BASES BIOFÍSICAS Y ESTRUCTURALES DEL ENSAMBLADO DE LA CÁPSIDE RETROVIRAL: VIRUS DE LA LEUCEMIA BOVINA



Gonzalo Obal

Unidad de Biofísica de Proteínas

Plataforma de Biología Estructural

Institut Pasteur de Montevideo

2014

Tesis de Doctorado en Ciencias Biológicas

Opción Biofísica

Programa de Desarrollo de las Ciencias Básicas (PEDECIBA)

Tutor:

Dr. Otto Pritsch

Unidad de Biofísica de Proteínas/Plataforma de Biología Estructural (Institut Pasteur Montevideo)

Tribunal:

Dr. Juan Arbiza

Sección Virología/Facultad de Ciencias (Universidad de la República)

Dra. Ana Denicola

Laboratorio de Fisicoquímica Biológica/Facultad de Ciencias (Universidad de la República)

Dra. Cecilia Fernández

Cátedra de Inmunología/Facultad de Química (Universidad de la República)

25 de Setiembre de 2014

"Nuestras virtudes y defectos son inseparables, como la fuerza y la materia. Cuando se separan se termina el hombre"

Nikola Tesla

RESUMEN

La organización supramolecular de la cápside retroviral, formada mediante el auto-ensamblado de una única proteína, representa una estructura fascinante, cuyas características debieron moldearse durante la evolución, sujetas a las restricciones biofísicas del ciclo replicativo de los retrovirus. En particular, su función de encerrar, contener, y proteger el genoma y enzimas replicativas de la partícula retroviral requiere que la estructura sea suficientemente estable para resistir fuera de la célula hospedero. A su vez, la estabilidad debe ser compatible con su posterior desensamblado, una vez que el virus entra en la célula, lo cual es necesario para continuar la infección. Por tanto, otra restricción evolutiva importante debe haber sido que las interacciones de la estructura fuesen reversibles. En este sentido, la elucidación de las bases moleculares y estructurales del auto-ensamblado de la proteína CA se ha convertido en un foco principal de atención en la investigación de la biología de retrovirus.

Actualmente, nuestro conocimiento de los factores fisicoquímicos que controlan el auto-ensamblado de CA es relativamente escaso, y la información disponible no permite predecir cómo se gatilla/modula el ensamblado. En particular, *in vitro* ninguna CA polimeriza en forma apreciable en condiciones fisiológicamente relevantes, por lo cual es aún una incógnita cómo se gatilla la polimerización en el virión. Por otro lado, si bien se ha avanzado enormemente en la caracterización de la estructura de la cápside, particularmente de los bloques de construcción hexamérico y pentamérico, aún quedan preguntas fundamentales sin responder. En particular, no ha sido posible elucidar la estructura 3D de la *lattice* hexamérica de la cápside, a una resolución adecuada (atómica) para definir las interacciones que median la asociación entre capsómeros.

En esta tesis, abordamos la caracterización biofísica y estructural del ensamblado de la cápside del Virus de la Leucemia Bovina (BLV), un oncoretrovirus del género *Deltatrovirus* (grupo BLV/HTLV), agente causal de la Leucemia Bovina Enzoótica. Específicamente, nos enfocamos en el estudio de las propiedades de auto-ensamblado *in vitro*, así como la elucidación de la estructura 3D cristalina de la proteína completa y sus dominios independientes NTD y CTD. Cabe notar que no existía información previa disponible en literatura sobre ninguno de estos dos aspectos, y de hecho tampoco existían datos específicos sobre otras CA *Deltaretrovirales*.

En el eje biofísico de este trabajo, analizamos el proceso de auto-ensamblado de la proteína. Pusimos a punto la producción de la proteína recombinante nativa (intacta) expresada en *Escherichia coli*, y luego optimizamos un ensayo turbidimétrico para evaluar la cinética de polimerización. Usando este ensayo, evaluamos el efecto de un amplio rango de condiciones y factores, con el objetivo de establecer cuáles son los requerimientos para gatillar y modular el ensamblado. Encontramos que CA_{BLV} permanece como

monómero, y auto-ensambla en forma fuertemente dependiente de la temperatura, no siendo apreciable a menos de 20 °C y promoviéndose la reacción hasta alcanzar un óptimo cerca de los 38°C (de hecho, la temperatura normal del hospedero bovino). Interesantemente, la reacción se inhibe casi completamente a fuerza iónica fisiológica. Sin embargo, la adición de ARN o agentes de *"crowding"* a efectos de mimetizar la presencia del ARN viral y elevada concentración de macromoléculas del interior del virión respectivamente, disparan dramáticamente el ensamblado. Los análisis mediante microscopía electrónica revelaron que CA_{BLV} auto-ensambla *in vitro* formando mallas bidimensionales. En presencia de iones divalentes (como fostato), CA_{BLV} auto-ensambla muy rápidamente formando estructuras esféricas, lo cual podría implicar la inducción de la formación de pentámeros. Sorprendentemente, en un tampón fisiológico y a 38oC, la adición de ARN o agentes de crowding no sólo "levantó" el efecto inhibitorio de la fuerza iónica, sino que alteró el hábito de ensamblado, promoviendo la formación (casi exclusiva) de partículas esféricas, similares a la cápside nativa del BLV en forma y tamaño. De esta forma, establecimos que en un contexto fisiológico, las condiciones intrínsecas del interior del virión son el principal determinante del ensamblado y formación de una estructura adecuada, con las características típicas de la cápside.

En el eje estructural, elucidamos la estructura cristalina de la proteína CA nativa, a una resolución de 2,75Å. La proteína intacta cristalizó formando un bloque de construcción hexamérico, organizado en un arreglo bidimensional (plano) que mimetiza el ensamblado hexagonal característico de la *lattice* hexagonal auténtica. Más específicamente, el cristal consta de "capas" 2D de hexámeros, apiladas unas sobre otras. En cada capa, los hexámeros se asocian lateralmente con sus vecinos, en forma análoga a la cápside madura nativa, por lo cual la estructura cristalina refleja fielmente el arreglo de los capsómeros en la cápside nativa.

Interesantemente, la estructura reveló que el capsómero hexagonal es asimétrico. Esta asimetría surge del hecho que el hexámero es completamente flexible, tanto a nivel del anillo central de NTDs, como del cinturón del CTDs circundante. En el cristal, el hexámero aparece distorsionado, con ambos anillos "retorcidos" en relación al eje central (eje de simetría 6; de hecho, *quasi*-6 debido a la asimetría) que vincula los seis protómeros. Este hecho es notable, ya que sugiere que el capsómero es estructuralmente flexible, pudiendo adoptar una amplia variedad de conformaciones. El giro o "twist" entre ambos anillos parece ser permitido/favorecido por el *linker* relativamente largo (12 residuos), revelando un nivel de plasticidad no observada en ninguna otra CA retroviral. Por otra parte, el empaquetamiento específico del hexámero en el cristal reveló la presencia de una interfaz trimérica, formada por la conjunción de los CTDs de tres hexámeros vecinos. Así, cada hexámero contacta con su vecinos a través de 6 interfaces CTD-CTD triméricas. Interesantemente, observamos que existen 3 tipos de contactos triméricos idénticos; en otras palabras, los contactos entre CTDs son similares o *quasi*-equivalentes entre sí (resultando en 3 ejes *quasi*-3). Esto otorga al hexámero la capacidad de adoptar distintas declinaciones respecto a los vecinos adyacentes. Esta interfaz se consideraba exclusiva de una *lattice* hexamérica curva, como ocurre en

cápsides auténticas o ensambles *in vitro* cilíndricos. Sin embargo, demostramos que la interfaz existe y es necesaria para mantener una estructura plana. Por su parte, los contactos entre hexámeros via la interfaz dimérica CTD-CTD se conservan relativamente intactos, aunque pueden adoptar pequeños cambios según el ambiente local. La figura que emerge es un capsómero hexamérico completamente flexible, cuya estructura responde coordinadamente, en forma alostérica, y donde el anillo de NTDs canaliza los cambios transmitidos desde el cinturón de CTDs. Esto sugiere una manera de generar curvatura mediante adaptabilidad del capsómero a ambientes de interacción locales variables. De hecho, la cápside es una estructura asimétrica cuya curvatura varía continuamente.

