y la proporción de los distintos subtipos presentes en cada grupo se muestra en la Tabla 3B.

Tabla 3B. Proporción de subtipos en los grupos de tropismo del lazo V3. En cada casilla se encuentra el porcentaje del subtipo correspondiente dentro de cada grupo. Entre paréntesis se indica el número total de secuencias de cada grupo o subtipo. La última fila indica el porcentaje de cada subtipo tomando en cuenta las 985 secuencias. El grupo OTROS contiene los subtipos F, N, G, O y U. Nótese que al igual que en Gp41, el grupo R5 tiene menos contenido de subtipos B que los grupos restantes.

Tropismo\Subtipo	A	B	C	D	OTROS
R5 (746)	11% (79)	69% (515)	18% (132)	<1% (5)	2% (15)
X4 (166)	3% (4)	77% (128)	1% (2)	19% (31)	<1% (1)
R5X4 (73)	1% (1)	92% (67)	1% (1)	5% (3)	1% (1)
TODAS (985)	8% (84)	72% (710)	14% (135)	4% (39)	2% (17)

Se desarrolló un análisis idéntico al aplicado al ectodominio de Gp41, resultando que las posiciones con la correlación más alta en relación al tropismo fueron la **11** y la **25** (308 y 322 respectivamente según la numeración de HXBc2), lo cual es consistente con los hallazgos previos y la noción de que una presión selectiva *in vivo* por el cambio en el uso del correceptor está mediada por la habilidad del virus por interactuar con distintas moléculas correceptoras (tropismo viral). La variabilidad del lazo V3 se aprecia en la Figura 34, mientras que las distribuciones de carga e hidropatía se encuentran en la Figura 35, donde se observa un comportamiento similar al del ectodominio de Gp41.

VI.6 Tablas y Figuras

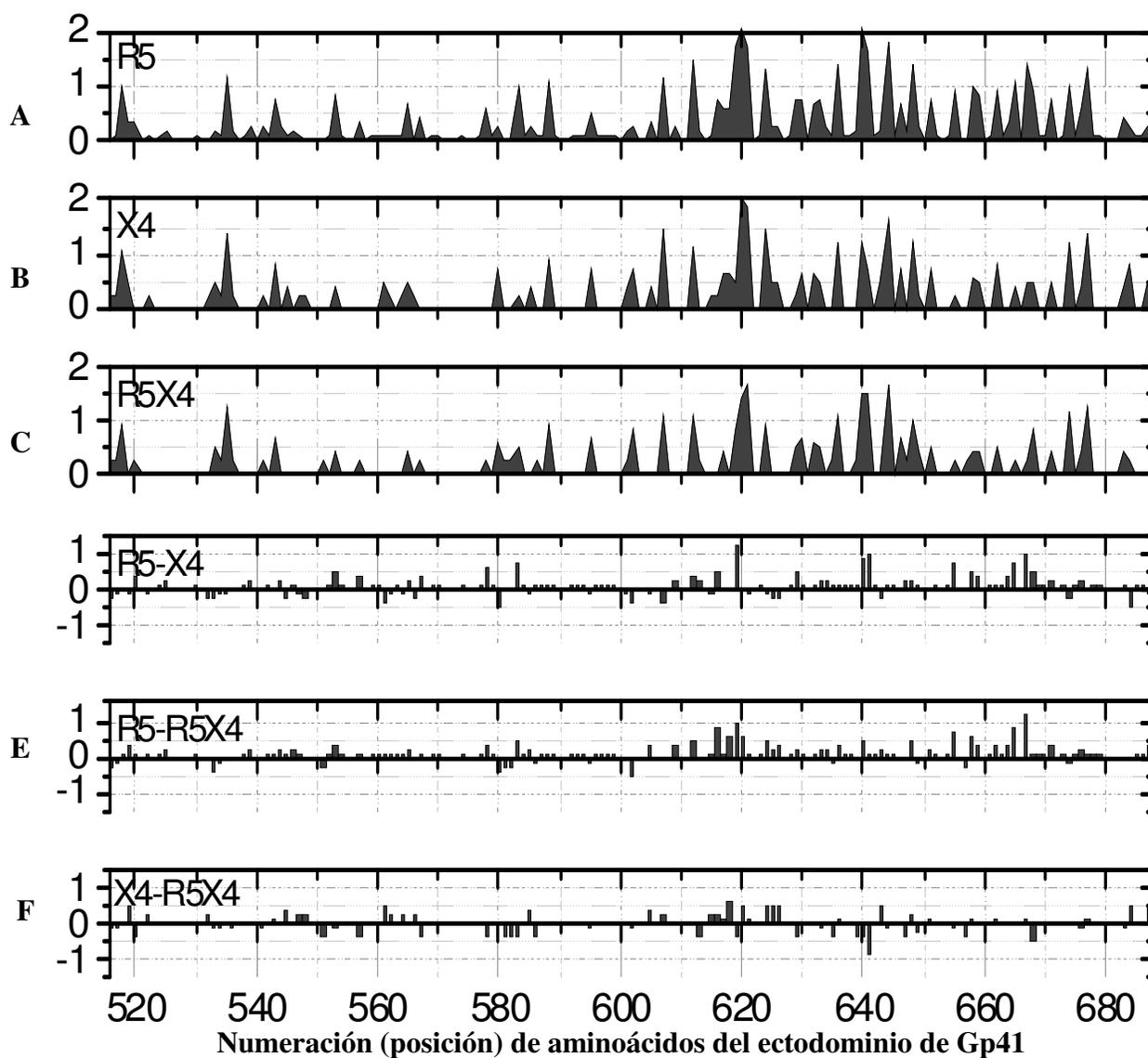


Figura 25. Gráficas A, B y C: distribución de valores de entropía (eje vertical) por posición de aminoácidos (eje horizontal) del ectodominio de la gp41 para las cepas virales tipo R5, X4 y R5X4 respectivamente. Gráficas D, E y F: distribución de las diferencias de entropía entre cada dos grupos R5-X4, X4-R5X4 y R5-R5X4 respectivamente. Numeración aminoacídica de acuerdo al gene precursor gp160 de la cepa de referencia del VIH (HXBc2).

Tabla 4. Índice hidropático promedio con desviación estándar y valores de p en la prueba U de Mann y χ^2 , indicando el error que se tendría al suponer la hipótesis nula según la prueba; Mann: las distribuciones comparadas son independientes, χ^2 : existe una correlación entre el carácter hidropático y el tropismo. Los guiones indican que la prueba no se pudo realizar.

Posición	Índice Hidropático Promedio (SD)			p (prueba χ^2) hidrofóbicos vs. hidrofílicos			p (U Mann)		
	R5	X4	R5X4	R5 vs. X4	R5 vs. R5X4	X4 vs. R5X4	R5 vs. X4	R5 vs. R5X4	X4 vs. R5X4
518	3.23 (1.08)	3.16 (1.02)	2.76 (1.07)	-	-	-	0.293	0.127	0.392
535	3.87 (1.07)	3.41 (1.08)	3.59 (1.06)	0.675	0.717	-	0.009	0.125	0.622
553	-2.11 (1.39)	-3.16 (0.92)	-3.05 (1.05)	-	-	-	0.006	0.031	1.000
561	1.83 (0.25)	1.63 (0.64)	1.80 (0.00)	0.013	-	0.36	0.243	1.000	1.000
578	1.20 (1.15)	1.80 (0.00)	1.59 (0.72)	0.033	0.247	0.24	0.041	0.337	0.429
580	4.20 (0.00)	4.33 (0.15)	4.30 (0.15)	-	-	-	9.47E-06	0.001	0.705
583	4.20 (0.47)	4.22 (0.07)	4.03 (0.68)	-	-	-	0.172	0.141	1.000
588	-3.97 (0.60)	-3.96 (0.30)	-4.12 (0.36)	-	-	-	0.497	0.410	0.296
602	3.61 (1.13)	2.53 (2.85)	2.58 (3.02)	0.012	0.041	0.89	0.006	0.267	0.977
607	-0.03 (2.12)	-0.44 (1.96)	0.10 (1.69)	0.202	0.539	0.66	0.341	0.986	0.451
612	-0.26 (1.67)	0.99 (1.52)	0.10 (1.26)	0.004	0.611	0.13	0.007	0.240	0.224
618	-0.90 (0.56)	-0.96 (0.68)	-0.80 (0.00)	-	-	-	0.973	0.545	0.785
619	0.20 (3.39)	3.34 (1.93)	2.79 (2.59)	5.00E-04	0.017	0.38	0.001	0.019	0.675
620	-2.47 (1.59)	-2.66 (1.66)	-2.81 (1.25)	0.3707	0.602	0.3778	0.589	0.672	0.965
621	-3.02 (1.16)	-1.85 (2.13)	-2.48 (2.15)	0.002	0.014	0.825	0.008	0.997	0.227
624	-3.26 (0.81)	-3.11 (1.06)	-3.24 (0.89)	-	-	-	0.852	1.000	1.000
625	-3.51 (0.06)	-3.48 (0.08)	-3.50 (0.00)	-	-	-	0.123	1.000	1.000
626	2.00 (0.44)	1.73 (0.67)	1.90 (0.00)	0.013	-	0.362	0.086	0.634	1.000
629	2.01 (1.32)	2.06 (0.65)	1.61 (1.70)	0.461	0.3959	0.24	0.669	0.416	0.893
636	-2.52 (1.32)	-2.61 (1.37)	-2.59 (1.34)	-	-	0.204	0.896	1.000	1.000
640	-2.46 (1.44)	-1.47 (1.42)	-2.02 (1.63)	-	-	-	0.013	0.350	0.562
641	1.87 (2.65)	3.24 (1.67)	2.60 (2.76)	0.031	0.295	0.393	0.264	0.302	0.687
643	-1.16 (0.00)	-1.66 (0.76)	-1.30 (0.00)	0.4613	0.5232	-	0.004	1.000	0.238
644	-2.34 (1.85)	-1.19 (2.07)	-2.17 (1.69)	0.142	0.7268	0.3825	0.296	0.795	0.596
648	-2.96 (1.95)	-3.09 (1.95)	-3.24 (0.89)	0.8397	0.3196	0.3778	0.312	0.681	0.379
655	-3.67 (1.29)	-3.54 (1.45)	-3.87 (0.12)	0.36	0.606	0.378	0.807	0.827	1.000

658	-3.10 (2.00)	-3.60 (0.18)	-3.57 (0.15)	0.254	0.322	-	0.680	0.402	0.673
662	-0.83 (2.66)	-2.2 (2.39)	-3.06 (1.53)	0.0642	0.0064	0.2545	0.073	0.012	0.519
665	-3.05 (1.40)	-3.85 (0.14)	-3.64 (0.89)	-	-	-	0.109	0.159	1.000
667	-0.29 (2.57)	1.31 (1.44)	1.33 (1.65)	0.02297	0.0225	0.7243	0.017	0.047	1.000
668	-1.86 (1.42)	-1.31 (1.08)	-1.67 (1.36)	0.6718	0.7136	-	0.153	0.551	0.662
674	-2.75 (1.23)	-2.65 (1.30)	-2.83 (1.22)	-	-	-	0.666	0.92	0.739
677	-3.57 (0.49)	-3.39 (0.73)	-3.41 (0.84)	-	-	-	0.264	0.912	0.542
684	4.43 (0.39)	4.31 (0.32)	4.44 (0.20)	-	-	-	0.011	0.927	0.342
687	2.13 (0.74)	2.39 (1.05)	1.90 (0.00)	-	-	-	0.363	0.592	0.238

Tabla 5. Índice hidropático. Prueba χ^2 fina. Segunda columna: comparación sólo entre R5 y X4. Tercera columna: comparación entre los tres grupos. Esta prueba hace una comparación más fina que la χ^2 , ya que no se hace una separación de dos rangos en el índice hidropático, sino que se compara valor por valor, aún cuando éstos son sólo hidrofóbicos (o hidrofílicos) y se mide la correlación.

Posición	χ^2_{kx2} R5-X4	χ^2_{kx3} R5-X4-R5X4
518	0.902	0.426
535	0.407	0.719
553	0.008	0.007
561	0.042	0.122
578	0.104	0.226
580	-	-
583	0.461	0.479
588	0.802	0.943
602	0.014	0.007
607	0.287	0.308
612	0.068	0.243
618	0.738	0.704
619	0.005	0.006
620	0.545	0.783
621	0.008	0.023
624	0.514	0.808
625	0.562	0.739
626	0.032	0.078
629	0.704	0.750
636	0.827	0.986
640	0.091	0.269
641	0.148	0.292
643	<0.001	<0.001
644	0.127	0.338
648	0.816	0.478
655	0.068	0.255
658	0.264	0.215
662	0.281	0.066
665	0.045	0.058
667	0.119	0.090
668	0.202	<0.001
674	0.737	0.925
677	0.598	0.740
684	0.564	0.742
687	0.220	0.223

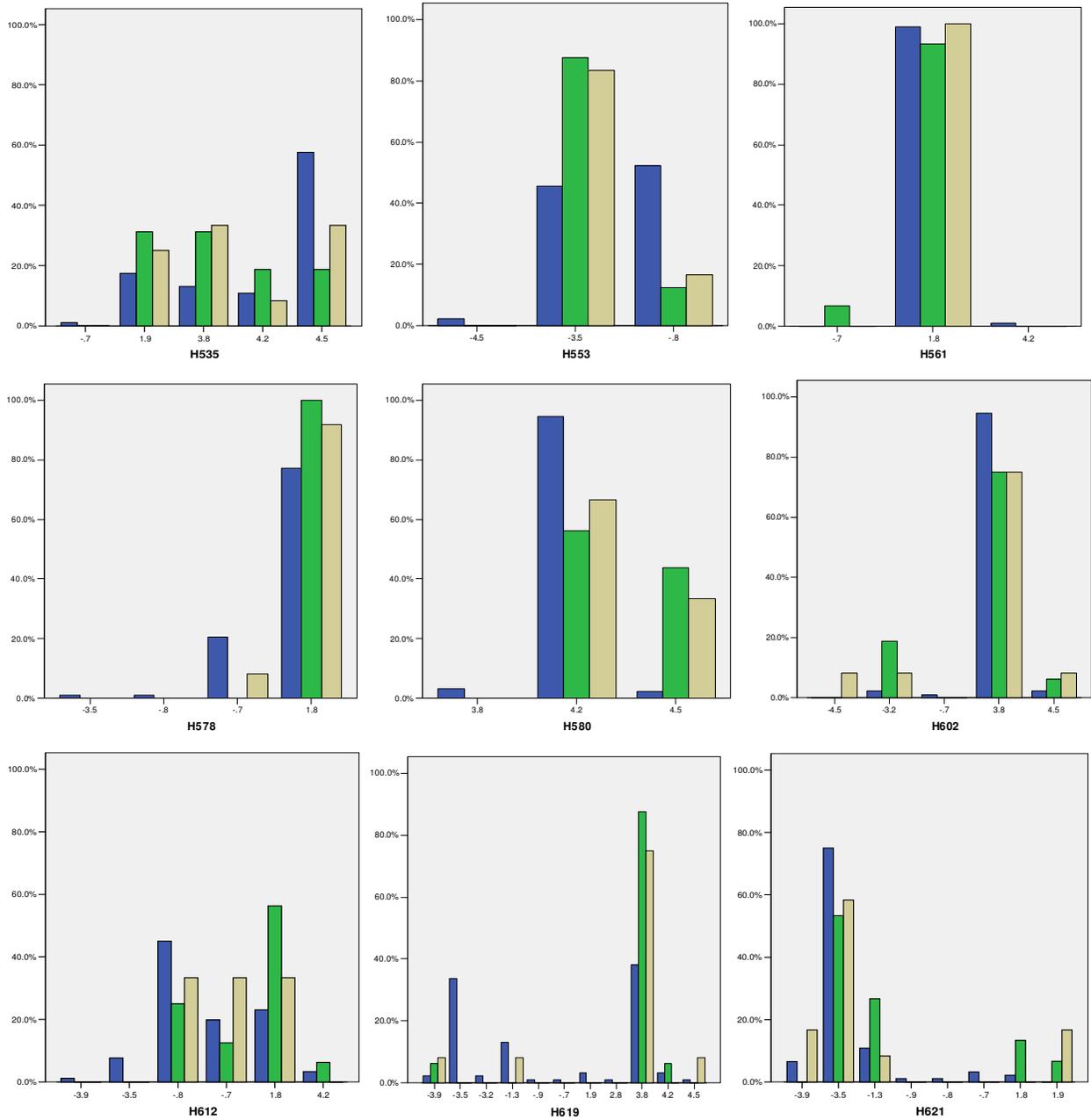


Figura 27. Distribución de frecuencias (eje vertical) de índice hidropático (eje horizontal) de todas las posiciones que resultaron significativas en las Tablas 4 y 5. Azul: R5, Verde: X4, Amarillo: R5X4. Continúa en la página siguiente.