Así, por primera vez para BLV (y retrovirus en general), pudimos establecer un vínculo funcional entre las condiciones de la partícula viral y la inducción de la polimerización y morfología supramolecular de la cápside. Por otro lado, logramos establecer las bases estructurales del ensamble hexagonal de la cápside madura, y demostramos a nivel atómico, (por primera en retrovirus) que el capsómero interacciona mediante los principios de *quasi*-equivalencia descritos inicialmente por Caspar y Klug, típicos de las cápsides icosaédricas. Esta información, abre la puerta a una mejor comprensión de la manera en que los capsómeros otorgan a la cápside su exquisita metastabilidad. Más allá de BLV, los resultados de esta tesis son relevantes para retrovirus en general, ya que aportan luz sobre aspectos que, al presente, son poco comprendidos y/o aún no fueron resueltos

Índice de Figuras Capítulo 1

Figura 1. Esquema de la organización del genoma retroviral	11
Figura 2. Estructura básica de la partícula retroviral	13
Figura 3. Ciclo replicativo de los retrovirus	16
Figura 4. Esquema de los destinos de la cápside retroviral en la célula infectada	19
Figura 5. La maquinaria ESCRT cellular es usada para el brote	
de las partículas retrovirales	23
Figura 6. Arquitectura modular de las proteínas Gag	25
Figura 7. Unión de Gag a la membrane plasmática	28
Figura 8. Estructura de la <i>lattice</i> inmadura de Gag	31
Figura 9. Diferencias en la organización estructural del virión maduro en inmaduro de HIV-1	34
Figura 10. Efecto del procesamiento proteolítico de Gag en la unión MA-CA	35
Figura 11. Morfologías representativas de cápsides retrovirales	38
Figura 12. Estructuras cristaolográficas del capsómero pentamérico y hexamérico	40
Figura 13. Quasi-equivalencia de los contactos en protómeros del capsómero retroviral	41
Figura 14. La <i>lattice</i> 2D (plana) de	
CA de HIV-1	43
Figura 15. Contactos entre hexámeros	
en la <i>lattice</i> curva	45

CAPÍTULO 1

Biología Celular y Estructural del Ensamblado de los Retrovirus

1.1 LOS RETROVIRUS

Los retrovirus son una familia diversa y fascinante de virus ARN, algunos de los cuales ocasionan desórdenes inmunológicos, neurológicos, y varios tipos de cáncer en humanos, animales, y plantas, y son uno de los patógenos más profundamente estudiados⁽¹⁸⁻²¹⁾. Su nombre se debe a la transcriptasa reversa, enzima fundamental para su replicación, que genera ADN a partir del ARN genómico viral⁽¹⁸⁾. Su descubrimiento revolucionó el dogma central de la biología, y cambió radicalmente la visión de la plasticidad génica. Las patologías ocasionadas por retrovirus se reconocieron hace más 100 años: el adenocarcinoma pulmonar ovino (ocasionado por el JSRV^a) fue reportado en 1825⁽²²⁾, y la leucemia bovina (ocasionada por el BLV^b) en 1870⁽²³⁾. En 1908 se reportó el origen viral de la leucemia en gallinas (ocasionado por el RSV^{c}), y se confirmó en 1911⁽²⁴⁾. El hallazgo de un retrovirus oncogénico fue un hito sin precedentes, seguido del aislamiento del Virus de la Leucemia Murina (MLV^e) cuatro décadas después^(25, 26). El primer retrovirus capaz de infectar humanos se aisló en 1980 por Robert Gallo y se denominó HTLV-1^f, agente causal de la leucemia de células T del adulto (ATLL^g) y la enfermedad neurológica paraparesis espástica tropical/mielopatía asociada al HTLV (HAM/TSP^h)^(27, 28). El pasado año 2013 marcó el 30^{mo} aniversario del primer reporte del aislamiento por Barré-Sinousi y Robert Gallo del virus HIV-1ⁱ, agente causal del síndrome de inmunodeficiencia adquirida⁽²⁹⁾, que ocasionó una epidemia devastadora (y desconocida) en la década de 1980^(30, 31). Además de su papel patogénico, a lo largo de la evolución algunos retrovirus se integraron en el ADN germinal de varios hospederos, y actualmente constituyen cerca del 8% del genoma humano⁽³²⁾. Más aún, proteínas fusogénicas codificadas por *retrovirus endógenos* juegan un papel esencial en el desarrollo de la placenta^(33, 34).

Deltaretrovirus y el Virus de la Leucemia Bovina

El género *Deltaretrovirus* incluye tres virus altamente relacionados: el BLV, los tipos I y II del HTLV (HTLV-1, y HTLV-2), y I y II del Virus de la Leucemia de células T de simios (STLV-1 y STLV- 2^{j})^(35, 36). El HTLV-1 es el más ampliamente estudiado, ya que infecta células T CD4+ de humanos, y afecta ~10-20 millones de personas a nivel mundial están infectadas con HTLV-1. El BLV es otro miembro importante de

^(a)Jaasiekte Sheep Retrovirus; ^(b)Bovine Leukemia Virus; ^(c)Rous Sarcoma Virus; ^(e)Murine Leukemia Virus; ^(f)Human Tcell Leukemia Virus Type I; ^(g)Adult T-cell Leukemia; ^(h)HTLV-associated myelopathy/tropical spastic paraparesis; ⁽ⁱ⁾Human Immunodeficiency Virus Type I. ^(j) Simian Immunodeficiency Virus;