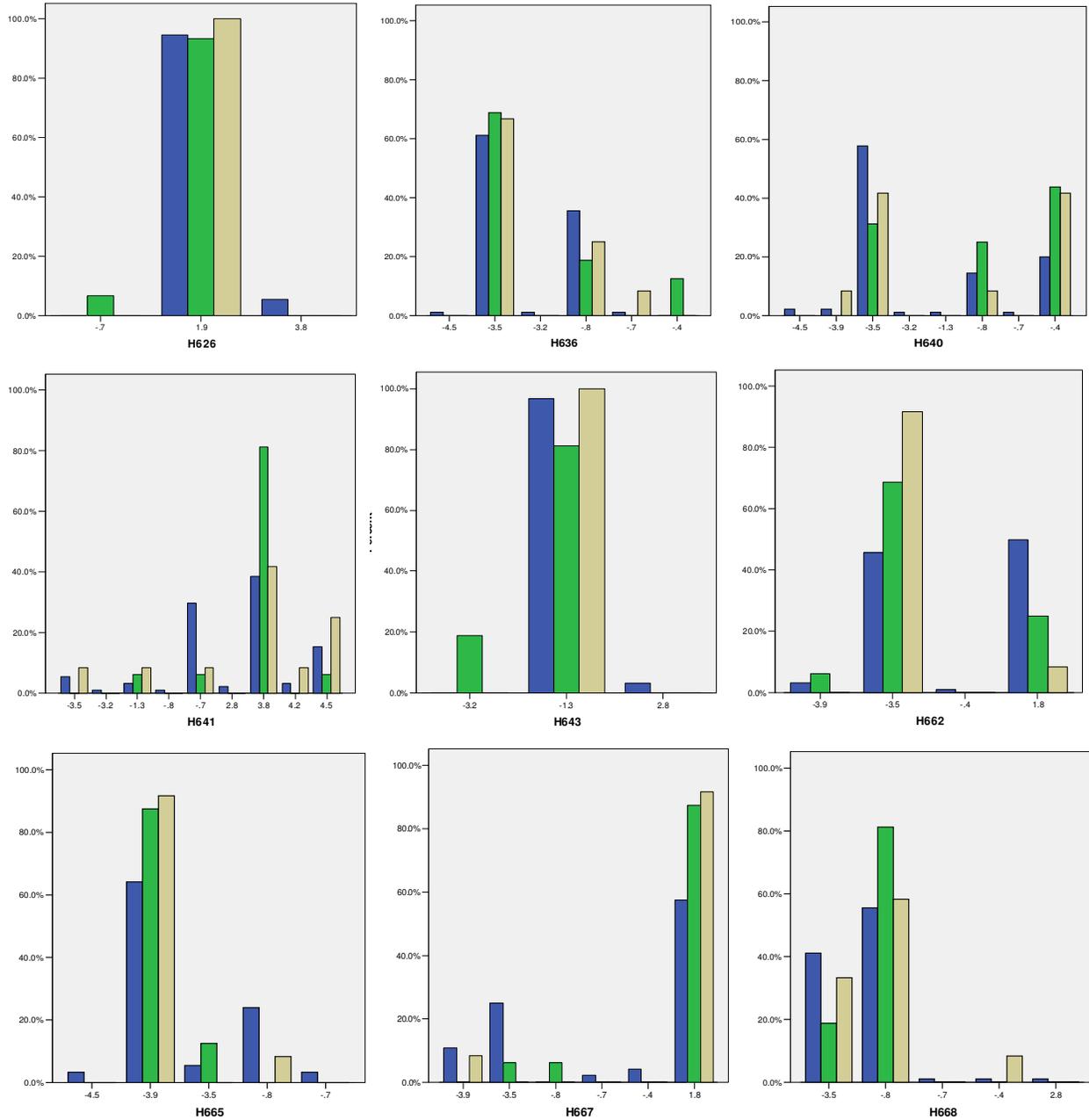


Figura 27 (continuación página anterior). Distribución de frecuencias (eje vertical) de índice hidropático (eje horizontal) de todas las posiciones que resultaron significativas (18) en las Tablas 4 y 5. Azul: R5, Verde: X4, Amarillo: R5X4.

Tabla 6. Carga eléctrica promedio con desviación estándar y valores de p en la prueba U Mann y X^2 . El valor de p indica el error que se tendría al suponer la hipótesis nula según la prueba Mann: las distribuciones comparadas son independientes y distintas, X^2 : existe una relación entre ser cargado y el tropismo del virus en esa posición. Los guiones indican que la prueba no se pudo realizar debido a que las distribuciones eran idénticas, por lo que no tiene sentido suponer diferencias.

Posición	Carga Eléctrica Promedio (SD)			p (prueba χ^2 cargados vs. neutros (-1, +1 vs. 0))			p (U Mann)		
	R5	X4	R5X4	R5 vs. X4	R5 vs. R5X4	X4 vs. R5X4	R5 vs. X4	R5 vs. R5X4	X4 vs. R5X4
518	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-	-	-	1.000	1.000	1.000
535	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-	-	-	1.000	1.000	1.000
553	0.01 (0.18)	0.00 (0.00)	0.00 (0.00)	0.464	0.526	-	0.964	0.978	1.000
561	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-	-	-	1.000	1.000	1.000
578	-0.01 (0.10)	0.00 (0.00)	0.00 (0.00)	0.675	0.717	-	1.000	1.000	1.000
580	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-	-	-	1.000	1.000	1.000
583	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-	-	-	1.000	1.000	1.000
588	0.88 (0.33)	0.81 (0.54)	0.92 (0.29)	0.495	0.703	0.832	0.735	1.000	0.853
602	0.01 (0.09)	0.11 (0.24)	0.13 (0.32)	-	-	-	0.023	0.035	1.000
607	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-	-	-	1.000	1.000	1.000
612	-0.03 (0.23)	0.00 (0.00)	0.00 (0.00)	0.337	0.405	-	0.945	0.963	1.000
618	0.00 (0.00)	-0.06 (0.25)	-0.80 (0.00)	0.016	-	0.378	0.148	1.000	1.000
619	-0.02 (0.29)	0.06 (0.25)	0.08 (0.29)	0.653	0.873	0.832	0.313	0.266	1.000
620	0.33 (0.59)	0.37 (0.62)	-0.58 (0.51)	0.742	0.405	0.662	0.751	0.205	0.51
621	-0.57 (0.62)	-0.33 (0.49)	-0.33 (0.77)	0.007	0.838	0.085	0.082	0.342	0.783
624	-0.59 (0.51)	-0.37 (0.50)	-0.50 (0.52)	0.055	0.382	0.508	0.103	0.539	0.702
625	0.01 (0.18)	0.04 (0.15)	0.00 (0.00)	0.526	0.523	0.362	0.419	0.977	1.000
626	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-	-	-	1.000	1.000	1.000
629	0.01 (0.10)	0.00 (0.00)	0.00 (0.00)	-	-	-	1.000	1.000	1.000
636	-0.36 (0.52)	-0.50 (0.52)	-0.58 (0.52)	0.454	0.227	0.662	0.382	0.203	0.718
640	-0.25 (0.54)	-0.31 (0.48)	-0.25 (0.62)	0.739	0.679	0.569	0.784	1.000	1.000
641	-0.05 (0.24)	0.00 (0.00)	-0.08 (0.29)	0.29	0.822	0.24	0.579	1.000	0.429
643	0.00 (0.00)	0.11 (0.24)	0.00 (0.00)	-	-	-	0.003	1.000	0.238
644	0.27 (0.52)	0.11 (0.29)	0.18 (0.40)	0.108	0.287	0.735	0.176	0.673	0.807
648	-0.67 (0.65)	-0.50 (0.82)	0.83 (0.39)	0.94	0.741	0.755	0.425	0.557	0.281
655	0.80 (0.45)	0.94 (0.25)	0.83 (0.58)	0.552	0.231	0.24	0.352	0.685	0.746

658	0.38 (0.48)	0.25 (0.45)	0.17 (0.39)	-	-	-	0.404	0.205	0.673
662	-0.41 (0.56)	-0.63 (0.62)	-0.83 (0.39)	0.045	0.021	0.595	0.115	0.014	0.492
665	0.66 (0.50)	0.87 (0.34)	0.92 (0.29)	0.675	0.717	-	0.137	0.115	1.000
667	0.05 (0.40)	0.00 (0.00)	0.08 (0.29)	0.082	0.472	0.24	0.533	1.000	0.429
668	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-	-	-	1.000	1.000	1.000
674	-0.63 (0.48)	-0.44 (0.51)	-0.58 (0.51)	0.146	0.751	0.445	0.172	1.000	0.704
677	0.33 (0.51)	0.32 (0.45)	0.33 (0.49)	0.916	0.71	0.82	0.977	0.983	1.000
684	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-	-	-	1.000	1.000	1.000
687	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-	-	-	1.000	1.000	1.000

Tabla 7. Carga eléctrica. Prueba χ^2 comparando aminoácidos cargados positivamente contra cargados negativamente.

Posición	<i>p</i> (prueba χ^2) cargados + vs. - (+1,0 vs. -1)		
	R5 vs. X4	R5 vs. R5X4	X4 vs. R5X4
518	-	-	-
535	-	-	-
553	0.675	0.717	-
561	-	-	-
578	0.675	0.717	-
580	-	-	-
583	-	-	-
588	0.017	-	0.378
602	-	-	-
607	-	-	-
612	0.393	0.459	-
618	0.016	-	0.378
619	0.340	0.408	-
620	0.714	0.199	0.445
621	0.030	0.382	0.381
624	0.081	0.470	0.508
625	0.683	0.715	-
626	-	-	-
629	-	-	-
636	0.357	0.172	0.662
640	0.920	0.814	0.907
641	0.337	0.693	0.240
643	-	-	-
644	0.476	0.541	-
648	0.482	0.616	0.378
655	0.552	0.231	0.240
658	-	-	-
662	0.074	0.012	0.378
665	0.675	0.717	-
667	0.340	0.408	-
668	-	-	-
674	0.146	0.751	0.445
677	0.547	0.602	-
684	-	-	-
687	-	-	-

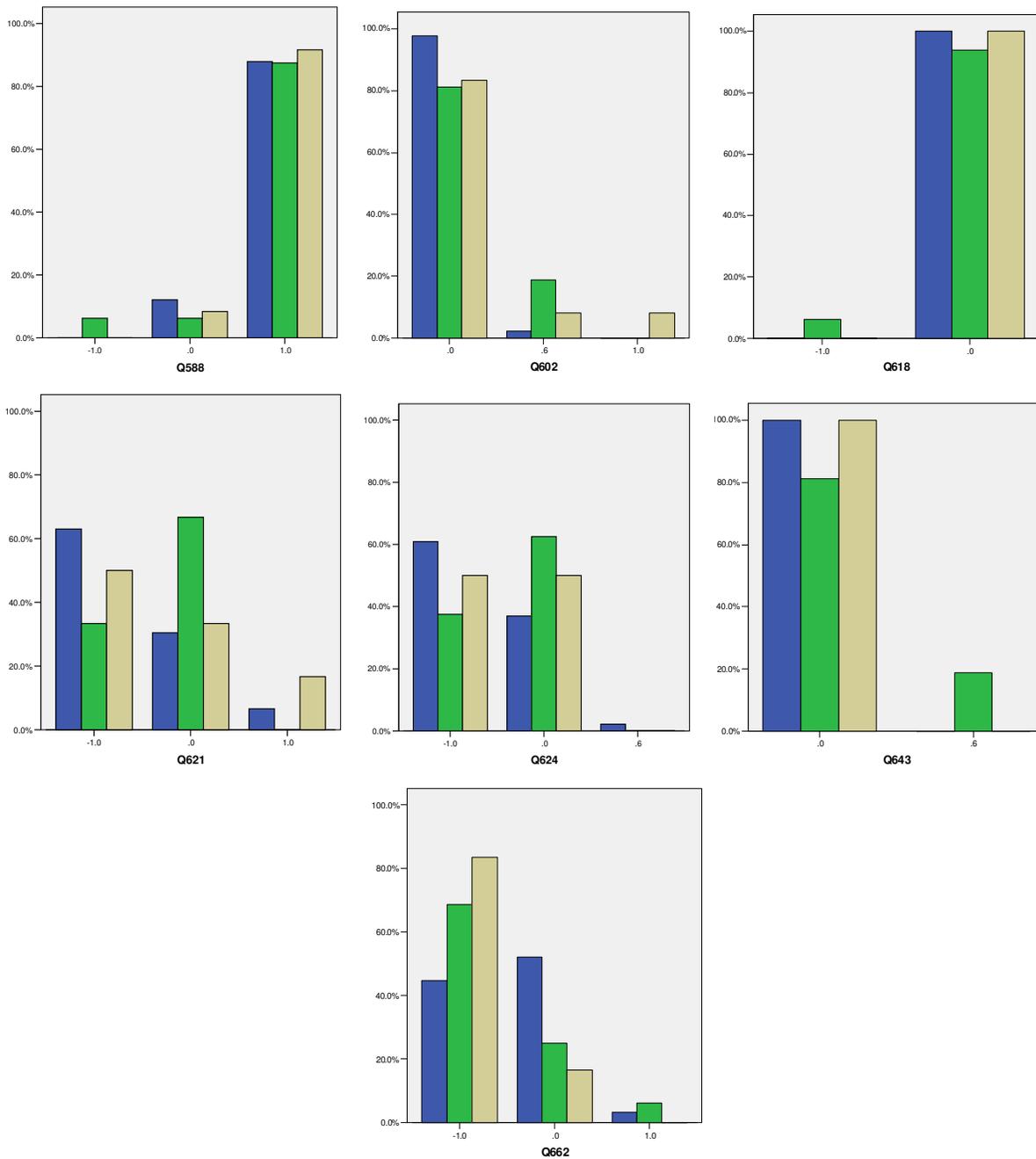


Figura 28. Distribuciones de frecuencia (eje vertical) de carga eléctrica Q (eje horizontal) de las posiciones estadísticamente significativas resultantes de las Tablas 6 y 7. Azul: R5, Verde: X4, Amarillo: R5X4.