los deltaretrovirus, muy similar al HTLV, por lo cual se denominan conjuntamente grupo HTLV/BLV. Es el agente etiológico de una enfermedad linfoproliferativa denominada Leucosis Bovina Enzoótica (EBL^a), que afecta las células B CD5+ del bovinos y puede ocasionar linfocitosis y/o tumores malignos^(4, 37). De hecho, es la enfermedad neoplásica más frecuente del ganado bovino lechero^(38, 39). Además del ganado en pie, sólo las ovejas desarrollan leucemia al infectarse con BLV, mientras que en humanos no se considera probable la infección⁽⁴⁰⁾. La enfermedad exhibe una latencia prolongada que puede cursar en forma asintomática, si bien puede diagnosticarse mediante serología⁽⁴¹⁾. La enfermedad es crónica y altamente contagiosa, con desarrollo de tumores a partir de los tres años, mientras que al cabo de este tiempo aproximadamente 30-70% de los animales infectados pueden desarrollar una linfocitosis persistente⁽⁴²⁾. Esta etapa es considerada benigna, y se caracteriza por acumulación de células B no transformadas. Una minoría de animales, cerca de un 0.1-5%, desarrolla linfomas malignos de origen clonal, usualmente luego de 5-8 años del curso de la infección⁽⁴³⁾. El BLV se transmite horizontalmente a través del contacto directo con células infectadas en la sangre. Su prevalencia es alta en varias regiones, incluyendo Estados Unidos y América Latina. Actualmente no existe cura para la infección con BLV, y no se ha avanzado casi nada en la generación de una vacuna. La enfermedad produce pérdidas económicas directas por su forma subclínica, incluyendo disminución en la producción lechera e infertilidad. Asimismo, es limitante para la exportación de ganado en pie o productos tales como el semen. Varios países han emprendido campañas de control y/o erradicación de la LBE, siendo el paradigma el éxito logrado por la Unión Europea mediante sacrificio de los animales seropositivos, siendo actualmente libres de LBE la mayoría de sus integrantes. Muchos países poseen programas voluntarios de control y políticas restrictivas a la importación de animales en pie. En Uruguay la LBE se conoce hace muchos años. Los estudios epidemiológicos llevados a cabo hace dos décadas revelan para el ganado lechero una prevalencia de más del 70%, con animales positivos en ~100% de los establecimientos ($^{(44)}$ y datos no publicados de nuestro laboratorio). Actualmente se reconoce su afección en el status sanitario del País, por lo cual en 2007 el Ministerio de Ganadería, Agricultura, y Pesca decretó un programa sanitario para la declaración de predios oficialmente libres de LBE.

Organización Genómica de los Retrovirus

La familia *Retroviridae* consta de dos subfamilias, *Spumaretrovirinae* y *Orthoretrovirinae*, cuya característica saliente es la transcripción reversa de su genoma ARN simple hebra a un intermediario ADN doble hebra, el cual se integra covalentemente en el ADN genómico de la célula hospedero (ver ANEXO 1). La organización genómica es similar entre retrovirus (ver Fig. 1). El genoma de los retrovirus consiste de dos copias idénticas de ARN simple hebra, relativamente pequeño (9.3 Kb en HIV-1, 8.7 Kb en BLV),

^(a) Enzootic Bovine Leukemia.

lineal y no segmentado, con un casquete 5'-metilguanosina y su extremo 3' poliadenilado^(18, 35). La hebra de ARN es de polaridad positiva (misma polaridad del ARN mensajero) por lo que puede usarse directamente como molde para la transcripción. El genoma es pseudo-diploide, ya que si bien cada partícula posee dos copias de su material genético, éstas dan lugar a la producción de una única copia de ADN en la célula infectada.⁽⁴⁵⁻⁴⁸⁾. Ambas copias son importantes para la replicación y cada una posee toda la información genética necesaria para llevar a cabo la replicación viral, lo cual constituye un mecanismo de seguridad.. El ARN está flanqueado por dos regiones idénticas, denominadas secuencia repetida

terminal (LTR^a). Ambas LTR 5' y 3', contienen las secuencias de inicio y terminación de la transcripción, intensificadores de la expresión génica, así como maduración posttranscripcional. La secuencia 5'LTR contiene una región no traducida (5`UTR^b) que actúa como promotor transcripcional para la ARN polimerasa II, una señal para el empaquetamiento del genoma denominada sitio psi (Ψ) , y un sitio de unión a cebador (PBS^c) para el comienzo de la transcripción, El 3'LTR actúa como terminador y contiene una secuencia de polipurinas que permite la replicación de la segunda cadena. El genoma presenta varios posibles marcos de lectura abiertos que codifican las proteínas estructurales, glicoproteínas de superficie, y enzimas replicativas, así como marcos de lecturas adicionales para proteínas accesorias. Existen tres regiones codificantes principales conservadas: gag, ⁽³²⁾-pol, y env (en la mayoría de retrovirus, pro se considera parte de *pol*). El gen *gag* codifica para la principal proteína estructural del virus,



Fig. 1. Esquema de la organización del genoma retroviral. Se muestra el genoma de BLV, el cual es representativo de la mayoría de retrovirus complejos (los simples no codifican proteínas accesorias). El ARN está flanqueado por la secuencia LTR y contiene los marcos de lectura abiertos para los genes gag, pro, pol, y env. El ARN es transcrito desde la 5'LTR hasta la 3'LTR, el sirve como molde para la producción de los poliproteínas Gag y Gag-Pro-Pol. El procesamiento proteolítico de estas poliproteínas (el cual ocurre en los viriones que brotaron de la célula infectada) da lugar a las proteínas Matriz (MA), Cápside (CA), Nucleocápside (NC), Proteasa (PRT), Transcriptasa reversa (RT), e integrasa (IN). Mediante empalme alternativo se generan los ARNm que codifican para el resto de proteínas. La escisión de una gran segmento en la región gag-pro-pol genera el ARNm de Env, la cual será luego escindida en las subunidades SU y TM. La escisión de un segundo segmento genera el ARNm para las proteínas accesorias Tax/Rex. Otros eventos de empalme alternativo dan lugar a las proteínas R3 y G4. El genoma de HIV, similar en organización, posee tres genes principales (gag, pol, y env) comunes a todos los retrovirus, y posee seis genes accesorios (tat, rev, vpu, vpr, vif, y nef). Tomado de ⁽⁴⁾.

denominada Gag (del inglés: Group specific antigens), una poliproteína que es procesada proteolíticamente en la etapa final de formación de la partícula viral (se discute ampliamente más adelante). El gen pro-pol codifica otra poliproteína, precursora de las dos enzimas replicativas esenciales -transcriptasa reversa (TR) e integrasa (IN)- así como la proteasa ⁽³²⁾ responsable del clivado de ambas poliproteínas. El gen gag se localiza en el extremo 5' del ARN, corriente arriba de pro-pol. La traducción de pro-pol es interesante ya que ocurre a través de un "salteo" del codón de terminación de gag, resultando en la expresión de ~5% de la poliproteína Gag-Pol⁽⁴⁹⁾. En BLV, el gen gag-pro (pr66^{gag-prt}) se traduce mediante supresión del codón de terminación, a través de un corrimiento de marco, por un ARNt de lisina^(50, 51). El gen *env* da lugar a ⁽⁴⁾la proteína de envoltura, la cual interacciona con el receptor específico de la célula hospedero y activa la fusión de membranas. Env se sintetiza como una proteína precursora, la cual es clivada por una proteasa celular (furina) en sus dos subunidades: superficie (SU) y trasmembrana (TM). Los géneros de retrovirus que sólo codifican para gag, pro, pol, y env, se denominan simples. Algunos géneros -particularmente el grupo HTLV/BLV v Lentivirus- se denominan *compleios*, va que además de estos genes en común, poseen genes accesorios que codifican proteínas reguladoras de la expresión génica⁽⁴⁾. HIV contiene varios genes accesorios (tat, rev, vif, nef, vpr, y vpu) esenciales para su ciclo replicativo. Los genomas de HTLV y BLV, si bien presentan diferencias, poseen una secuencia característica denominada pX localizada entre env v 3'LTR que codifica las proteínas regulatorias Tax y Rex. Tax es la proteína más ampliamente estudiada debido a su papel clave en la replicación y el potencial oncogénico de ambos retrovirus, mientras que Rex controla el transporte y traducción del ARN viral⁽⁵²⁻⁵⁴⁾. En BLV la región pX expresa las proteínas R3 y G4 que contribuyen al mantenimiento de la carga viral, mientras que HTLV-1 expresa p12^I y p13^{II} (funcionalmente homólogas a R3 y G4 respectivamente) y el regulador transcripcional p30^{II(55, 56)}.

Estructura y Composición de la Partícula Retroviral

La estructura de virión es común a la mayoría de retrovirus, incluyendo el BLV (ver Fig 2). Existen dos formas estructural y funcionalmente diferentes de la partícula retroviral: i) una forma *inmadura*, no infectiva, y es la que posee el virus inmediatamente luego de brotar de la célula hospedero; y ii) una vez libre, poco tiempo después, tiene lugar un dramático reordenamiento estructural que genera la partícula *madura*, con plena capacidad infectiva. Esta maduración se verá en detalle más adelante, ya que involucra el objeto de estudio de esta Tesis, la formación de la cápside. Ahora se describe, a efectos de aportar el marco general, la organización típica de la partícula retroviral madura.

Los viriones -esto es, las partículas virales poseen una morfología más o menos esférica, con un diámetro de muy variable entre distintos retrovirus, típicamente descrita entre 100-120 nanómetros (145 nm en HIV-1, es la determinación más precisa mediante crioelectrón-microscopía) ^(18, 57, 58). La partícula de BLV muestra un diámetro muy variable, entre 60-125 nm^(59, 60). Son virus envueltos, esto es, poseen una envoltura lipídica que deriva de la membrana lipídica de la célula hospedero. La membrana viral deriva de la célula hospedero infectada, y es adquirida cuando la nueva partícula viral brota al medio extracelular, llevándose consigo como envoltura parte de la bicapa lipídica. Sin embargo, su composición lipídica es diferente, ya que el ensamblado de la partícula viral ocurre en microdominios de la membrana plasmática rica en lípidos (así como proteínas) requeridos para reclutar los componentes virales en la bicapa lipídica.

En la membrana se encuentra inserta la glicoproteína Env, compuesta de las gliproteínas SU (glicoproteína mayor de superficie) v TM (glicoproteína transmembrana). La glicoproteína Env forma un trímero de tres subunidades SU, unidas a tres subunidades TM que permanecen ancladas a la membrana. En HIV-1, Env (gp160) consiste de gp120 (SU) y gp41 (TM). En la Env de BLV (gp72), SU y TM se denominan gp51 y gp30 respectivamente. El interior de la partícula posee las tres proteínas estructurales, derivadas de la proteólisis de la poliproteína Gag: Matriz (MA), Cápside (CA), y Nucleocápside (NC). Gag representa ~50% de la masa total del virus, y 3/4 del componente proteico de la partícula viral una vez ensamblada y lista para ser liberada de la célula hospedero. En los viriones de HIV-1 se estiman ~5000 moléculas de Gag, de las cuales ~2% están ubiquitiniladas. En BLV Gag se denomina Pr44^{gag}, la cual se escinde en sus tres dominios: la proteína p15



Fig. 2. Estructura básica de la partícula retroviral. Se muestra en forma esquemática la organización de la partícula del BLV. Dos copias idénticas de ARN simple hebra constituyen el genoma, empaquetadas formando un complejo ribonucleoproteico con la proteína NC (p12). Este "core" se encuentra dentro de la cápside -formada por la proteína CA (p24)-, estructura que encierra totalmente el genoma, protegiéndolo. Dentro de la cápside también se encuentran las enzimas replicativas RT e IN. La proteína MA se encuentra de hecho anclada a la bicapa lipídica (se ve con detalle más adelante), tapizando gran parte de la superficie de la cara citosólica de la misma. La membrana lipídica viral deriva de la célula hospedero, aunque posee una composición lipídica y proteica específica. Anclada a la misma, se encuentra la proteína de envoltura Env) responsable de la entrada del virus a la célula hospedero, y que costa de las subunidades SU y TM. La representación es una figura clásica en la literatura del BLV. Su sencillez, aunque totalmente ilustrativa del ordenamiento de los componentes básicos, refleja el escaso conocimiento estructural disponible BLV y HTLV. Tomado de ⁽⁴⁾.

(Matriz, MA), p24 (Cápside, CA), y p12 (Nucleocápside, NC). Matriz se encuentra unida al lado interno de la bicapa lipídica de la envoltura viral, "tapizando" gran parte de la superficie interna de la membrana, y en BLV es subsecuentemente escindida en tres fragmentos, p10 (conteniendo el miristoílo), un producto de 7 aa, y p4⁽⁶¹⁾. La proteína Cápside forma una estructura supramolecular denominada cápside, compuesta por varios cientos de copias auto-ensambladas. La cápside encierra completamente y protege las dos hebras de ARN genómico. Ambas hebras forman un complejo ribonucleoproteico compacto con ~1,500-2,000 moléculas de NC, una proteína altamente básica que cumple una función análoga a las histonas, permitiendo el empaquetamiento compacto del ARN. ^(62, 63). Dentro de la cápside, estimándose ~100 moléculas en total. El resto del componente proteico es aportado principalmente por Env, si bien existen sólo ~14 trímeros en la superficie de la membrana. La proteína accesoria vpr se encuentra en cantidades relativamente grandes, ~200 moléculas por partícula. Las dos copias de ARN representan 1/3 partes de la masa del virión.

Notablemente, en la partícula retroviral también se encuentran varios componentes de la célula hospedero, los cuales son reclutados específicamente (revisado $en^{(64-67)}$). Posiblemente uno de los más importantes es la única molécula de ARN transferencia (ARNt) incorporado en cada partícula retroviral, la cual permanece asociada via su extremo 3' terminal, al extremo 5' del ARN retroviral. Este ARNt cumple un papel clave en la replicación actuando como cebador para la síntesis de ADN por la $RT^{(18)}$. Otros ácidos nucleicos presentes incluyen varios ARNt, ARN ribosomales, y algunos ARNm. Una diversidad de proteínas también es incorporada, la mayoría mediante interacciones con Gag durante el ensamblado, y algunas colaboran en el empaquetamiento del ARNt⁽⁶⁸⁻⁷²⁾. Una fracción importante son glicoproteínas de membrana, las cuales median la unión ("attachment") inicial del virus a la célula blanco, así como modular la respuesta inmune. El número de componentes citosólicos incorporados es sumamente diverso, algunos de los más relevantes son: i) chaperonas celulares (ciclofilina A, proteínas de choque térmico, etc.), muy abundantes (aprox. 200 moléculas por partícula); ii) actina, la cual se encuentra unida a la proteína CA (en HIV-1), iii) proteínas de sistemas de tráfico celular, (componentes y reguladores del citoesqueleto, microtúbulos, transporte vesicular); iv) proteínas nucleares, (histonas, enzimas de reparación y modificación de ADN, e incluso proteínas antivirales, cuyo papel no se conoce cabalmente); v) quinasas, la actividad quinasa está relacionada con la capacidad de reclutamiento de proteínas celulares, así como en el proceso de ensamblado y desensamblado de la partícula retroviral, vi) moléculas del complejo principal de histocompatibilidad humano, muy abundantes en la superficie de la membrana, ampliamente mayoritarios a Env. Un dato relevante en este sentido es que en HIV-1 se han identificado más de 500 proteínas celulares necesarias para cumplir exitosamente las distintas etapas de la infección⁽⁷³⁻⁸²⁾. El papel del complejo arsenal de proteínas accesorias celulares recién comienza a elucidarse.