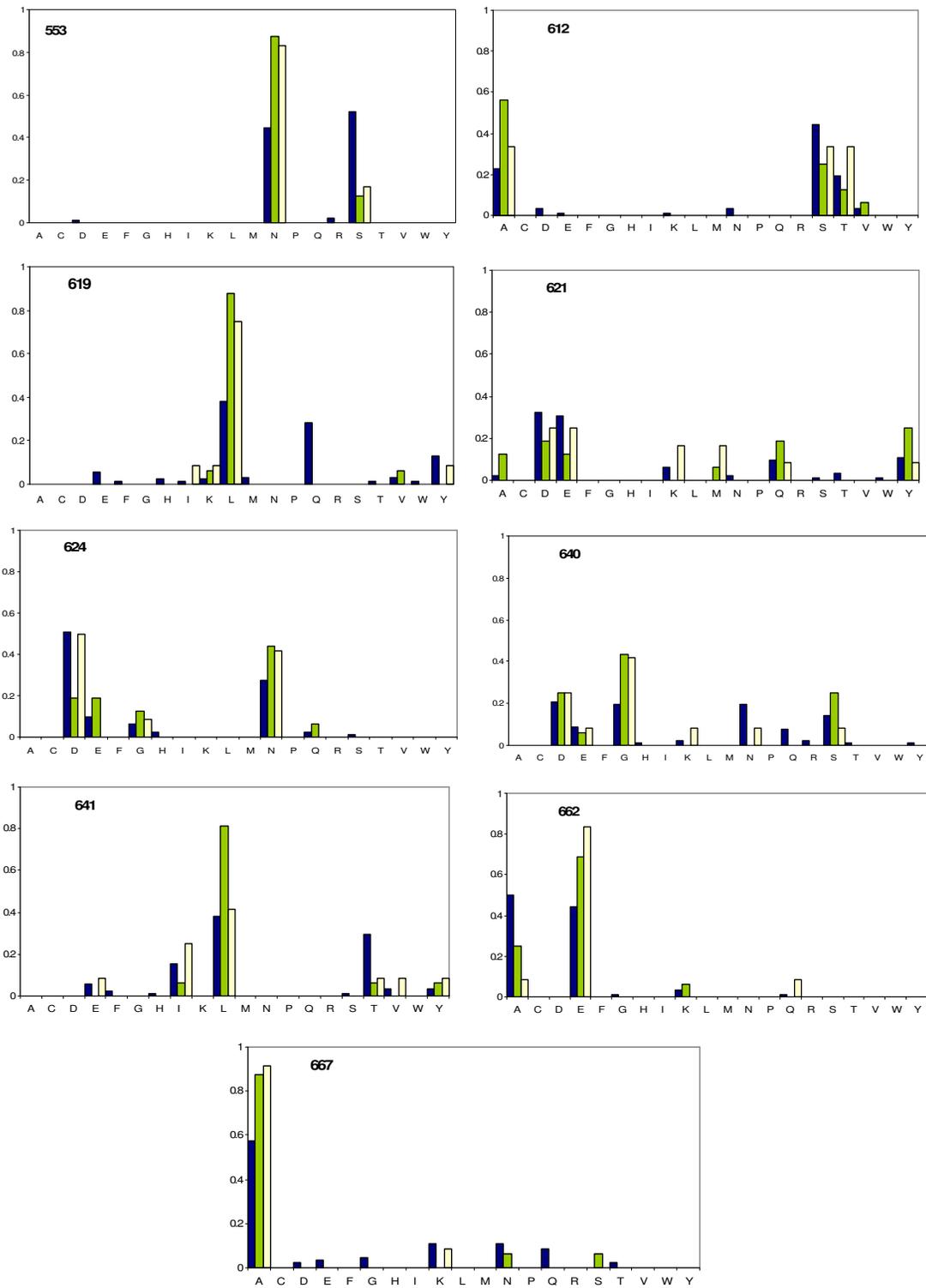


Figura 29. Gráficas de barra (eje vertical: frecuencia de aparición) que despliegan la composición de aminoácidos en orden alfabético de cada una de las 9 posiciones que son relevantes por su alta correlación con el tropismo viral. Estas posiciones también se enlistan en la Tabla 10 y Figura 33. Azul: R5, Verde: X4, Amarillo: R5X4.

Tabla 8. Arriba: valores p arrojados por las pruebas U de Mann-Whitney y χ^2 correspondientes a cada una de las 20 posiciones significativas según de las Tablas 4 y 5, sólo para las secuencias del subtipo B. Abajo: Grados de significancia (valores de p) obtenidos al aplicar las pruebas a los valores conjuntos de las parejas. El color cian y melón indican la propiedad considerada, hidropatía o carga, respectivamente.

TODOS LOS SUBTIPOS			
PAREJAS	Mann		
	R5vsX4	R5vsR5X4	X4vsR5X4
612-619(HI)	0.000043	0.01	0.287
619-621(HI)	0.001	0.135	0.427
621-624(Q)	0.015	0.226	0.509
640-641(HI)	0.016	0.313	0.677
662-667(HI)	0.712	0.726	0.567
χ^2			
	R5vsX4	R5vsR5X4	X4vsR5X4
612-619(HI)	0.000012	0.0403	0.111
619-621(HI)	0.00014	0.0083	0.5514
621-624(Q)	0.0011	0.440	0.091
640-641(HI)	0.119	0.444	0.638
662-667(HI)	0.798	0.725	0.643

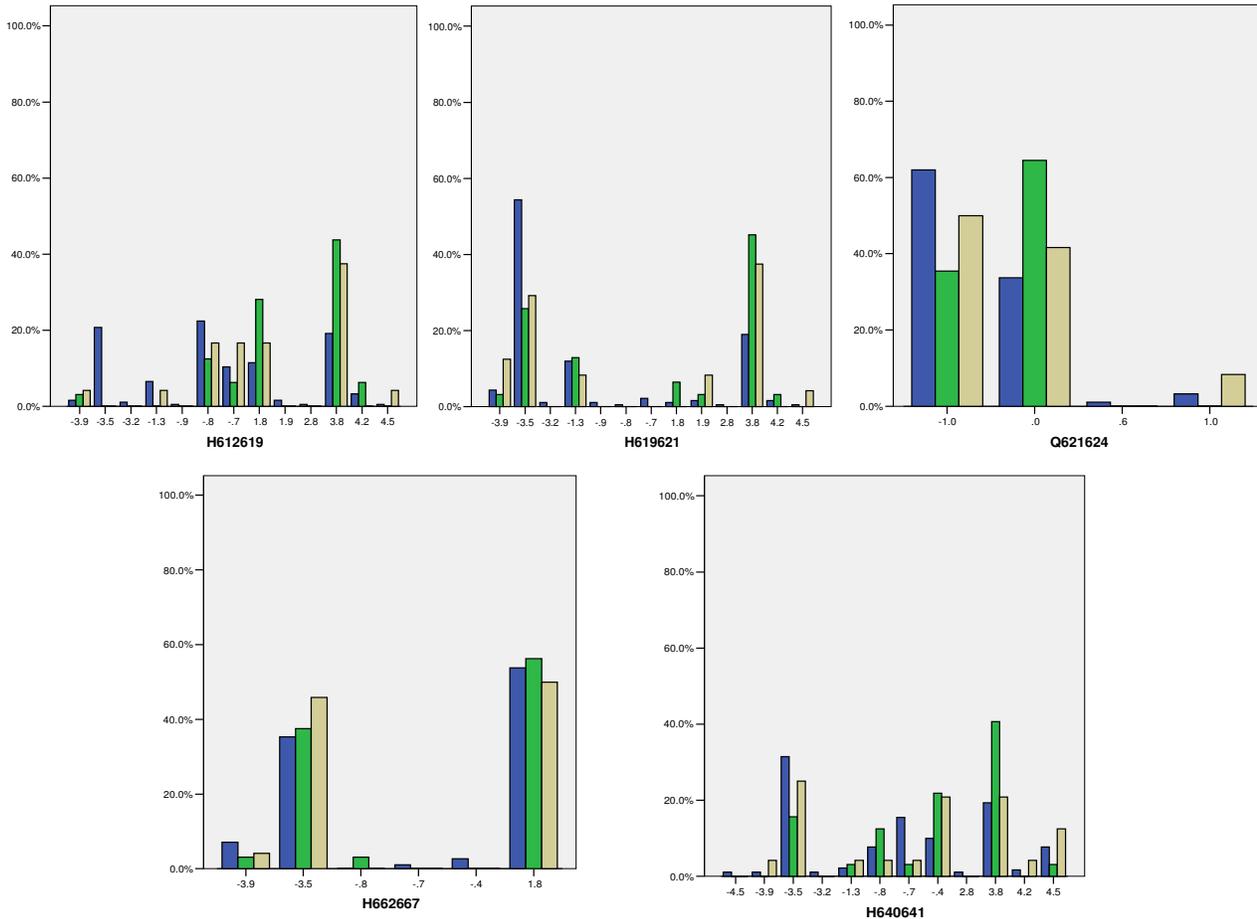


Figura 30. Distribuciones de frecuencias (eje vertical) de Hidropatía y Carga (eje horizontal) tomando en cuenta las parejas de todos los subtipos analizadas en la Tabla aquí mostrada. Azul: R5, Verde: X4, Amarillo: R5X4. Eje vertical: frecuencia porcentual. Eje horizontal: valores de índice hidropático o carga (según el caso) que se encontraron en cada posición, la leyenda bajo cada gráfico indica con una letra H o Q si es distribución de hidropatía o de carga respectivamente, mientras que los números indican las parejas consideradas.

Tabla 9. Arriba: valores p arrojados por las pruebas U de Mann-Whitney y χ^2 correspondientes a cada una de las 20 posiciones significativas según de las Tablas 4 y 5, sólo para las secuencias del subtipo B. Abajo: Grados de significancia (valores de p) obtenidos al aplicar las pruebas a los valores conjuntos de las parejas. Los colores cian y melón en cada renglón y las letras HI o Q entre paréntesis a la derecha de la numeración de aminoácidos indican resultados para hidropatía o carga respectivamente.

POSICIONES INDIVIDUALES SOLO SUBTIPO B						
Aminoácido	Mann			χ^2		
	R5vsX4	R5vsR5X4	X4vsR5X4	R5vsX4	R5vsR5X4	X4vsR5X4
535(HI)	0.230	0.24	0.935	-	-	-
553(HI)	0.828	0.458	0.414	-	-	-
561(HI)	0.039	1.000	0.394	0.037	-	0.381
578(HI)	1.000	1.000	1.000	-	-	-
580(HI)	0.000021	0.002	0.592	-	-	-
602(HI)	0.048	1.000	0.414	-	-	-
602(Q)	1.0	1.000	1.000	-	-	-
612(HI)	0.030	0.682	0.097	0.023	0.867	0.111
618(Q)	0.048	1.000	0.414	0.046	-	0.402
619(HI)	0.038	0.653	0.414	0.063	0.126	-
621(HI)	0.044	0.886	0.398	0.021	0.037	1
621(Q)	0.777	0.864	1.000	0.076	0.351	0.035
624(Q)	0.441	0.871	0.468	0.219	0.867	0.456
626(HI)	0.076	-	-	0.039	-	-
640(HI)	0.432	0.343	0.742	-	-	-
641(HI)	0.364	0.814	0.612	0.304	0.814	0.306
643(Q)	0.00048	1	0.135	0.00043	-	0.125
662(HI)	0.603	0.288	0.135	0.653	0.176	0.125
662(Q)	0.911	0.111	0.135	0.781	0.149	0.125
667(HI)	0.802	0.338	0.414	0.814	0.333	0.402

PAREJAS SOLO SUBTIPO B						
Parejas	Mann			χ^2		
	R5vsX4	R5vsR5X4	X4vsR5X4	R5vsX4	R5vsR5X4	X4vsR5X4
578-580(HI)	0.292	0.494	0.857	-	-	-
602-612(HI)	0.299	0.942	0.371	0.057	0.893	0.148
612-619(HI)	0.024	0.263	0.371	0.005	0.307	0.148
619-621(HI)	0.074	0.431	0.670	0.052	0.1	1.00
640-641(HI)	0.341	0.578	0.819	0.565	0.888	0.601
641-643(HI)	0.886	0.728	0.845	0.565	0.888	0.601
662-667(HI)	0.648	0.728	0.526	0.72	0.752	0.604
621-624(Q)	0.438	0.977	0.559	0.035	0.588	0.005

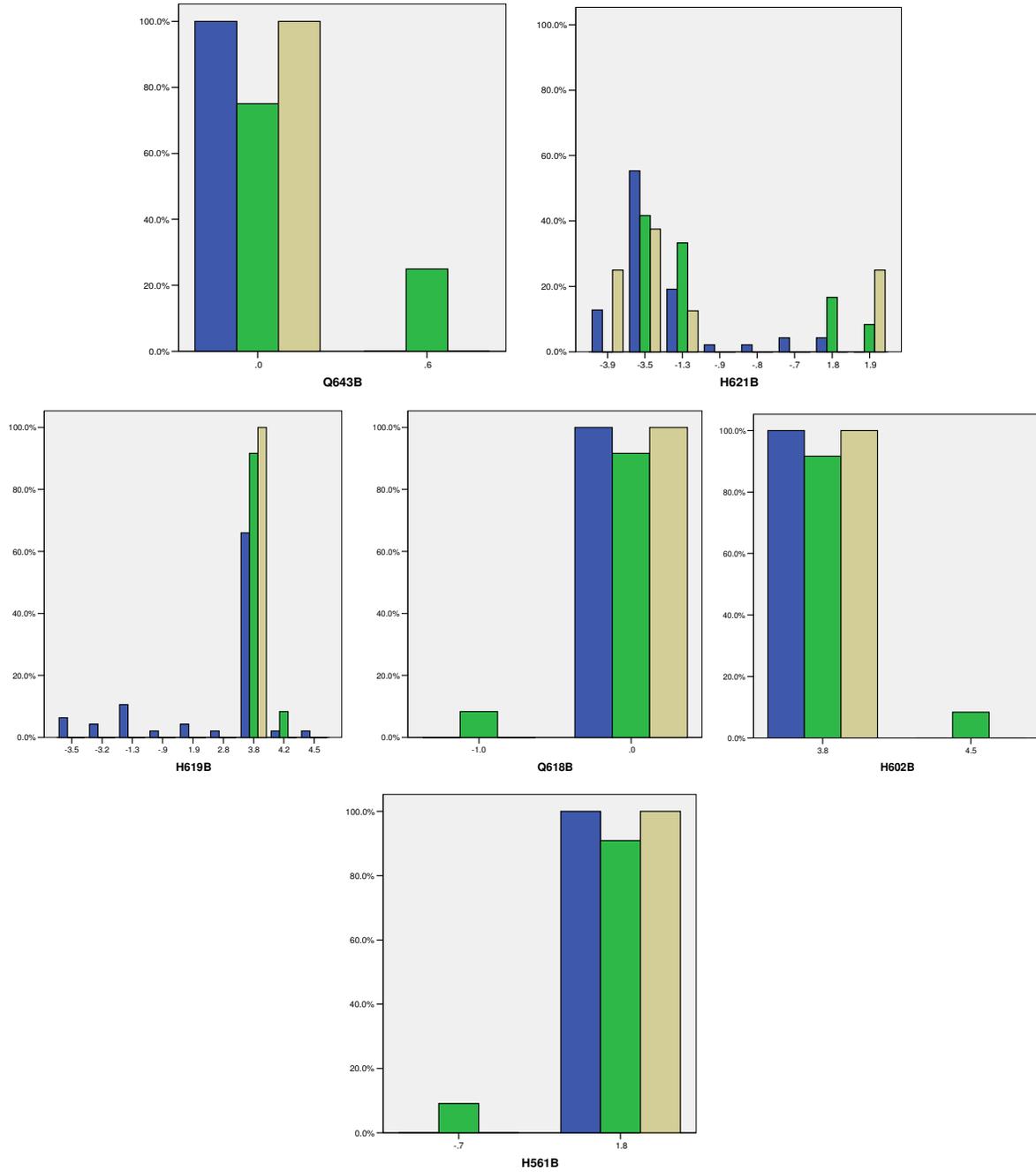


Figura 31. Distribuciones individuales de Carga e Hidropatía para las posiciones significativas del panel superior de la Tabla 9 (Subtipo B). Azul: R5, Verde: X4, Amarillo: R5X4. Eje vertical: frecuencia porcentual. Eje horizontal: valores de índice hidropático o carga (según el caso) que se encontraron en cada posición, la leyenda bajo cada gráfico indica con una letra H o Q si es distribución de hidropatía o de carga respectivamente, mientras que los números indican la posición considerada.