Ciclo Replicativo de los Retrovirus

La fase temprana: desde el ingreso a la célula a la integración del genoma viral. La mayoría de partículas que ingresan a una nueva célula no generan una infección productiva, y sólo una minoría (1/60.000) de la progenie serán infectivas⁽⁸³⁻⁸⁵⁾. El ciclo comienza con la interacción de la partícula con la célula hospedero (Fig 3). Esta unión ("binding") o adsorción inicial ocurre mediante interacciones relativamente inespecíficas a través de proteínas derivadas de la célula hospedero incorporadas en la membrana viral. La adsorción permite a la partícula unirse a receptores específicos en la superficie celular a través de su proteína de envoltura Env. La glicoproteína Env contiene la maquinaria de fusión que permite la entrada ("entry") del virus al citoplasma celular, sorteando la primera gran barrera: la membrana plasmática. Durante su síntesis, la forma precursora de Env trimeriza y luego es escindida por proteasas celulares de la famila furina en sus subunidades SU y TM, adquiriendo entonces capacidad fusogénica. La trimerización v escisión son fundamentales para la infectividad del futuro virión. En BLV, gp72 es escindida en gp51 y gp30 por convertasas tipo subtilisina/kexina (furinas)⁽⁸⁶⁾. El trímero de subunidades SU de Env contiene el sitio de unión al receptor (CD4 en células T para HIV-1, aún desconocido en BLV) y co-receptor, mientras que las tres subunidades TM ancladas a la bicapa son responsables de dirigir la fusión entre membranas⁽⁸⁷⁾. La unión a ambos, receptor y coreceptor, gatilla un cambio conformacional del trímero TM que expone su dominio hidrofóbico N-terminal (denominado péptido de fusión), el cual se inserta entonces en la membrana celular, permitiendo así la fusión entre las membranas viral y plasmática.

Luego de la fusión, los componentes de la partícula son liberados al citoplasma celular. Allí, la cápside conteniendo el ARN viral en su interior comienza a desensamblarse, proceso conocido como <u>desensamblado</u> o <u>desnudamiento</u> ("*uncoating*"). Este evento es crucial para el inicio de la infección, y existen diferencias importantes en el mecanismo utilizado por distintos retrovirus. Los factores que controlan/modulan este proceso son todavía muy pobremente conocidos, pero comienzan a elucidarse algunos aspectos clave. En particular, el escenario actual cambió la concepción clásica (bastante trivial), de la cápside vista como una mera envoltura, rápidamente descartada luego de entrar a una nueva célula. Esta visión, apoyada por el hallazgo de que ~40% de CA se encuentra en forma libre dentro de la partícula, consideraba a la cápside como una estructura metaestable en equilibrio con proteína CA no polimerizada⁽⁸⁸⁻⁹⁷⁾. Del punto de vista de la acción de masas, es lógico suponer que la cápside se desensamblaría al diluirse en el citoplasma⁽⁹²⁾.

No se ha establecido el sitio preciso de desensamblado, y las hipótesis abarcan desde el punto de ingreso en la membrana, hasta el contacto con el complejo del poro nuclear (NPC^a). En otras palabras, cualquier punto del citoplasma previo a la integración. Esta controversia deriva de la dificultad experimental para analizar este proceso dinámico y complejo. Existen tres grandes hipótesis: i) el



Fig. 3. Ciclo replicativo de los retrovirus. (A) Se muestran en forma esquemática diferentes eventos en el ciclo replicativo de HIV-1, el más ampliamente caracterizado, como ejemplo representativo del ciclo retroviral. En la replicación pueden diferenciarse diferentes etapas: i) adsorción ("attachment"), ii) entrada o fusión ("entry or fusión"), iii) desensamblado ("uncoating"), iv) transcripción reversa, v) integración, vi) traducción, vii) ensamblado ("assembly"), viii) brote ("budding"), ix) maduración. La etapa o fase temprana abarca los pasos iniciales de entrada a la célula hospedero, retro-transcripción del ARN genómico viral a ADN doble hebra, e integración de este ADN copia en el genoma celular. La etapa o fase tardía comienza con la síntesis y reclutamiento de los componentes virales, ensamblado de las nuevas partículas, y finalmente salida (brote) de la progenie de la célula hospedero. En la fase temprana de la replicación La entrada del virus a la célula hospedero requiere de la unión a un receptor específico en la membrana celular, y posterior fusión de la membrana vira l en la membrana plasmática (o en endosomas; no se muestra). La cápside conteniendo el genoma viral y enzimas replicativas se libera entonces en el citoplasma en donde tiene lugar la transcripción reversa del ARN y tránsito desde el citoplasma al núcleo. El ADN se integra en el ADN celular y dirige la expresión de moléculas de ARN que servirán tanto como genoma de la nueva progenie viral como de molde para la síntesis de proteínas. Se muestra en forma esquemática la organización en dominios de Gag (la cual se discute con detalle más adelante). Todos los componentes virales son transportados hasta los sitios de ensamblado de la futura progenie de partículas en proceso de ensamblado. Algunos factores celulares también son reclutados durante este proceso, ya sea porque son necesarios para completar el ciclo, o porque serán incorporados en las futuras partículas virales, siendo necesarios para comenzar la infección en una nueva célula hospedero. Tomado de ^(2, 6, 12, 14).

desensamblado ocurre rápidamente luego del contacto con el citoplasma, ubicándolo cerca de la bicapa⁽⁹²⁻⁹⁵⁾; ii) la cápside se desensamblaría progresivamente -a la vez que comenzaría la retro-transcripciónmientras es transportada hasta el núcleo ^(3, 96); y iii) la cápside se mantendría intacta hasta alcanzar los complejos de poro nuclear, en donde comenzaría la retro-transcripción⁽⁹⁷⁾. Las dos últimas hipótesis cuentan con creciente evidencia experimental, y de hecho, el bloqueo farmacológico de la RT genera una acumulación de cápsides intactas junto a los poros nucleares⁽⁹⁸⁾. A su vez, la estabilidad intríseca de la cápside parece estar evolutivamente adaptada a su función, ya que un aumento o disminución afecta profundamente la retro-transcripción y la entrada del ADN viral al núcleo^(89, 99-102). Esto concuerda con la baja tolerancia a la introducción de mutaciones en CA, las cuales típicamente generan viriones no infectivos^(3, 92, 103).

El ARN genómico viral es convertido a ADN mediante retro-transcripción (la reacción característica de los retrovirus) por parte de la RT, que sintetiza ADN copia doble hebra a partir del ARN genómico viral. La síntesis de ADN por la RT depende de un ARNt (específicamente incorporado en la partícula viral) unido al PBS del ARN genómico viral, el cual sirve como cebador. La retro-transcripción puede introducir errores ya que la TR -en comparación con las polimerasas celulares- es de fidelidad relativamente baja y no tiene actividad de corrección ("proof-reading"), lo cual genera la alta mutabilidad de virus como HIV-1. En BLV, la producción de partículas virales es sumamente baja, y se considera que su genoma se multiplica principalmente a través de la proliferación de las células B infectadas, durante la mitosis. Esto proporciona una vía adecuada para la propagación del genoma retroviral, reduciéndose al mínimo la introducción de mutaciones. Esta estrategia se adapta al ``estilo de vida'' del virus, donde la infección induce una expansión inicial de las células B, que es seguida de un estado de latencia prolongada⁽³⁷⁾. Un conjunto creciente de evidencias indican que el desensamblado esta acoplado al inicio de la retro-transcripción del ARN, si bien no se ha establecido si ésta ocurre antes o durante el desensamblado⁽¹⁰⁴⁻¹⁰⁶⁾. Desafortunadamente, se conoce casi nada sobre los factores que gatillan la retrotranscripción, si bien se cree que la exposición de la cápside a factores citoplasmáticos pudiera cumplir un $papel^{(107)}$. En particular, la cápside es permeable al pasaje de pequeñas moléculas desde el citoplasma, por lo que la entrada de deoxi-ribonucleótidos del citoplasma podría gatillar la retro-transcripción⁽¹⁰⁸⁻¹¹⁰⁾. Es probable que la formación de híbrido ARN-ADN o incluso ADN doble hebra generados por la RT desestabilicen la cápside y promueva su ruptura⁽¹¹¹⁾. Ambos mecanismos se influencian mutuamente, y la presencia de actividad RT se correlaciona con una pérdida de estabilidad de la cápside⁽¹¹¹⁻¹¹⁵⁾. Más aún, si el desensamblado de la cápside ocurre muy rápido o muy lentamente, la retrotranscripción se afecta e incluso puede inhibirse, revelando que existe un momento y velocidad óptima de desensamblado para gatillar la retrotranscripción^(116, 117). Mediante mutagénesis y/o uso de drogas capaces de desestabilizar o estabilizar la cápside, se ha comprobado que tanto la hipo-estabilidad como la hiper-estabilidad se asocia a un bloqueo de la retrotranscripción y/o afectación de la integración^(89, 104, 118, 119). En HIV-1 (el modelo más ampliamente estudiado) el desensamblado parece requerir de la participación de componentes celulares, sugiriendo que el proceso no es espontáneo. Dentro del virión de HIV-1, y luego del procesamiento de Gag, la quinasa celular ERK2 incorporada en la partícula, fosforila la Ser16 del motivo conservado Ser16-Pro17 de la proteína CA^(120, 121). La fosforilación es blanco de la peptidil-prolil *cis-trans* isomerasa Pin1, la cual induce el desensamblado⁽¹²²⁾. El papel de Pin1 es sinérgico con el de la ciclofilina A, (CypA, otra peptidil-prolil *cis-trans* isomerasa) que cumple un papel clave modulando en forma positiva o negativa el desensamblado de la cápside de HIV-1, dependiendo de la célula hospedero⁽¹²³⁻¹²⁷⁾. En células T, por ejemplo, CypA estabiliza la cápside y promueve la infección⁽¹²³⁾. Es probable que la influencia en el desensamblado de la retro-transcripción y factores celulares sea diferente entre distintos retrovirus.

El desensamblado genera partículas subvirales conocidas como complejos de transcripción reversa (RTCs^a) v de pre-intregración (PICs^b), aunque su composición precisa no se conoce. De hecho, la transición entre RTC y PIC ocurre progresivamente con el desensamblado de la cápside y muy probablemente implica múltiples complejos. En términos generales, se reconocen como RTCs los complejos con actividad RT elevada y ARN viral, mientras que los PICs son complejos competentes para integrarse en el genoma blanco, y conteniendo ADN doble hebra viral^(3, 128-130). Se ha demostrado la presencia PR, RT, IN, y Vpr en RTCs y PICs, mientras que las proteínas estructurales parecen perderse completamente⁽¹³¹⁻¹³³⁾. Varios factores celulares interaccionan con, o son incorporados en, los RTCs así como en los PICs^(92, 133). Durante la retro-transcripción, los RTCs/PICs son transportados hasta el núcleo a través de los microtúbulos, en donde interaccionan con los NPCs^(90, 134). La retro-transcripción ocurre más tardíamente (8-12 horas) que el transporte hacia el núcleo (minutos-2h), lo cual apoya la idea que la síntesis de ADN viral ocurre principalmente en RTCs unidos al poro nuclear. La evidencia más reciente sugiere que el desensamblado funcional y temporalmente correcto de la cápside es esencial para que el transporte de los RTCs y PICs prosiga correctamente hasta alcanzar el núcleo⁽¹⁰⁴⁾. La importancia del éxito temporal y funcional del desensamblado para la infectividad de los retrovirus ha generado un gran interés en desarrollar drogas que afecten el proceso.

En este sentido, es interesante notar que la estructura de la cápside es, de hecho, un blanco evolutivamente conservado de varios mecanismos de restricción antivirales sumamente efectivos (Fig. 4). La proteína TRIM5 α^c es un actor principal de la inmunidad innata intracelular que media la restricción contra infecciones retrovirales⁽¹³⁵⁻¹³⁹⁾. El mecanismo de acción de TRIM5 α (aunque parcialmente caracterizado) es sumamente interesante, ya que reconoce la cápside intacta, esto es, su organización estructural característica. De esta forma, TRIM5 α hace blanco sobre la cápside desde las etapas más tempranas luego de su ingreso al citoplasma, donde su estructura esta mayoritariamente íntegra, siendo capaz de responder inmediatamente al ingreso de retrovirus. La unión de TRIM5 α a la cápside promueve su

desensamblado rápido y prematuro, afectando y/o bloqueando la retro-transcripcion y favoreciendo su degradación proteosomal⁽¹⁴⁰⁾. La acción de TRIM5 α es específica de especie; así TRIM5 α es responsable de la incapacidad de HIV-1 de infectar células de varios monos del viejo mundo, y de la resistencia de células humanas a la infección con MLV y EIAV^(137-139, 141, 142).

Para alcanzar el genoma de la célula hospedero y poder integrarse al mismo, el genoma retroviral debe ingresar al núcleo a través de los poros nucleares. La *integración* del ADN doble hebra viral es el evento que marca el establecimiento exitoso de la infección. La enzima IN es un componente clave de los PICs, podría tener un papel en la translocación a través del poro nuclear, y permanece íntimamente asociada al ADN viral hasta el momento de su integración en el(los) cromosoma(s). CA también parece ser

necesaria para el pasaje a través del poro nuclear, y de hecho, se ha demostrado que la CA de HIV-1 se une a dos proteínas del poro nuclear -NUP358 y NUP153^{(126, 143,} ¹⁴⁴⁾. Una vez integrado, el ADN viral se denomina provirus. La función de la integrasa es fundamental para proseguir hacia la etapa tardía del ciclo. La integración es crucial para la infección productiva ya que se considera que el provirus integrado es el único sustrato que sirve como molde para la producción de nuevas moléculas de ARN viral que servirán como ARNm y ARN genómico^{(126,} ¹⁴⁴⁻¹⁴⁹⁾. Se sabe que BLV si bien se integra aleatoriamente en el genoma celular, lo hace predominantemente en regiones no codificantes⁽¹⁵⁰⁾. Por el contrario, HIV-1 se integra mayoritariamente cerca de genes activos. Esto posiblemente refleje las distintas estrategias infectivas de ambos retrovirus⁽¹⁵¹⁾. Retro-transcripción, entrada nuclear, e integración, son eventos íntimamente ligados al desnudamiento, el cual también es regulado por proteínas



Fig. 4. Esquema de los destinos de la cápside retroviral en la célula infectada. Luego de la fusión, la cápside entra en contacto con el citoplasma de la célula hospedero, pudiendo tomar diferentes caminos. La infección productiva se produce cuando la cápside en transportada exitosamente hasta el núcleo, y durante ese camino ocurre el desensamblado progresivo hasta alcanzar el NPC. Allí culmina el desensamblado y culmina (o empieza) la retrotranscripción, seguida del pasaje al núcleo del ADN viral y finalmente su integración. Por otro lado, una vez en el citoplasma, la cápside también se convierte en blanco del sistema inmune innato intracelular, a través de un conjunto de proteínas de restricción anti-retroviral. Este grupo incluye los factores Fv1 -inhibidor de la entrada al núcleo del ADN viral, TRIM5 y TRIMCypA – ambos aceleran el desensamblado de la cápside. Tanto el desensamblado prematuro/disminución de la estabilidad del "core" como la inhibición de la retro-transcripción/aumento de la estabilidad y consecuente inhibición del desensamblado, resultan en el bloqueo de los pasos subsecuentes y pérdida de infectividad. Tomado de ⁽¹⁻³⁾.

celulares.