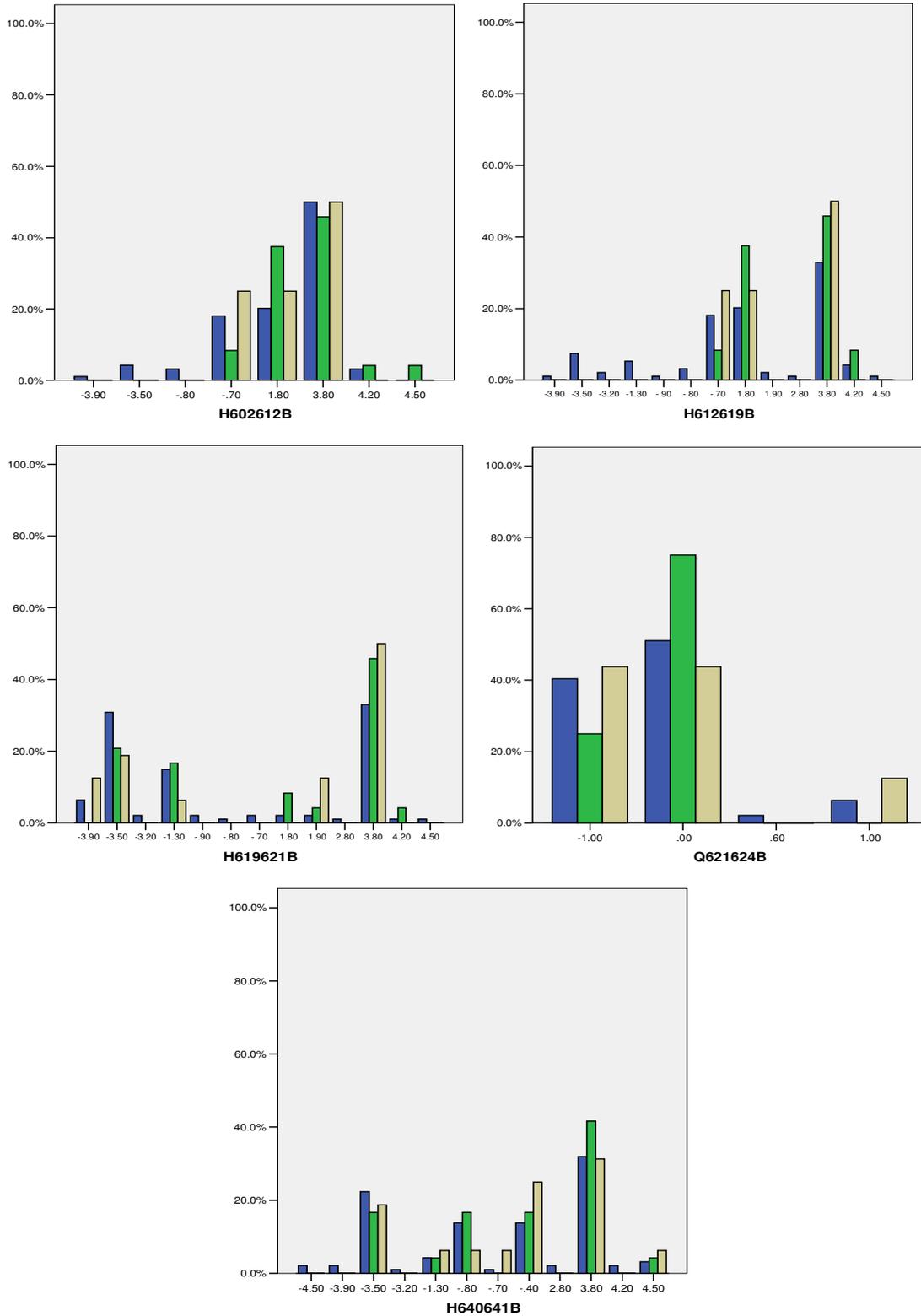


Figura 32. Distribuciones de Hidropatía y Carga para las parejas significativas del panel inferior de la Tabla 8 (Subtipo B). Azul: R5, Verde: X4, Amarillo: R5X4. Eje vertical: frecuencia porcentual. Eje horizontal: valores de índice hidropático o carga (según el caso) que se encontraron en cada pareja, la leyenda bajo cada gráfico indica con una letra H o Q si es distribución de hidropatía o de carga respectivamente, mientras que los números indican la pareja considerada.

Tabla 10. Posiciones más relevantes del ectodominio de gp41 que mostraron tener una alta correlación con el uso del correceptor. Se enlistan su localización espacial y observaciones más importantes.

Posición	Localización y disposición espacial en el complejo fusogénico	Observaciones
553	Hélice N. Parcialmente expuesta.	En R5 se sustituye N por S el 52% de los casos. En X4 es N el 87% de las veces.
612	Rizo.	En X4 hay una frecuencia alta de Alanina (apolar) mientras que en R5 es una Serina o Treonina (polares). R5 y mixtas son más hidrofílicas que X4. Está junto a un sitio de glicosilación. Podría participar en la unión con gp120. Posición no existente en los virus VIH-2 y VIS.
619	Rizo y comienzo de la hélice C. Totalmente expuesta.	Leucina 88% en X4 y 75% en R5X4. Esta frecuencia baja hasta 38% en R5. Experimentos realizados muestran que una sustitución por Alanina no influye en la entrada vira a pesar de que altera de manera importante la estructura de la cadena lateral [Jacobs et. al. 2005], lo que deja abierta la funcionalidad de este aminoácido en el tropismo infeccioso. Se encuentra totalmente expuesta.
621	Comienzo de la hélice C.	Aún cuando la distribución de aminoácidos es muy amplia, se observan patrones definidos de acumulación de carga, siendo los R5 y R5X4 más negativos y los X4 más neutros. Interactúa con una Q también variable (591) en la hélice N. Por su proximidad con la 619, podrían actuar en sinergia en el reconocimiento del correceptor.
624	Hélice C. Junto a un sitio de glicosilación.	Se observan patrones de distribución de carga diferentes entre R5 que tienden a ser cargados y X4 que tienden a ser neutros.
640	Hélice C. Muy expuesta.	R5 tiende a ser mucho más variable e hidrofílica, aunque los tres grupos presentan valores muy altos de entropía. Vecina inmediata de la 641, con la que podría actuar en sinergia.
641	Hélice C. Relativamente expuesta.	Distribución de aminoácidos similar a la posición 619. X4 es mucho más conservada, en R5 hay más Treonina T. Interacciona con una Leucina muy conservada (568) en la hélice N.
662	Hélice C. Expuesta.	Significamente distintas en carga y composición de aas. Alta frecuencia de E (cargado negativamente) en X4 y R5X4 y mayor frecuencia de Alanina (no polar) en R5. Espacialmente cercana a 658. X4 tiende a ser más hidrofílica que R5.
667	Final de la hélice C. Relativamente expuesta.	Las X4 y R5X4 tienden a ser más hidrofóbicas que R5 (87% vs. 58% de Alaninas respectivamente). Interacciona con la posición variable 546 (Serina la mayoría de las veces) en una hélice N.

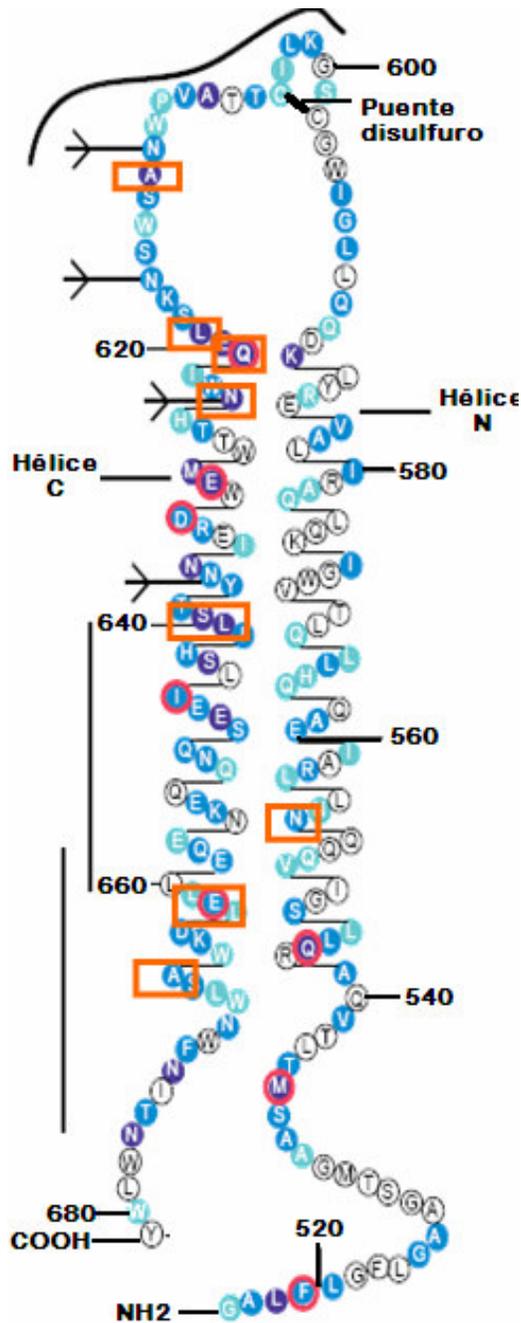


Figura 33. Diagrama del monómero de gp41. Se enmarcan en color naranja las posiciones de la tabla anterior (Tabla 9). Sólo una se encuentra en la hélice N (553) y el resto (ocho) en el lado amino terminal, una en el lazo (612), dos en el comienzo de la hélice N (619, 621), 3 dentro de ella (640, 641, 662), y una fuera (667), cercana a la parte transmembranal de la proteína. Las flechas del lado izquierdo indican los sitios de glicosilación.

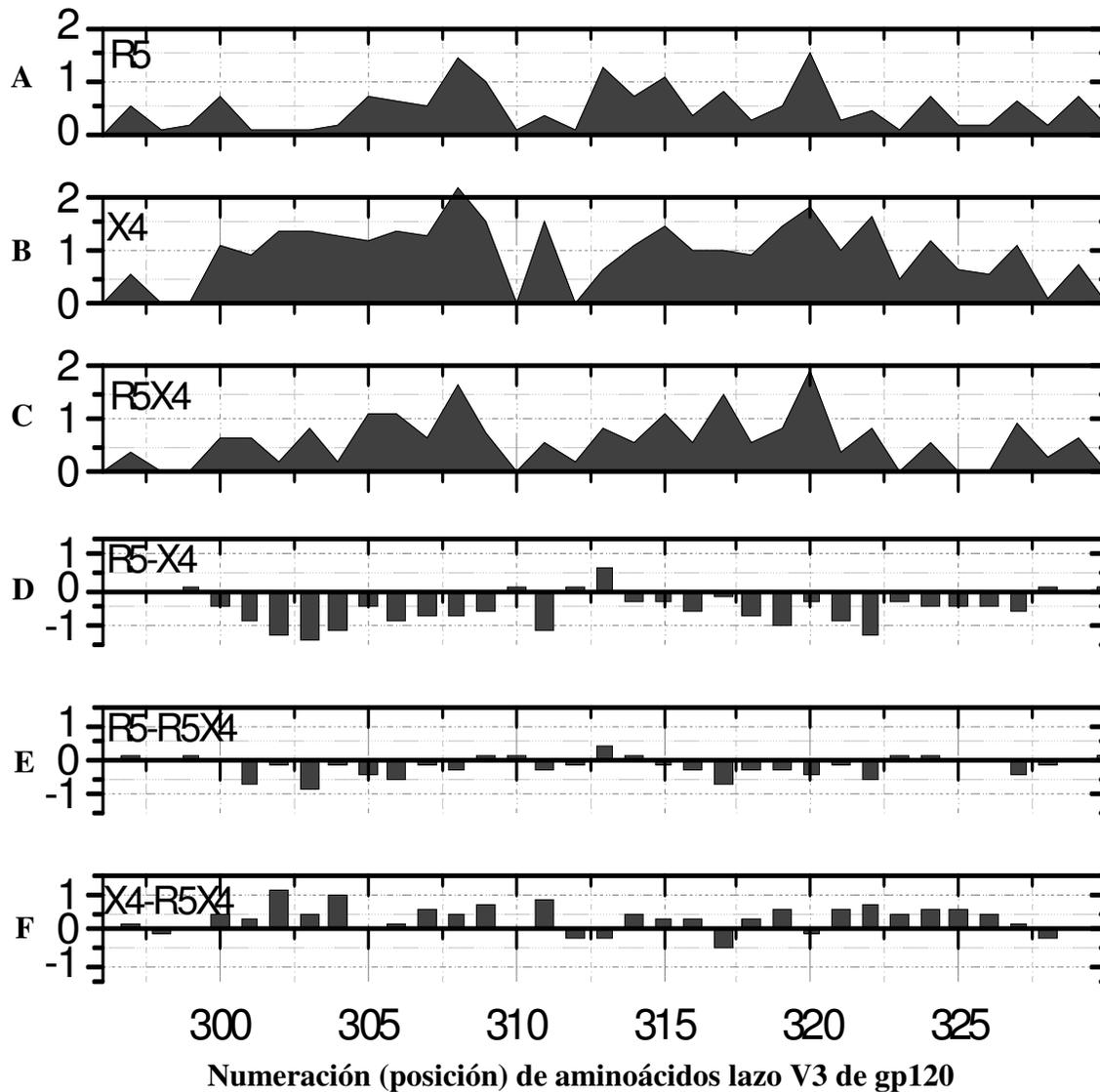


Figura 34. Gráficas A, B y C: distribución de entropías (eje vertical) por posición de aminoácidos (eje horizontal) del lazo hipervariable V3 de gp120 para las cepas virales tipo R5, X4 y R5X4, respectivamente. Gráficas D, E y F: distribución de las diferencias de entropía (eje vertical) entre cada dos grupos R5-X4, X4-R5X4 y R5-R5X4. Numeración de acuerdo a el gene precursor gp160 de la cepa de referencia del VIH: HXBc2.

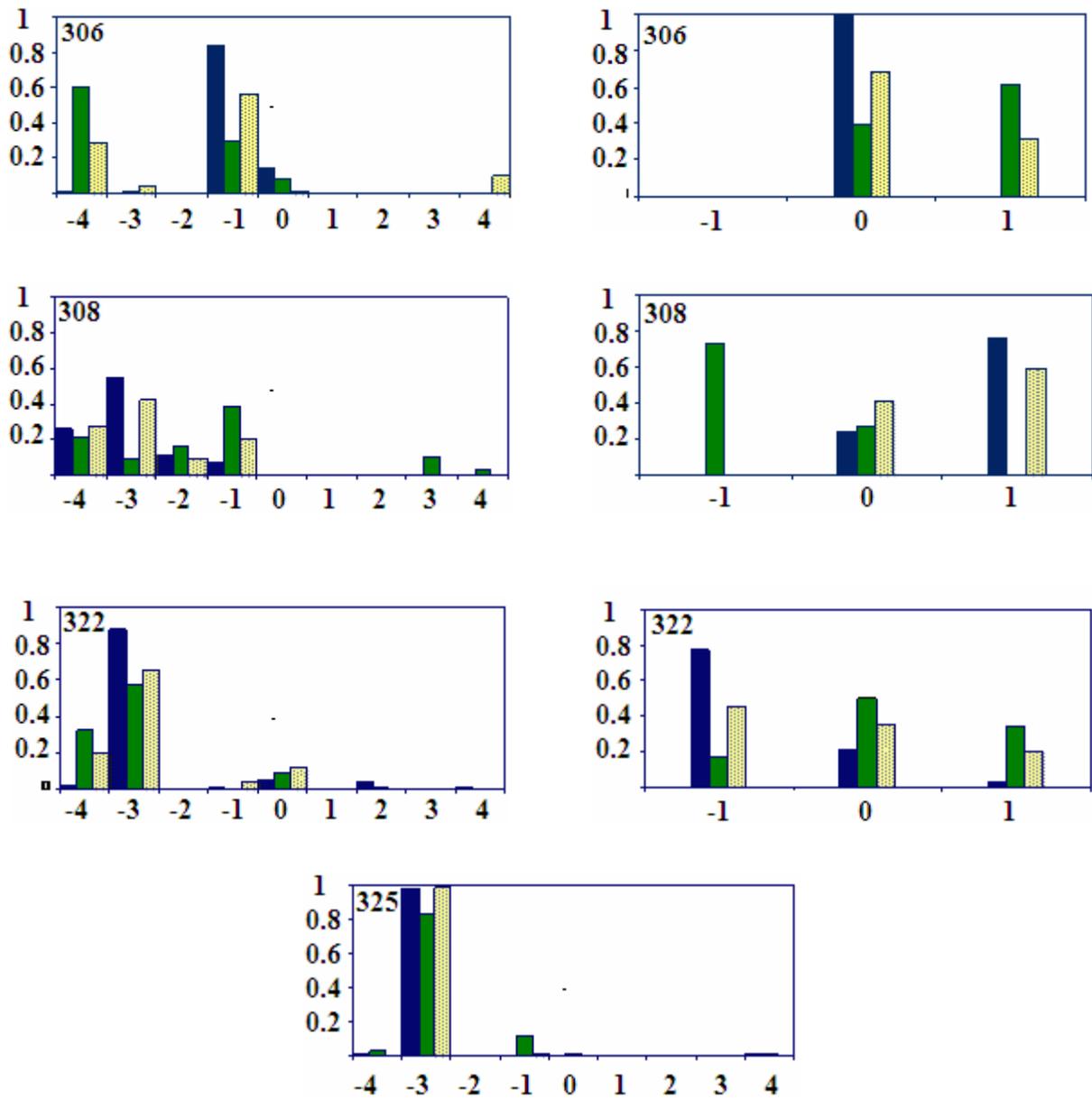


Figura 35. Histogramas de Hidropatía y Carga de las posiciones que resultaron ser más significativas en el análisis sobre Gp120-V3. Eje horizontal: índice hidropático [-4,4] o carga [-1,1], eje vertical: frecuencia de aparición. También se muestra una de las posiciones que resultaron ser significativamente espurias (325). En la esquina superior izquierda de cada gráfica se encuentra la posición respecto a HXBc2. Las posiciones 11 y 25 de la regla 11KR/25K corresponden a la 308 y 322 respectivamente. Azul: R5, Verde: X4, Amarillo: R5X4.

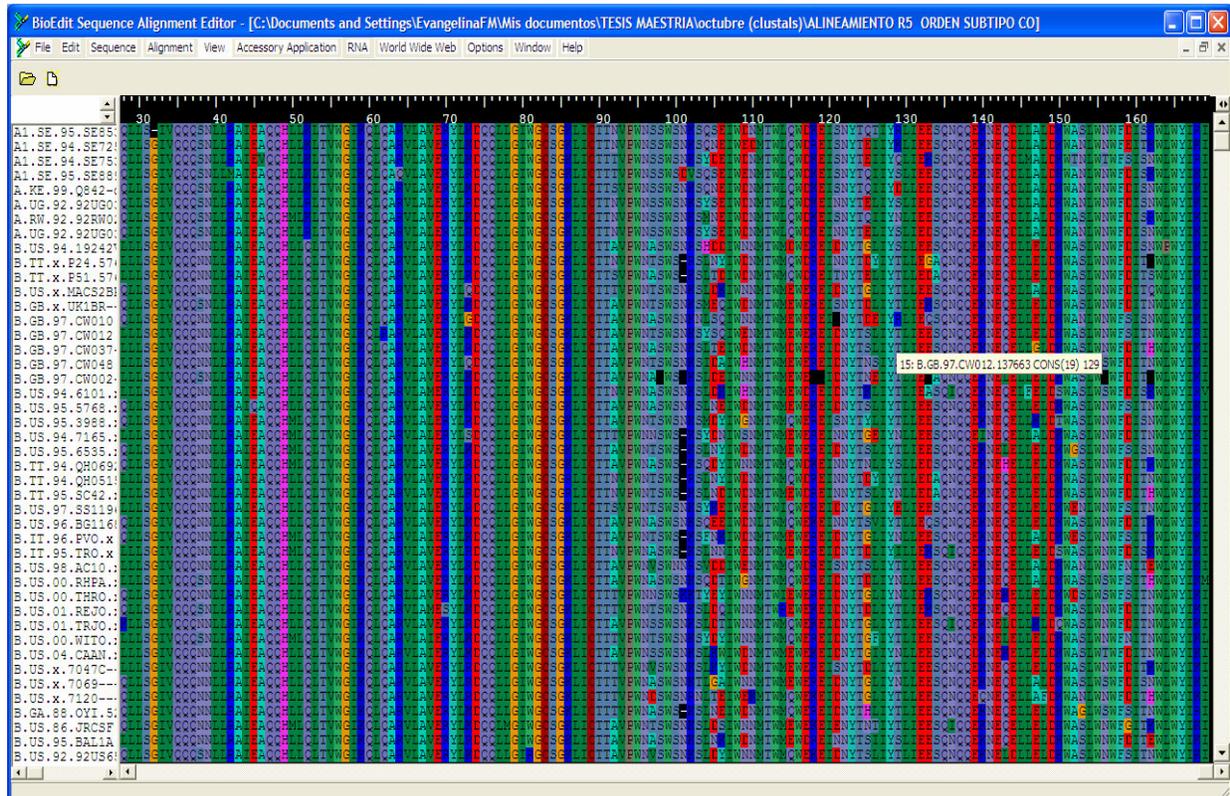
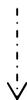


Figura 36. Hoja de trabajo en el programa *BioEdit 7.0.0* donde se muestran algunas de las secuencias alineadas y consensuadas del grupo R5. Del lado izquierdo se observa el código de identificación de cada secuencia. Los colores sólo intentan diferenciar entre un aminoácido y otro según sus propiedades fisicoquímicas. Se puede apreciar la variabilidad de aminoácidos entre una secuencia y otra, desde la posición (columna) 100 hasta la 170 aproximadamente, región correspondiente a la región de la hélice C de Gp41-e; mientras que las columnas de la izquierda representan sitios muy conservados correspondientes a la región de la hélice N.

Tabla 11 (siguientes cuatro páginas). Secuencias pertenecientes a los grupos X4, R5X4 y R5 respectivamente, después de haber hecho los consensos (indicados con la terminación CONS(<número de cepas consensuadas>) en el código de cada aislado). Los residuos van desde la posición 516 hasta la 684 del ectodominio de la proteína gp41 del VIH-1 (de acuerdo a la numeración de gp160 de la cepa de referencia HXBc2).



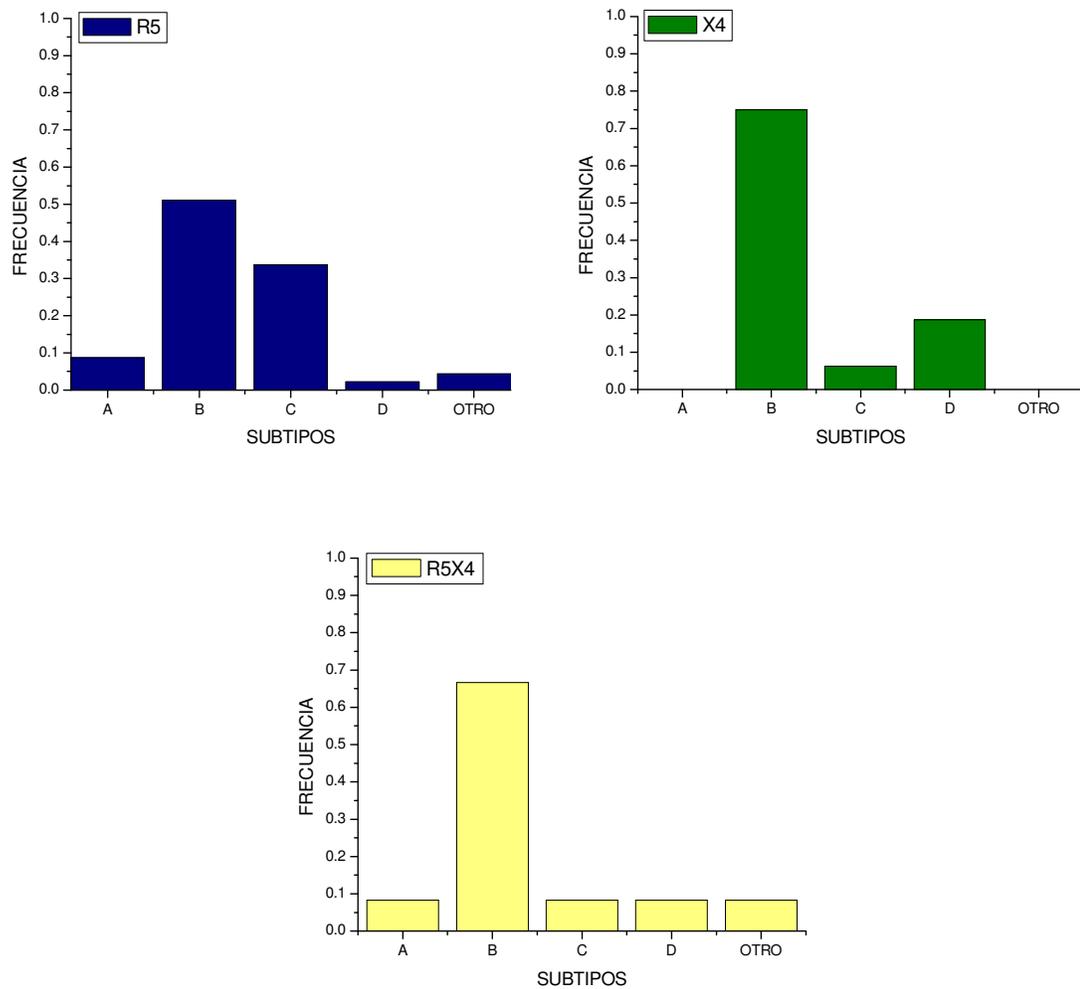


Figura 38. Distribución de subtipos dentro de cada grupo de secuencias de Gp41-e. Nótese en el grupo R5 la altura del subtipo B y C en comparación con los grupos X4 y R5X4.

Tabla 12. Tendencia en hidropatía (posiciones en color cian) y carga (posiciones en color melón) de cada subtipo sin tomar en cuenta el grupo de tropismo de las secuencias de Gp41. El guión entre paréntesis indica carga negativa y el “0” un índice hidropático igual a cero.

Subtipo	619	621	624	640	641
A	H-fílicos	(-)	(-)	H-fílicos	H-fóbicos
B	H-fóbicos	neutros	(-) & neutros	H-fílicos & 0	H-fóbicos
C	H-fílicos	(-)	(-)	H-fílicos	H-fílicos
D	H-fóbicos	(-)	neutros	0	H-fóbicos

1 HIV Sequence Search, Los Alamos National Laboratory. http://www.hiv.lanl.gov/components/hiv-db/combined_search_s/search.html

2 Schneider, T.D. and R.M. Stephens. Sequence Logos: A new Way to Display Consensus Sequences. *Nucleic Acids Res.* 1990 18: 6097-6100.

3 Thompson, J.D., Higgins, D.G. and Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994 22: 4673-4680.

4 Shannon, C.E. A Mathematical Theory of Communication. *The Bell System Technical Journal* 1948 27: 379-423, 623-656.

5 Pierce, J.R. *An Introduction to Information Theory: Symbols, Signals and Noise*, Dover Publications, Inc., 1980. New York. second edition.

Discusión

Analizando las propiedades fisicoquímicas de los aminoácidos del ectodominio de Gp41 de un conjunto de secuencias del VIH-1 pudimos identificar 9 posiciones que experimentan sustituciones no conservativas de tal forma que están relacionadas con el tropismo viral. Siete de esas posiciones están localizadas en un fragmento de ~50 aminoácidos comprendidos entre la parte C-terminal del lazo y la región de la hélice C que concentra residuos glicosilados y altamente variables. Se identificó también una posición más en la hélice N y otra cercana al segmento transmembranal. Además, cuatro de esas posiciones están asociadas de manera bastante significativa con el uso del correceptor en el subtipo B, lo cual permite pensar que podrían representar un papel determinante en la interacción con el correceptor en el ectodominio de Gp41 (Gp41-e).

Los resultados obtenidos en este trabajo indican que la interacción de *Env* con CXCR4 puede estar asociada con una tendencia hidrofóbico/no cargado de aminoácidos individuales y aminoácidos acompañados (cúmulos) en la parte C-terminal del lazo y la región de la hélice C de Gp41-e. Por otro lado, los residuos hidrofílicos/cargados negativamente podrían influenciar la interacción de *Env* con CCR5.

Resulta interesante que las posiciones 619-621 comprenden la región más variable de Gp41-e, aunque la posición 620 experimenta sustituciones conservativas que mantienen el carácter altamente hidrofílico del residuo en ese sitio.

Las sustituciones aminoacídicas descritas aquí no alteran el patrón de glicosilación característico de esta región, que se ha asociado a la evasión inmune¹.

Respecto al análisis de las divergencias en la entropía, las posiciones 619, 640 y 641 mostraron los valores más altos entre R5 y X4, lo cual refleja una variabilidad más alta de las cepas R5 comparadas con las X4 y R5X4. Lo anterior es un comportamiento congruente con el hecho de que los virus R5 dominan en un período de tiempo largo en los individuos infectados, teniendo la oportunidad de mutar para adaptarse a las condiciones inmunológicas y fisiológicas variables durante ese tiempo. El caso de X4 es contrario, pues sus cepas se detectan una fracción de alrededor de un 50% de los individuos infectados, lo que sugiere que requieren de un medio más restringido. El caso del grupo R5X4 resultó interesante debido a que en la mayoría de las posiciones relevantes estuvo relacionado con X4, lo cual es consistente con observaciones sobre el parecido de R5X4 con X4 asociado con una progresión a la enfermedad².

En el contexto del modelo de fusión de membranas inducido por *Env*, nuestros resultados sugieren que algunos residuos de Gp41-e podrían contribuir en la elección del correceptor por algún mecanismo indirecto como influenciar la conformación de los sitios de unión al correceptor en gp120, o por medio de un contacto directo con la molécula correceptora, ya sea formando un sitio de enlace en conjunción con gp120 o independiente de ésta en un paso subsiguiente en el proceso de fusión.

Se ha confirmado que un área bastante extensa de Gp41-e está involucrada en la interfase no covalente con la subunidad gp120 en el estado innato³², sugiriendo la hipótesis de que la hidropatía y carga de los aminoácidos en dicha interfase influenciarían la conformación local de gp120 y por lo tanto su interacción con la molécula correceptora. De hecho, recientemente varios grupos han reportado que T-20, un péptido compuesto por una fracción de la secuencia de la hélice C de Gp41-e, se une a gp120 en una región cercana a la base del

lazo V3³. Lo anterior es consistente con la región de Gp41-e donde se encontraron los sitios correlacionados con el tropismo del VIH-1.

La naturaleza transitoria de los cambios conformacionales de Gp41 inducidos después del acoplamiento de gp120 y su interacción con CD4 permanecen aún desconocidos. Sin embargo, después de la interacción de gp120 con CD4, la estructura metaestable de Gp41 es liberada y los residuos que antes se encontraban ocluidos estarían expuestos y en contacto cercano con la membrana blanco⁴.