La fase tardía: desde la producción y ensamblado de los componentes virales, a la salida del virión. La *transcripción* se inicia en el promotor y es llevada a cabo por la ARN polimerasa II así como otros factores celulares, mientras que la traducción se realiza en el citoplasma por la maquinaria de síntesis de proteínas celular^(6, 11, 152). Los transcritos de ARN viral interaccionan con la maquinaria celular de procesamiento de ARN en el núcleo de igual manera que los ARN mensajeros (ARNm) celulares⁽³⁵⁾. Parte del ARN viral sufre empalme alternativo, dando lugar a ARNm que codifican para las glicoproteínas Env y factores virales accesorios. Una proporción de los ARNs completos, no procesados, son exportados desde el núcleo al citosol mediante la vía de tráfico nucleocitoplasmático de la célula en colaboración con proteínas virales accesorias, para luego ser empaquetados en las partículas que están siendo ensambladas⁽³⁶⁾. En el caso de los retrovirus complejos, esta misma población de ARN también actúa como ARNm molde para la síntesis de proteínas ^(6, 7). Los ARN virales (no procesados y total/parcialmente procesados) dan lugar a las proteínas estructurales y enzimas (Gag y Gag-Pro-Pol)⁽⁶⁾.

Los retrovirus se ensamblan en, y brotan desde, la membrana plasmática de la célula infectada. Por tanto, todos los componentes del virión sintetizados deben ser transportados al sitio de ensamblado en la membrana plasmática. Mediante microscopía con células vivas se ha determinado una media de ~10min para el ensamblado de una partícula viral, si bien existen diferencias entre distintos retrovirus^(153, 154). El ensamblado y egreso de la célula infectada es un proceso complejo que requiere de la acción coordinada de los componentes virales y factores celulares. Gag (y también Gag-Pro-Pol) es traducida en el citoplasma, pero no forma ensamblados de gran tamaño hasta que alcanza la membrana plasmática⁽¹⁵³⁾. Sin embargo, en algún momento y lugar durante su transporte, algunas moléculas de Gag soluble pueden formar complejos con el ARN genómico viral. Estos pequeños complejos ribonucleoproteicos ARNg: Gag anteceden a la extensa polimerización que ocurre posteriormente en la membrana, y parecen ser necesarios para la correcta incorporación de las únicas dos copias que deberá reclutar cada nueva partícula (se discute más adelante). La síntesis coordinada de la proteína estructural y enzimas replicativas como componentes de las poliproteínas Gag y Gag-Pro-Pol (sintetizadas en una relación ~20:1) permite regular la estequiometria de los componentes de la partícula viral⁽¹⁵⁵⁻¹⁵⁷⁾. Gag y Gag-Pro-Pol son transportadas a la cara citosólica de la membrana plasmática a través de un mecanismo poco caracterizado pero que involucra el tráfico vía citoesqueleto. La llegada de Gag a la cara citosólica de la membrana plasmática es un evento crucial del ciclo replicativo, ya que marca el inicio de un foco de ensamblado de una futura partícula viral⁽⁶⁾. Miles de moléculas de Gag (y Gag-Pro-Pol) se unen a la membrana e interaccionan entre sí, formando un parche denso, visible por microscopía electrónica. Este foco dirige el ensamblado y promueve el reclutamiento del resto de componentes virales. El ARN genómico dimeriza y se une a unas pocas moléculas de Gag en el citoplasma, y estos complejos ribonucleoproteicos viajan entonces a la membrana 19 plasmática en donde inician la nucleación del ensamblado^(5, 158, 159). El ARN cumple entonces una función importante en promover el ensamblado. El proceso por el cual Gag se asocia y recluta a la Env presente en la membrana (específicamente el trímero de SU/TM) aún no es claro. Algunas evidencias sugieren una incorporación "pasiva" a través de la localización preferencial de Env en los microdominios de membrana de composición lipídica específica, a los cuales Gag se une y comienza el ensamblado de la nueva partícula^(160, 161). También se ha reportado interacción directa entre la cola citoplasmática de TM y el dominio MA de Gag, mientras que otros trabajos sugieren que factores celulares también podrían mediar la interacción Gag-Env^(162, 163). La membrana plasmática es la primera barrera de la infección y también es la última que el virus debe pasar para poder egresar. La salida o brote tiene lugar mediante un proceso denominado fisión en el cual la partícula naciente es separada físicamente de la célula. Para un mismo retrovirus pueden existir variaciones en el ensamblado/brote dependiendo del tipo celular. Por ejemplo, HIV-1 se replica en macrófagos por largos períodos de tiempo con escasa citotoxidad, en marcado contraste con lo que ocurre en células T⁽¹⁶⁴⁾. La interacción homotípica entre las moléculas de Gag induce curvatura en la membrana, la cual continúa hasta que adquiere la forma de una forma esférica, que contiene todos los componentes de la partícula viral. La membrana lipídica de las partículas retrovirales proviene de la membrana celular. El ensamblado ocurre en microdominios de la membrana cuya composición lipídica es diferente del resto de la membrana. Estos microdominios (denominados lipid rafts⁽¹⁶⁵⁾) aportan un ambiente particular en la bicapa que propicia el reclutamiento de, y las interacciones entre, los componente virales y de éstos con factores celulares.

Recientemente se demostró que los retrovirus utilizan la maquinaria ESCRT celular para completar la polimerización de Gag y catalizar la salida de la célula hospedero (Fig. 5) ⁽¹⁶⁶⁻¹⁶⁹⁾. ESCRT es un conjunto de más de 30 proteínas celulares involucradas en la maquinaria de ubiquitinilación, así como en la escisión de los "puentes" de membrana remanentes durante la liberación de vesículas a los cuerpos multivesiculares endosomales y en la separación célula madre-célula hija durante la división celular^(170, 171). Este proceso de escisión es topológicamente equivalente al proceso de fisión entre la membrana viral y celular. El complejo funcional básico de la maquinaria ESCRT es el heterotetrámero ESCRT-I, el cual a través de las subunidades ALIX y Tsg101 recluta una variedad de factores que finalmente llevan a cabo la fisión. La poliproteína Gag de muchos retrovirus contiene una región conservada denominada "dominio tardío" en la que se encuentran los motivos PPXY, PTAP, o YXXL, a través de los cuales reclutan la maquinaria ESCRT. Diferentes retrovirus usan diferentes combinaciones de estos motivos, localizados en distintas regiones de Gag⁽¹⁷²⁻¹⁷⁶⁾. La Gag de HIV-1 contiene un dominio (p6^{Gag}) que posee los motivos PTAP e YPXL⁽¹⁷⁷⁻¹⁷⁹⁾. El dominio denominado PTAP se une a la subunidad Tsg101 del heterotetrámero ESCRT-I, mientras que YPXL se une al factor ALIX. Los deltaretrovirus BLV y HTLV-1/2 contiene un motivo

también en MA⁽¹⁸⁰⁻¹⁸²⁾. La interacción con Tsg101 y/o ALIX del complejo ESCRT-I promueve el reclutamiento de las ATPasas VPS4A/B y las proteínas ESCRT-III de las familias CHMP1, CHMP2, y CHMP4. Las proteínas ESCRT-III polimerizan sobre la membrana formando un filamento espiral que se cierra sobre sí mismo, dando lugar finalmente a una estructura circular tipo "domo". Este "domo" parece ofrecer una superficie con la cual interaccionaría la membrana celular, separándose entonces de la membrana viral, y produciéndose así la separación virus-célula.