La unión a las regiones N o C de los péptidos inhibidores demuestra que ambas regiones están accesibles antes de la formación post-fusión del paquete de seis hélices³² por lo que en este corto período de tiempo y estado poco entendido de Gp41, bien puede establecer interacciones directas con los receptores membranales. Las regiones en la parte amino terminal y los lazos extracelulares de CCR5 y CXCR4 son críticas para la función del correceptor, lo que sugiere que tales dominios son importantes para la inducción de los cambios conformacionales fusogénicos en *Env*^{5,6}.

Es sumamente interesante notar que CXCR4 contiene una especie de surco en la región que se localiza entre el segundo y tercer lazo, definido por las secuencias LWWVF en ECL2 (segundo lazo extra celular, de las siglas en inglés Extra Cellular Loop) y SFILLEII en ECL3. Estas regiones podrían representar posibles sitios para la interacción con residuos hidrofóbicos de Gp41-e en virus X4. Esta hipótesis está soportada por observaciones de que las interacciones que determinan la actividad del correceptor del ECL2 de CXCR4 no son de carácter electrostático ya que son independientes de los residuos cargados⁷, y los virus X4 tienden a contener aminoácidos neutros en los sitios que resultaron relevantes en nuestro análisis. De manera contraria, CCR5 contiene una región muy hidrofílica y cargada positivamente (QWKNFQ) en ECL2 y (NNLGFFEQ) en ECL3, lo que sugiere que los virus R5 podrían tener la capacidad de interactuar con esta región ya que sus aminoácidos tienden a ser hidrofílicos y a estar cargados negativamente. Más aún, debido a que CCR5 contiene residuos con carga positiva en la parte amino terminal situados en el fondo del surco que forman ésta sección con el ECL1, y cargados negativamente en la sección superior a la anterior (sección V.5), nuestros resultados sugieren que la carga negativa propia de los aminoácidos presentes en las posiciones 621 y 624 del grupo R5 de gp41-e, podrían repelerse con el extremo amino negativo de CCR5 y abrirse paso para ser atraídos por las cargas positivas que se encuentran al fondo, cercanas a la parte transmembranal.

Aunque en el grupo R5 el subtipo B está presente sólo en un 47%, el análisis sobre este subtipo arrojó el mismo comportamiento que el primer análisis donde se incluyeron todos los subtipos. Esto debido a que la significancia observada en los sitios relevantes continua observándose aún dentro del subtipo B. Además es de suma importancia conocer el comportamiento restringido a este subtipo ya que es el preponderante en el continente americano.

Por último, el método que se utilizó para encontrar sitios clave asociados con el tropismo viral, se pudo valorar aplicándolo también a Gp120-V3 y comparando los resultados obtenidos para Gp41-e con estos últimos, quedando clara la consistencia y similitud del comportamiento observado en ambas regiones de las proteínas de envoltura.

1 Johnson, W. E., J. M. Sauvron, and R. C. Desrosiers. Conserved, N-linked carbohydrates of human immunodeficiency virus type 1 gp41 are largely dispensable for viral replication. *J. Virol.* 2001 75:11426–11436.

2 Simmons G, Wilkinson D, Reeves JD, Dittmar MT, Beddows S, Weber J, Carnegie G, Desselberger U, Gray PW, Weiss RA, Clapham PR. Primary, syncytium-inducing human immunodeficiency virus type 1 isolates are dual-tropic and most can use either Lestr or CCR5 as coreceptors for virus entry. *J Virol.* 1996 70(12):8355-60.

3 Yuan W, Craig S, Si Z, Farzan M, Sodroski J. CD4-induced T-20 binding to human immunodeficiency virus type 1 gp120 blocks interaction with the CXCR4 coreceptor. *Journal of Virology* 2004, Vol. 78, No. 10 pp. 5448-5457

4 S. Wang, J. York, W. Shu, O. Stoller, H. Nunberg and M. Lu, Interhelical Interactions in the gp41 Core: Implications for Activation of HIV-1 Membrane Fusion, *Biochemistry* 2002 41, 7283-7292.

5 Dragic T, Trkola A, Lin SW, Nagashima KA, Kajumo F, Zhao L, Olson WC, Wu L, Mackay CR, Allaway GP, Sakmar TP, Moore JP, Maddon PJ. Amino-terminal substitutions in the CCR5 coreceptor impair gp120 binding and human immunodeficiency virus type 1 entry. *J Virol.* 1998 72(1):279-85.

6 Wu L, LaRosa G, Kassam N, Gordon CJ, Heath H, Ruffing N, Chen H, Humblias J, Samson M, Parmentier M, Moore JP, Mackay CR. Interaction of chemokine receptor CCR5 with its ligands: multiple domains for HIV-1 gp120 binding and a single domain for chemokine binding. *J Exp Med.* 1997 186(8):1373-81.

7 Wang, W.-K., T. Dudek, Y. J. Zhao, H. Brumblay, M. Essex, and T.-H. Lee. CCR5 co-receptor utilization involves a highly conserved arginine residue of human immunodeficiency virus type 1 gp120. *Proc. Natl. Acad. Sci. USA* 1998 95:5740-5745.

Conclusiones

Las posiciones que mostraron estar más correlacionadas con el tropismo infeccioso del VIH-1 fueron la **612**, **619** y **641**, además de las parejas **619/621** en el índice hidropático y **621/624** para la carga, esto independiente de si se consideran todos los subtipos o sólo el B. En general se encontró una tendencia de los virus X4 por tener aminoácidos hidrofóbicos y neutros, mientras que los del virus R5 tienden a ser más hidrofílicos y cargados negativamente. Por lo anterior se concluye que la subunidad transmembranal del virus, Gp41, está involucrada en el proceso de reconocimiento del correceptor. Estas posiciones podrían participar en el “switch fenotípico” R5→X4 asociado a una progresión más rápida hacia el SIDA.

También la modulación de la función de reconocimiento de gp120 por distintas regiones de gp41 podría ser un factor potencial en la elección del correceptor.

Se encontraron sustituciones no conservativas asociadas al tropismo independientemente del subtipo genético, ya que la distribución de subtipos en los grupos fue de 75% B en X4, mientras que los subtipos A y C son casi exclusivos de las cepas R5. El comportamiento observado dentro del subtipo B comprueba que no es la cantidad de secuencias de éste subtipo en cada grupo lo que determina las diferencias observadas sobre todas las secuencias entre los tres grupos de tropismo analizados.

Estudios recientes indicaron que el fenotipo del virus es dominante respecto del genotipo en relación al uso del correceptor y que los virus X4 eran muy improbables de encontrar en el subtipo C, dando pie a suponer que el genotipo puede otorgar propiedades biológicas importantes tales como virulencia, tropismo por tejidos y transmisión¹. Sin embargo, estudios más recientes encontraron una frecuencia más alta de virus X4 en el subtipo C^{2,3}. Por otro lado, se encontraron virus con diferentes tropismos dentro del subtipo B. En todos los casos los pacientes con virus X4 o R5X4 fueron sintomáticos o tuvieron un peor pronóstico que los pacientes con virus R5. Por lo tanto, aunque la secuencia de aminoácidos refleja relaciones filogenéticas, la presión selectiva de ambientes fisiológicos e inmunológicos parecen ser cruciales en el desarrollo de una variabilidad hacia el equilibrio.

Algunas propiedades fisicoquímicas de aminoácidos de *Env* parecen representar mutaciones adaptables determinadas por características patogénicas. En este sentido, nuestros hallazgos sugieren que los residuos altamente correlacionados pueden participar en la determinación del tropismo infeccioso del VIH-1 y en características patogénicas tanto en el subtipo B como en el resto. Esto debido a que las características de los aminoácidos presentes en las cepas R5-B son similares a las de las cepas C (que son la mayoría R5), mientras que las cepas X4-B son similares a las cepas D (que en nuestra muestra son todas X4).

Es interesante notar que la posición 612 no existe en el VIH-2 y en el VIS (Virus de Inmunodeficiencia de los Simios) los cuales son mucho menos agresivos que el VIH-1 y su capacidad infecciosa es mucho menor.

Estas observaciones pueden agregar una pieza de evidencia en el estudio del mecanismo de fusión de membranas inducido por el VIH, así como apoyar el pronóstico clínico de los pacientes mediante la determinación del tropismo viral y complementar la predicción del fenotipo viral basada en el lazo V3 de gp120, en donde al aplicar el mismo análisis encontramos resultados consistentes.

Como trabajo posterior, sería interesante verificar si anticuerpos monoclonales o alguna proteína dirigida hacia los aminoácidos en los sitios 612, 619, 621, 624 y 641 son capaces de inhibir la fusión o si las mutaciones del carácter hidropático de dichos aminoácidos son capaces de alterar el tropismo infeccioso viral, lo cual daría pie a posibles aplicaciones terapéuticas.

1 Zhang, H., G. Orti, Q. Du, J. He, C. Kankasa, G. Bhat, and C. Wood, Phylogenetic and phenotypic analysis of HIV type 1 env gp120 in cases of subtype C mother-to-child transmission, *AIDS Res. Hum. Retroviruses* 2002 18: 1415-1423.

2 Cilliers, T., J. Nhlapo, M. Coetzer, D. Orlovic, T. Ketas, W. C. Olson, J. P. Moore, A. Trkola, and L. Morris. The CCR5 and CXCR4 Coreceptors Are Both Used by Human Immunodeficiency Virus Type 1 Primary Isolates from Subtype C. *J. Virol.* 2003 77: 4449-4456.

3 Pollakis, G., S. Kang, A. Kliphuis, M. I. M. Chalaby, J. Goudsmit, and W. A. Paxton. N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization. *J. Biol. Chem.* 2001 276: 13433-13441.

Apéndice A

Proteínas: conceptos básicos^{1,2}

Cuando miramos una célula al microscopio o analizamos su actividad eléctrica o bioquímica, en esencia estamos observando proteínas. Éstas constituyen la mayor parte del peso seco de una célula y ejecutan prácticamente todas las funciones celulares.

Desde un punto de vista químico, las proteínas son las moléculas estructuralmente más complejas y funcionalmente más sofisticadas que se conocen. Esto quizá no resulta sorprendente cuando uno se da cuenta de que la estructura y las propiedades químicas de cada proteína se han ido desarrollando y afinando a lo largo de miles de millones de años de historia evolutiva.

En las proteínas hay 20 tipos de aminoácidos, cada uno con propiedades fisicoquímicas diferentes (Tabla A1). Una molécula de proteína está formada por una larga cadena de estos aminoácidos, cada uno de los cuales está unido a su vecino mediante un enlace covalente peptídico. Cada tipo de proteína tiene una secuencia única de aminoácidos, que debe ser exactamente la misma de una molécula a otra del mismo tipo.

La secuencia repetida de átomos a lo largo de la cadena se conoce como esqueleto polipeptídico. Unidas a esta cadena repetitiva están aquellas partes de los aminoácidos que no participan en la formación del enlace peptídico y que confieren a cada uno de los aminoácidos sus propiedades características: las 20 cadenas laterales diferentes. Algunas de estas cadenas laterales son apolares e hidrofóbicas (que presentan una repulsión por la molécula del agua), otras están cargadas positiva o negativamente, algunas son reactivas, etc. En la Tabla A1 se muestran sus estructuras atómicas y las propiedades esenciales de cada aminoácido con sus abreviaturas.

Los átomos se comportan casi como si fueran pesadas esferas de un radio determinado (su radio de van der Waals). El requisito de que dos átomos no pueden solaparse limita el número posible de ángulos de enlace en una cadena polipeptídica. Estas fuerzas y otras interacciones estéricas reducen severamente el número de disposiciones tridimensionales (o conformaciones) posibles. A pesar de ello, una larga cadena flexible, como la de una proteína, puede plegarse en un número enorme de posibilidades.

Sin embargo, el plegamiento de una cadena proteica también está condicionado por muchos tipos diferentes de enlaces no covalentes débiles, formados entre partes de la propia cadena. Por tanto, la estabilidad de cada forma plegada se verá afectada por la fuerza combinada de un gran número de estos enlaces.

Existe una cuarta fuerza débil que también desempeña un papel central en la determinación de la forma de una proteína. Las moléculas hidrofóbicas, incluyendo las cadenas laterales apolares de algunos aminoácidos, tienden a unirse unas a otras cuando se hallan en un entorno acuoso, minimizando así su efecto interruptor de la red de moléculas de agua unidas entre sí por enlaces de hidrógeno. Por lo tanto, un factor importante que condiciona el plegamiento de cualquier proteína es la distribución de sus aminoácidos polares y apolares. Las cadenas laterales apolares (hidrofóbicas) de una proteína –correspondientes a aminoácidos tales como la fenilalanina, la leucina, la valina o el triptófano- tienden a aglutinarse en el interior de la molécula (de manera parecida a como las gotitas de aceite coalescen formando una sola gota grande). Ello les permite evitar el contacto con el agua que

las rodea en el interior de la célula. Por el contrario, las cadenas laterales polares –como las correspondientes a la arginina, la glutamina o la histidina- tienden a disponerse cerca de la superficie exterior de la molécula, donde pueden formar enlaces de hidrógeno con el agua y con otras moléculas polares. Generalmente, cuando los aminoácidos polares se esconden dentro de una proteína, establecen enlaces de hidrógeno con otros aminoácidos polares o con el esqueleto polipeptídico.

Como resultado de todas estas interacciones, cada tipo de proteína tiene una estructura tridimensional determinada por el orden de los aminoácidos de su cadena. Generalmente, la estructura plegada final, o conformación, que adopta una cadena polipeptídica es la que minimiza la energía libre de Gibbs.

La estructura de las proteínas es muy compleja, razón por la cual resulta conveniente describirla considerando distintos niveles de organización:

Estructura primaria: se refiere al número e identidad de los aminoácidos que componen la molécula y al ordenamiento o secuencia de esas unidades en la cadena polipeptídica. La unión peptídica sólo permite formar estructuras lineales; por ello, las cadenas no presentan ramificaciones.

Estructura secundaria: a medida que la longitud de las cadenas va aumentando y en función de las condiciones fisicoquímicas del medio, se conforma este nivel de estructura, que es la disposición espacial regular, repetitiva, que puede adoptar la cadena polipeptídica, generalmente mantenida por enlaces de hidrógeno. En este caso suelen darse dos tipos de ordenamiento local (Figura A1):

Hélice alfa. Las cadenas de aminoácidos tienen varios centros polares y, debido a esto, la fibra suele enrollarse dando lugar a una hélice que se estabiliza formando puentes de hidrógeno intramoleculares.

Lámina beta. Las cadenas de péptidos se unen formando filas paralelas que se estabilizan de manera intermolecular mediante puentes de hidrógeno.

Estructura terciaria: ésta es la estructura de la mayoría de las proteínas globulares, aparece a partir de que la hélice alfa se vuelve a enrollar. Es una arquitectura tridimensional completa que se debe a las fuerzas de atracción o repulsión electrostática, a enlaces de hidrógeno, a fuerzas de Van der Waals y a puentes disulfuro.

Estructura cuaternaria: son estructuras de carácter oligomérico, que están compuestas por varias cadenas separadas pero entrelazadas en estructura terciaria. Se aplica sólo a proteínas constituidas por dos o más cadenas polipeptídicas y se refiere a la disposición espacial de esas cadenas y a los enlaces que se establecen entre ellas (puentes de hidrógeno, atracciones electrostáticas, interacciones hidrofóbicas, puentes disulfuro entre cisteínas de cadenas diferentes, etc.). Un ejemplo de este tipo de estructura es la hemoglobina que está compuesta por cuatro subunidades de mioglobina.

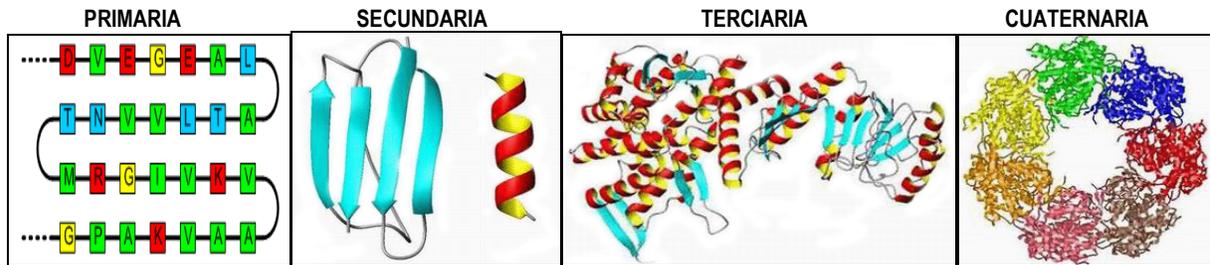


Figura A1. De izquierda a derecha: los cuatro niveles de estructura en las proteínas.

A.1 Las hélices alfa y láminas beta: patrones comunes de plegamiento

Cuando se comparan las estructuras tridimensionales de muchas moléculas proteicas, se hace evidente que, aunque la conformación global de cada proteína es única, a menudo se encuentran dos patrones de plegamiento en algunas zonas de las proteínas. Ambos fueron descubiertos hace más de 50 años, estudiando el pelo y la seda. El primer patrón descubierto, llamado *hélice α* , fue hallado en la proteína α -queratina, que es muy abundante en la piel y sus derivados. Un año después se halló una segunda estructura plegada, llamada *lámina β* , en la proteína fibroína, el principal constituyente de la seda. Estos dos patrones son particularmente comunes porque resultan de la unión de los grupos N-H y C=O del esqueleto polipeptídico mediante enlaces de hidrógeno, sin que intervengan las cadenas laterales de los aminoácidos. Por lo tanto, muchas secuencias diferentes de aminoácidos pueden formar estas estructuras. En cada caso, la cadena proteica adopta una conformación regular repetitiva.

El núcleo de muchas proteínas contiene extensas regiones de lámina β . Estas láminas β pueden estar formadas por cadenas polipeptídicas vecinas dispuestas paralelamente o por una cadena que se pliega una y otra vez sobre sí misma, de modo que cada sección de la cadena se dispone en dirección contraria a sus vecinas.

La hélice α se genera cuando una cadena polipeptídica se enrolla sobre sí misma formando un cilindro rígido. Entre cada cuatro enlaces peptídicos se establece un enlace de hidrógeno, uniendo el C=O de un enlace peptídico con el N-H de otro. Esto da lugar a una hélice regular con una vuelta completa cada 3.6 aminoácidos.

En las proteínas localizadas en las membranas celulares, como las proteínas transportadoras y los receptores, son especialmente abundantes regiones cortas de hélice α . Es muy raro que las subunidades se unan formando una línea recta, por lo que normalmente esta disposición da lugar a una hélice –una estructura regular que recuerda a una escalera de caracol. Dependiendo del giro de la escalera, una hélice puede ser dextrógira o levógira. Las hélices son comunes en las estructuras biológicas, tanto si las subunidades son moléculas pequeñas unidas entre sí covalentemente (los aminoácidos en una hélice α) como si son grandes moléculas proteicas. Esto no debe resultar sorprendente. Una hélice no es una estructura excepcional; se genera disponiendo muchas subunidades similares una a continuación de otra, de forma que la relación de cada una respecto de la anterior se repite estrictamente.

A.2 Glicosilación proteica

La glicosilación es la modificación covalente más frecuente en las proteínas. Ocurre por la unión de una o más cadenas oligosacáridas a la secuencia aminoacídica, confiriéndole a la proteína diferentes cualidades estructurales y funcionales. La glicosilación contribuye, por ejemplo, al correcto patrón de plegamiento de las proteínas, les confiere mayor resistencia a las proteasas, mayor estabilidad ante modificaciones físico-químicas del medio, entre otras propiedades. Por otra parte, los carbohidratos cumplen funciones muy importantes en el reconocimiento biológico, uniéndose a diferentes moléculas y mediando interacciones célula-célula.

La glicosilación o glicosilación proteica se trata del proceso de adición de carbohidratos a una proteína. La finalidad de las proteínas a las que se les ha añadido una cadena de glúcidos (**Glicoproteínas**), es ser secretada o formar parte de la superficie celular, aunque proteínas nucleares y citosólicas también pueden estar glicosiladas. Los carbohidratos de las glicoproteínas tienen funciones importante en el plegamiento de proteínas en el retículo endoplasmático rugoso (RER), en el destino de la proteína en los compartimientos intracelulares y en las interacciones célula-célula.

Hay dos tipos: N-Glicoproteínas y O-Glicoproteínas, dependiendo del lugar de adición de los carbohidratos:

- 1 N-glicoproteína: Los carbohidratos se unen al extremo N de la cadena lateral, del aminoácido asparagina.
- 2 O-Glicoproteína: El lugar de unión es el átomo de Oxígeno de los aminoácidos Serina y Treonina.

Los carbohidratos que se unen directamente a estos sitios son normalmente N-acetilglucosamina (GlcNAc) o N-acetilgalactosamina respectivamente (GalNAc).

La glicosilación se lleva a cabo en el interior del retículo endoplasmático rugoso. El primer paso es la translocación proteica (introducción de la proteína al interior del RER). Mientras se introduce, se hace la transferencia de un oligosacárido común, compuesto de 14 residuos (2 *N-acetilglucosaminas*, 3 glucosas y 9 manosas) a un aminoácido de asparagina de la proteína en síntesis. El oligosacárido es introducido al interior del RER, gracias a un lípido transportador de su membrana, llamado el dolicol fosfato.

Una vez transferido a la proteína, en el proceso de "maduración" de la proteína, este oligosacárido sufrirá unas modificaciones: perderá las 3 glucosas y 1 manosa. Posteriormente las glicoproteínas pasaran al Aparato de Golgi, donde sufrirán modificaciones adicionales.

Esto es para el *N-acetilglucosamina*, los *O-oligosacáridos* se forman con la adición de un único azúcar compuesto por pocos residuos monosacáridos. La diferencia con la O-glicosilación, es que esta última ocurre directamente en el aparato de Golgi, y directamente en la proteína postraduccionalmente, adicionándose con un grupo hidroxilo al residuo de serina o treonina. Por el contrario la N- glicosilación empieza en el RER y termina en el Golgi, donde se empaqueta y exporta desde el retículo trans-golgi como una proteína madura.

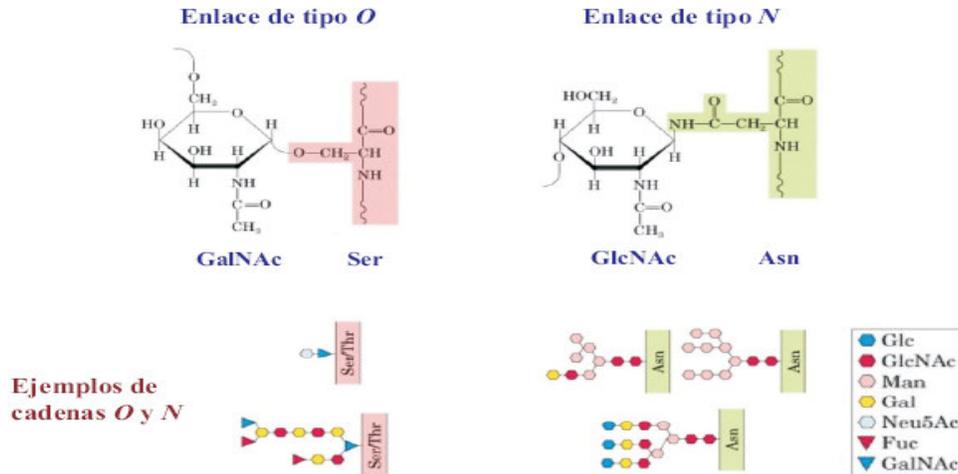
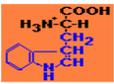
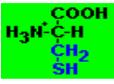
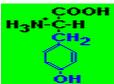
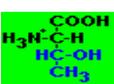
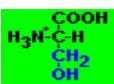
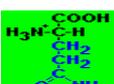
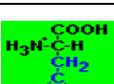
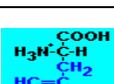
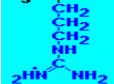
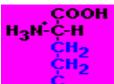


Figura A2. Esquema que intenta ilustrar la N- y la O-glicosilación. La mayoría de las proteínas presentan glicosilación del tipo N.

Tabla A1. Los 20 aminoácidos que forman las proteínas. En la lista se presentan las abreviaturas de una y de tres letras. La cuarta columna indica la polaridad de su cadena lateral, que es la que le confiere sus propiedades a cada aminoácido. El color del fondo de la estructura atómica indica el grupo al que pertenecen: naranja=apolares, verde=polares sin carga, fucsia=ácidos y cyan=básicos. Como se muestra, el número de cadenas laterales apolares es igual al de polares y iónicos.

Nombre Abreviatura	Estructura Atómica	Carga eléctrica [e] (a PH biológico)	Cadena Lateral	Índice Hidropático
Alanina Ala A	$\begin{array}{c} \text{COOH} \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_3 \end{array}$	0	Apolar	1.8
Glicina Gly G	$\begin{array}{c} \text{COOH} \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{H} \end{array}$	0	Apolar	-0.4
Valina Val V	$\begin{array}{c} \text{COOH} \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH} \\ \\ \text{H}_3\text{C}-\text{CH}_3 \end{array}$	0	Apolar	4.2
Leucina Leu L	$\begin{array}{c} \text{COOH} \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH} \\ \\ \text{H}_3\text{C}-\text{CH}_3 \end{array}$	0	Apolar	3.8
Isoleucina Ile I	$\begin{array}{c} \text{COOH} \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{H}_3\text{C}-\text{CH} \\ \\ \text{CH}_2 \\ \\ \text{CH}_3 \end{array}$	0	Apolar	4.5
Prolina Pro P	$\begin{array}{c} \text{COOH} \\ \\ \text{HN}-\text{C}-\text{H} \\ \\ \text{}^2\text{HC}-\text{CH}_2 \\ \\ \text{CH}_2 \end{array}$	0	Apolar	-1.6
Fenilalanina Phe F	$\begin{array}{c} \text{COOH} \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}_6\text{H}_5 \end{array}$	0	Apolar	2.8
Metionina Met M	$\begin{array}{c} \text{COOH} \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{S}-\text{CH}_2 \\ \\ \text{CH}_3 \end{array}$	0	Apolar	1.9

Triptófano Trp W		0	Apolar	-0.9
Cisteína Cys C		0	Apolar	2.5
Tirosina Tyr Y		0	Polar	-1.3
Treonina Thr T		0	Polar	-0.7
Serina Ser S		0	Polar	-0.8
Glutamina Gln Q		0	Polar	-3.5
Asparagina Asn N		0	Polar	-3.5
Histidina His H		0.6	Polar	-3.2
Lisina Lys K		1	Iónica	-3.9
Arginina ARg R		1	Iónica	-4.5
Ácido Glutámico Glu E		-1	Iónica	-3.5
Ácido Aspártico Asp D		-1	Iónica	-3.5

Apéndice B

Pruebas estadísticas³

B.1 La prueba χ^2

La prueba χ^2 es probablemente la prueba más usada de significancia estadística. Es una prueba no paramétrica, de modo que no involucra suposiciones acerca de una distribución normal de las poblaciones en estudio. La estadística χ^2 está determinada por $\chi^2 = (O - E)^2 / E$, en donde O es el valor observado y E el valor esperado de una cantidad bajo estudio. Dos aplicaciones importantes son la estimación de bondad de un ajuste y las tablas de contingencia. La forma básica de la prueba determina si las frecuencias observadas del valor de un parámetro particular o rango de valores difieren significativamente de las frecuencias esperadas de acuerdo a cierta teoría o hipótesis. Cuando la prueba se aplica a una serie N de valores observados O_i y esperados E_i , la prueba tiene la forma $\chi_i^2 = (O_i - E_i)^2 / E_i$, con $i=1, 2, \dots, N$. La estadística χ^2 estará dada por la sumatoria $\sum \chi_i^2$, y aparece tabulada en tablas que dependen de los grados de libertad estadísticos del sistema. El nivel de probabilidad resultante puede entonces compararse con el nivel de significancia escogido. Finalmente, puede realizarse la decisión de aceptar o rechazar la hipótesis nula H_0 referente a un problema específico.

En este trabajo, la prueba χ^2 se utiliza en el cálculo de tablas de contingencia. Las tablas de contingencia más simples de 2×2 , son pruebas de asociación entre categorías mutuamente exclusivas de una variable (dadas en los renglones de la tabla) y categorías mutuamente exclusivas de otra variable (dadas en las columnas de la tabla). Una tabla se expresa en forma de frecuencias que indican cómo se distribuye la frecuencia total entre las cuatro celdas de la tabla. La hipótesis nula a ser probada es de la forma:

H_0 : No existe relación entre las dos categorías de las variables.

Los grados de libertad de la prueba se determinan por $GL = (NC - 1) \times (NR - 1)$, donde NC es el número de columnas y NR el número de renglones, de manera que en una tabla de 2×2 , $GL = 1$. Por otro lado, si el número de frecuencias observadas en las celdas de la tabla son a, b, c, d, con un gran total $N = a + b + c + d$, entonces la estadística χ^2 puede calcularse de acuerdo con la fórmula $\chi^2 = N (ad-bc)^2 / [(a+b)(c+d)(a+c)(b+d)]$. El esquema anterior puede extenderse al caso más general de tablas de contingencia de $k \times n$. Es importante aclarar que la estadística χ^2 puede aplicarse a sistemas que involucran poblaciones con un número reducido de elementos.

B.2 La prueba U de Mann-Whitney

La prueba U de Mann-Whitney es una prueba no paramétrica utilizada para verificar si dos muestras de poblaciones provienen de la misma distribución. La hipótesis nula H_0 es que dos muestras han sido obtenidas de una sola población, y por tanto, sus distribuciones de probabilidad son idénticas. La prueba requiere que las dos muestras sean independientes y que las observaciones tengan forma numérica, de manera que pueda decidirse, dadas dos de ellas, cuál es mayor. La prueba involucra el cálculo de la estadística “U”, cuya distribución dada la hipótesis nula es conocida. La probabilidad asociada aparece tabulada para muestras pequeñas, la cual puede compararse con un nivel de significancia preestablecido. Cuando las muestras exceden a los 20 elementos, entonces puede utilizarse una distribución normal con bastante buena aproximación.

El cálculo de la estadística “U” para muestras pequeñas es muy sencillo. Para ello, se escoge la muestra para la cual los rangos de valores parecen ser más pequeños. Si se llama a esta muestra 1 y a la restante 2, entonces dada una observación en la muestra 1, se debe contar el número de observaciones en la muestra 2 que sean más pequeñas que ésta (contándose la mitad cuando son iguales). El número total de estas cuentas es U.

1 Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, James D. Watson Molecular Biology of The Cell 1994, 4ª edición.

2 Lehninger, David L. Nelson, Michael M. Cox, Principles of Biochemistry, Third edition, Worth publishers, 2004.

3 R.F. Mould, Introductory Medical Statistics, 3rd. Edition, Institute of Physics Publishing, Bristol, 1995.

BIBLIOGRAFIA

- 1 JUNAIDS (Joint United Nations Program on HIV-AIDS) Epidemic of AIDS 2005.
- 1 Richard A. Goldsby, Thomas J. Kindt, Barbara A. Osborne, Janis Kuby; Immunology, fifth edition 2003, McGrawHill.
- 1 Abul K. Abbas, Andrea H. Lichtman, Inmunología Celular y Molecular, quinta edición 2000, Elsevier España.
- 1 John M. Coffin, HIV Population Dynamics in Vivo: Implications for Genetic Variation, pathogenesis and Therapy, Science 267 1995.
- 1 E. Crivellato, A. Vacca and D. Ribatti, Trends in Immunology 25, Setting the stage: an anatomist's view or the immune system, 210-217 (2004).
- 1 Joanne York and Jack H., Role of Hydrophobic Residues in the Central Ectodomain of gp41 in Maintaining the Association between Human Immunodeficiency Virus Type 1 Envelope Glycoprotein Subunits gp120 and gp41. J Virology 2004, vol. 78 no. 9: 4921-4926.
- 1 W. R. Gallaher, J. M. Ball, R. F. Garry, A. M. Martin-Amedee and R. C. Montelaro, A General Model for the Surface Glycoproteins of HIV and Other Retroviruses, AIDS Research and human retroviruses vol. 11, num. 2, 1995.
- 1 Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, Nagashima KA, Cayanan C, Maddon PJ, Koup RA, Moore JP, Paxton WA. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. Nature 1996, 381: 667-673.
- 1 Deng HK, Unutmaz D, KewalRamani VN & Littman DR. "Expression cloning of new receptors used by simian and human immunodeficiency viruses". Nature 388, 296-300 (1997).
- 1 Lapham CK, Ouyang J, Chandrasekhar B, Nguyen NY, Dimitrov DS, Golding H. Evidence for cell-surface association between fusin and the CD4-gp120 complex in human cell lines. Science. 1996 Oct 25; 274 (5287):602-605.
- 1 Wu L, LaRosa G, Kassam N, Gordon CJ, Heath H, Ruffing N, Chen H, Humblías J, Samson M, Parmentier M, Moore JP, Mackay CR. 1997. Interaction of chemokine receptor CCR5 with its

ligands: multiple domains for HIV-1 gp120 binding and a single domain for chemokine binding. *J Exp Med.* 186 (8): 1373-1381.

1 EA. Berger, HIV entry and tropism: the chemokine receptor connection, *AIDS* 11 (Suppl. A): S3-S16.

1 T. Dragic, An Overview of the determinants of CCR5 and CXCR4 co-receptor function. *J. General Virol.* (2001), 82, 1807-1814.

1 Simmons G, Wilkinson D, Reeves JD, Dittmar MT, Beddows S, Weber J, Carnegie G, Desselberger U, Gray PW, Weiss RA, Clapham PR. Primary, syncytium-inducing human immunodeficiency virus type 1 isolates are dual-tropic and most can use either Lestr or CCR5 as coreceptors for virus entry. *J Virol.* 1996 70 (12): 8355-8360.

1 A. Berger, M. Murphy and M. Farber, Chemokine receptors as HIV-1 coreceptors: Roles in viral Entry, Tropism and Disease, *Annu. Rev. Immunol.* 1999 17, 657-700.

1 Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di Marzio P, Marmon S, Sutton RE, Hill CM, Davis CB, Peiper SC, Schall TJ, Littman DR, Landau NR. Identification of a major co-receptor for primary isolates of HIV-1. *Nature* 1996, 381: 661-666.

1 Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR. 1997. Change in coreceptor use correlates with disease progression in HIV-1 infected individuals. *J Exp Med.* 1997; 185 (4):621-628.

1 Scarlatti G, Tresoldi E, Bjorndal A, Fredriksson R, Colognesi C, Deng HK, Malnati MS, Plebani A, Siccardi AG, Littman DR, Fenyo EM, Lusso P. In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression. *Nat Med.* 1997; 3(11):1259-1265.

1 J. P. Moore, S. G. Kitchen, P. Pugach and J. A. Zack, The CCR5 and CXCR4 Coreceptors- Central to Understanding the Transmission and Pathogenesis of Human Immunodeficiency Virus Type 1 Infection, *AIDS Res. Hum. Retroviruses* 2004 20, 111-126.

1 Cleo G. Anastassopoulou and Leondios G. Kostrikis, Viral Correlatos of HIV-1 Disease, *Current HIV Research* 2005, 3, 113-132.

1 A. Jacobs, J. Sen, L. Rong and M. Caffrey, Alanine Scanning Mutants of the HIV gp41 Loop, *The Journal of Biological Chemistry* 2005, Vol. 280, No. 29, 27284-27288.

- 1 Burke, D. S. and F. E. McCutchan. 1997. Global distribution of human immunodeficiency virus type 1 clades. *AIDS: Biology, Diagnostics, Treatment and Prevention*. Ed. Lippincott-Raven Publishers, New York. 119 – 126.
- 1 Crandall K.A., D.A. Vasco, D. Posada and H. Imamichi. Advances in understanding the evolution of HIV. *AIDS* 1999. 13 (suppl): S39 – S47.
- 1 Roberts, J. D., K. Bedenek, and T. A. Kunkel. 1988. The accuracy of reverse transcription from HIV-1. *Science*. 242:1171 - 1173.
- 1 T. Leitner, B. Korber, M. Daniels, C. Calef, F. Foley, Subtypes & CRF's, Los Alamos National Laboratory, Los Alamos <http://hiv.lanl.gov/>
- 1 Osmanov S, Pattou C, Walter N, Schwardlander B and Esparza J. *J Acquir Immune Defic Syndr* 2002. 29: 184-190.
- 1 Blackard JT, Renjifo B, Fawzi W, Hertzmark E, Msamanga G, Mwakagile D, Hunter D, Spiegelman D, Sharghi N, Kagoma C and Essex M. *Virology* 2001. 287: 261-265.
- 1 Yang R, Kusagawa S, Zhang C, Xia X, Ben K and Takebe Y. *J Virol* 2003.77: 685-695.
- 1 Takebe Y, Motomura K, Tatsumi M, Lwin HH, Zaw M and Kusagawa S. *Aids* 2003. 17: 2077-2087.
- 1 Delgado E, Thomson MM, Villahermosa ML, Sierra M, Ocampo A, Miralles C, Rodríguez-Pérez R, Diz-Aren J, Ojea-de Castro R, Losada E, Cuevas MT, Vazquez-d Parga E, Carmona R, Perez-Álvarez L, Medrano L, Cuevas L, Tabeada JA and Najera R. *J Acquir Immune Defic Syndr* 2002. 29:536-543.
- 1 Ping Zhu, Jun Liu, Julian Bess Jr, Elena Chertova, Jeffrey D. Lifson, Henry Grisé, Gilad A. Ofek, Kenneth A. Taylor & Kenneth H. Roux. Distribution and three-dimensional structure of AIDS virus envelope spikes, *Nature* 2006 vol 441.
- 1 J. P. Moore, B. A. Jameson, R. A. Weiss and Q. J. Sattentau, "The HIV-cell fusion reaction" in *Viral Fusion Mechanism*, edited by J. Bentz, CRC Press, Boca Raton, 1993, pp. 233-289.
- 1 Alam SM, Paleos CA, Liao HX, Searce R, Robinson J, Haynes BF. An inducible HIV type 1 gp41 HR-2 peptide-binding site on HIV type 1 envelope gp120. *AIDS Res Hum Retroviruses* 2004;

20(8):836-845.

1 D. S. Dimitrov, X. Xiao, D.J. Chabot, C.C. Broder, HIV coreceptors, *J. Membrane Biol.* 1998, 16675-16690.

1 [Rizzuto CD, Wyatt R, Hernandez-Ramos N, Sun Y, Kwong PD, Hendrickson WA, Sodroski J](#). A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding, *Science* 1998 Jun 19; 280(5371):1949-53.

1 Kliger Y, Gallo SA, Peisajovich SG, Munoz-Barroso I, Avkin S, Blumenthal R, Shai Y. Mode of action of an antiviral peptide from HIV-1. Inhibition at a post-lipid mixing stage. *J Biol Chem.* 2001 Jan 12; 276(2): 1391-7.

1 Koshiba T, Chan DC. The prefusogenic intermediate of HIV-1 gp41 contains exposed C-peptide regions. *J Biol Chem* 2003 Feb 28;278(9):7573-7579.

1 Chang D-K, Cheng S-F, and Trivedi VD. Biophysical characterization of the structure of the amino-terminal region of gp41 of HIV-1. Implications on viral fusion mechanism. *J. Biol. Chem.* 1999 274(9): 5299–5309.

1 Picard L, Simmons G, Power CA, Meyer A, Weiss RA, Clapham PR. Multiple extracellular domains of CCR-5 contribute to human immunodeficiency virus type 1 entry and fusion. *J Virol.* 1997. 71(7):5003-11.

1 Lee B, Sharron M, Blanpain C, Doranz BJ, Vakili J, Setoh P, Berg E, Liu G, Guy HR, Durell SR, Parmentier M, Chang CN, Price K, Tsang M, Doms RW. Epitope mapping of CCR5 reveals multiple conformational states and distinct but overlapping structures involved in chemokine and coreceptor function. *J Biol Chem.* 1999 274(14):9617-26.

1 J.J. de Jong, A. de Ronde, W. Keulen, M. Tersmette and J. Goudsmit, *J. Virol.* 1992 66, 6777-6780.

1 G. Hoffmann, F. Seiller-Moiseiwitsch, J. Ahn, M. Walker and R. Swanstrom, Variability in the Human Immunodeficiency Virus Type 1 gp120 Env Protein Linked to Phenotype-Associated Changes in the V3 Loop, *J. Virol.* 2002 76, 3852-3864.

1 Olson WC, Rabut GE, Nagashima KA, Tran DN, Anselma DJ, Monard SP, Segal JP, Thompson DA, Kajumo F, Guo Y, Moore JP, Maddon PJ, Dragic T. Differential inhibition of human

immunodeficiency virus type 1 fusion, gp120 binding, and CC-chemokine activity by monoclonal antibodies to CCR5. *J Virol.* 1999 73(5):4145-55.

1 Doranz, B., Orsini, M., Turner, J., Hoffman, T., Berson, J., Hoxie, J., Peiper, S., Brass, L. & Doms, R., Identification of CXCR4 domains that support co-receptor and chemokine receptor functions, *J. Vir.* 1999 73, 2757-2761.

1 R. W. Sanders, B. Korber, Min Lu, B. Berkhout and J. P. Moore, Mutational Analyses and Natural Variability of the gp41 Ectodomain, LANL Database, <http://hiv-web.lanl.gov/content/hiv-db/REVIEWS/reviews.html>

1 R. Suntok and C. Chan, The Fusion Activity of HIV-1 gp41 Depends on Interhelical Interactions, *J. Biol. Chem.* 2005 280, 19852-19857.

1 David C. Chan, Deborah Fass, James M. Berger, and Peter S. Kim. Core structure of Gp41 from the HIV envelope glycoprotein. *Cell* 1997 89, 263-273.

1 O. M. Pello, J. M. Rodríguez-Frade, L. Martínez-Muñoz, M. Mellado. Modulación del tráfico leucocitario: Papel de las quimiocinas y de los opioides. *Inmunología* 2006 25 (1):39-49.

1 Endres, M.J., Clapham, P.R., Marsh, J., Ahuja, M., Turner, J.D., McKnight, A., Thomas, J.F., Stoebenau-Haggarty, B., Choe, S., Vance, P.J., Wells, T. N., Power, C.A. et al., CD4-independent infection by HIV-2 is mediated by fusin/CXCR4. 1996 87:745-756.

1 A. Koning, P. van Rij and H. Schuitemaker, Biological and Molecular Aspects of HIV-1 Coreceptor Usage, HIV Sequence Database Los Alamos,

1 R. Sahchez, A. Sali, *J. Comput. Phys.* 1999, 151:388-395.

1 Shuqun Liu, Shixiu Fan, Zhirong Sun, Structural and functional characterization of the human CCR5 receptor in complex with HIV gp120 envelope glycoprotein and CD4 receptor by molecular modeling studies, *J. Mol. Model* 2003, 9:329-336.

1 Shuqun Liu, Xiufan Shi, Ciquan Liu, Zhirong Sun, Characterize dynamic conformational space of human CCR5 extracellular domain by molecular modeling and molecular dynamics simulation, *J. Mol. Structure* 2004, 673:133-143.

1 HIV Sequence Search, Los Alamos National Laboratory. <http://www.hiv.lanl.gov/components/hiv->

db/combined_search_s/search.html

1 Schneider, T.D. and R.M. Stephens. Sequence Logos: A new Way to Display Consensus Sequences. *Nucleic Acids Res.* 1990 18: 6097-6100.

1 Thompson, J.D., Higgins, D.G. and Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994 22: 4673-4680.

1 Shannon, C.E. A Mathematical Theory of Communication. *The Bell System Technical Journal* 1948 27: 379-423, 623-656.

1 Pierce, J.R. *An Introduction to Information Theory: Symbols, Signals and Noise*, Dover Publications, Inc., 1980. New York. second edition.

1 Johnson, W. E., J. M. Sauvron, and R. C. Desrosiers. Conserved, N-linked carbohydrates of human immunodeficiency virus type 1 gp41 are largely dispensable for viral replication. *J. Virol.* 2001 75:11426–11436.

1 Simmons G, Wilkinson D, Reeves JD, Dittmar MT, Beddows S, Weber J, Carnegie G, Desselberger U, Gray PW, Weiss RA, Clapham PR. Primary, syncytium-inducing human immunodeficiency virus type 1 isolates are dual-tropic and most can use either Lestr or CCR5 as coreceptors for virus entry. *J Virol.* 1996 70(12):8355-60.

1 Yuan W, Craig S, Si Z, Farzan M, Sodroski J. CD4-induced T-20 binding to human immunodeficiency virus type 1 gp120 blocks interaction with the CXCR4 coreceptor. *Journal of Virology* 2004, Vol. 78, No. 10 pp. 5448-5457

1 S. Wang, J. York, W. Shu, O. Stoller, H. Nunberg and M. Lu, Interhelical Interactions in the gp41 Core: Implications for Activation of HIV-1 Membrane Fusion, *Biochemistry* 2002 41, 7283-7292.

1 Dragic T, Trkola A, Lin SW, Nagashima KA, Kajumo F, Zhao L, Olson WC, Wu L, Mackay CR, Allaway GP, Sakmar TP, Moore JP, Maddon PJ. Amino-terminal substitutions in the CCR5 coreceptor impair gp120 binding and human immunodeficiency virus type 1 entry. *J Virol.* 1998 72(1):279-85.

1 Wu L, LaRosa G, Kassam N, Gordon CJ, Heath H, Ruffing N, Chen H, Humblias J, Samson M, Parmentier M, Moore JP, Mackay CR. Interaction of chemokine receptor CCR5 with its ligands:

multiple domains for HIV-1 gp120 binding and a single domain for chemokine binding. *J Exp Med.* 1997 186(8):1373-81.

1 Wang, W.-K., T. Dudek, Y. J. Zhao, H. Brumblay, M. Essex, and T.-H. Lee. CCR5 co-receptor utilization involves a highly conserved arginine residue of human immunodeficiency virus type 1 gp120. *Proc. Natl. Acad. Sci. USA* 1998 95:5740-5745.

1 Zhang, H., G. Orti, Q. Du, J. He, C. Kankasa, G. Bhat, and C. Wood, Phylogenetic and phenotypic analysis of HIV type 1 env gp120 in cases of subtype C mother-to-child transmission, *AIDS Res. Hum. Retroviruses* 2002 18: 1415-1423.

1 Cilliers, T., J. Nhlapo, M. Coetzer, D. Orlovic, T. Ketas, W. C. Olson, J. P. Moore, A. Trkola, and L. Morris. The CCR5 and CXCR4 Coreceptors Are Both Used by Human Immunodeficiency Virus Type 1 Primary Isolates from Subtype C. *J. Virol.* 2003 77: 4449-4456.

1 Pollakis, G., S. Kang, A. Kliphuis, M. I. M. Chalaby, J. Goudsmit, and W. A. Paxton. N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization. *J. Biol. Chem.* 2001 276: 13433-13441.

1 Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, James D. Watson
Molecular Biology of The Cell 1994, 4^a edición.

1 Lehninger, David L. Nelson, Michael M. Cox, *Principles of Biochemistry*, Third edition, Worth publishers, 2004.

1 R.F. Mould, *Introductory Medical Statistics*, 3rd. Edition, Institute of Physics Publishing, Bristol, 1995.