1.2 ENSAMBLADO Y ARQUITECTURA DE LA PARTÍCULA RETROVIRAL INMADURA

Un paso previo a la producción de una partícula infecciosa

Un punto crucial del ciclo replicativo de los retrovirus es que el virión que sale exitosamente de la célula hospedero no es infeccioso. Esto responde al hecho que la organización estructural adoptada por Gag durante el ensamblado de la partícula no le proporciona capacidad infectiva. Por este motivo, los viriones que recién brotan de la célula infectada no son infecciosos por lo que se denominan *inmaduros*. Por tanto, el ciclo replicativo no culmina con la salida de la partícula de la célula infectada, sino que la formación de una partícula retroviral infecciosa procede en dos etapas. En la primera etapa se logra ensamblar una partícula inmadura que, aunque no infectiva, egresa de la célula portando todos los componentes que constituyen un retrovirus. En la segunda etapa, uno de estos componentes, la proteasa viral, escinde a Gag gatillando un reordenamiento estructural que determinará la adquisición de completa capacidad infectiva, completando así el ciclo replicativo. Las diferencias en la arquitectura de las partículas virales inmadura y madura es un reflejo de los distintos requerimientos funcionales de cada etapa del ciclo replicativo del retrovirus^(158, 183, 184).

En esta sección se proporcionarán las principales características estructurales de Gag y su papel en la arquitectura inmadura del virión. Esto servirá como marco para luego describir el reordenamiento de esta organización inicial, cuyo evento crucial es el ensamblado y formación de la cápside, cuyo éxito determina la adquisición de la capacidad infectiva.

Gag, una maquinaria molecular de ensamblado de retrovirus

El ensamblado de la partícula retroviral ocurre mediante una serie de eventos interdependientes, separados espacial y temporalmente, y orquestado por la poliproteína Gag, la principal proteína estructural de retrovirus. Como se vio anteriormente, Gag es la única proteína necesaria para ensamblar una partícula viral^(184, 185). Asimismo, Gag recluta en forma transitoria componentes celulares que son requeridos para llevar a cabo los distintos pasos de la infección, siendo uno de los más notorios el secuestro de la maquinaria de fisión vesicular ESCRT mencionado anteriormente⁽¹⁸⁶⁻¹⁸⁹⁾. La partícula viral que brota de la



Fig. 5. La maquinaria ESCRT celular es usada para el brote de las partículas retrovirales. (*izquierda***) El esquema muestra algunos elementos clave para llevar adelante la reacción de fisión. El ejemplo corresponde a HIV-1, en el cual el mecanismo está mejor caracterizado. Gag es responsable del reclutamiento de la subunidad Tsg101 del complejo tetramérico ESCRT-1 (en rojo) el cual se une a ubiquitina (en negro), y el dominio V de ALIX. Este reclutamiento inicial promueve la incorporación de las proteínas ESCRT-III de las familias CHMP1, CHMP2, y CHMP4. Estas proteínas polimerizan formando un espiral que se cierra sobre sí mismo, formando un domo sobre cuya superficie comienza a unirse la bicapa, estrechando así el "cuello" que separa la célula hospedero del virión. La ATPasa VPS4 aporta la energía para la reacción de fisión, la cual está ligada al desensamblado del "domo" formado por ESCRT-III. (***derecha***) El modelo actual (visto en más detalle) postula que ESCRT-III (en verde) formaría una estructura tipo anillo inicial que rodearía en sitio o foco de ensamblado de Gag (A), delimitando la extensión de la futura "vesícula". Este anillo de proteínas ESCRT-III actúa como núcleo para formación del filamento de ESCRT-III. De alguna manera, vía la curvatura intrínseca del espiral o via la atracción por factores presentes en la partícula (¿la propia Gag?) el filamento se cierra progresivamente mientras que la membrana celular acompaña la disminución superficial del domo (paso B). Este proceso continúa hasta que el domo se cierra adoptando una forma cónica, y ocurre la separación de ambas membranas. Tomado de ^(5, 7, 14).**

célula hospedero tiene un diámetro aproximado de 100-200nm, el cual varía entre distintos retrovirus. Está rodeada de la bicapa derivada de la célula, y su arquitectura interior carece de la simetría icosahédrica que

posee otros virus (siendo el prototipo los virus desnudos, esto es, sin envoltura lipídica), sino que consta de una malla de complejos hexaméricos formados por el auto-ensamblado de Gag sobre la superficie interna de la membrana viral. Esta malla (a la que llamaremos por su nombre común en literatura: *lattice*) está constituida tanto por Gag como por Gag-Pro-Pol, si bien Gag es mayoritaria (~20:1, reflejando la proporción en la que son producidas como se mencionó anteriormente). Las moléculas de Gag (y Gag-Pro-Pol) asociadas entre sí formando esta *lattice* permanecen unidas a la bicapa a través de su extremo MA Nterminal, mientras que el extremo NC se orienta hacia el interior, al centro de la partícula, otorgándole a la *lattice* una configuración radial. La *lattice* formada por Gag es la estructura sobre la cual se ensambla la partícula viral y la que permite el reclutamiento de todos los factores. Esta *lattice* inicial se conoce como **inmadura**, ya que corresponde a la organización de la partícula no infectiva.

Gag es una poliproteína de aproximadamente 55 kDa que consta de tres dominios funcionales principales, plegados independientemente, denominados Matriz (MA), Cápside (CA), y nucleocápside (NC) (Fig. 6). Esta organización es común a todas las Gag, si bien existen diferencias entre distintos retrovirus. La Gag de BLV posee la arquitectura básica de tres dominios, separados por el sitio de corte para la proteasa viral. HIV-1 además posee dos péptidos espaciadores SP1 (p1) y SP2 (p2), y un pequeño dominio (p6). Estos dominios accesorios también son liberados por la proteasa y cumplen funciones importantes en el ensamblado de Gag y posterior maduración hacia la forma infectiva, así como en la interacción con factores celulares^(8, 190). El extremo N-terminal de Gag (MA) es un dominio globular de ~15-17 kDa compuesto de cinco α -hélices, uno segmento 3₁₀ corto, y un arreglo mixto de tres hojas β (⁽¹⁹⁾⁻ y revisado en ⁽¹⁹⁵⁾). MA posee una secuencia de miristilación en su extremo amino 194) (¹MGXXX(S/T)XX⁸) que señala la adición de un grupo miristoílo (ácido mirístico) un ácido graso saturado⁽¹⁹⁶⁾. Esta reacción ocurre a nivel co-traduccional mediante la acción secuencial de una metionina amino-peptidasa que remueve la metionina inicial, y subsecuentemente una N-miristoil-transferasa que une vía enlace amida un motivo miristoílo al residuo Gly-2. CA es el dominio central de Gag, y consiste a su vez de dos dominios estructurales independientes. Sus características se escribirán ampliamente más adelante ya que es el objeto de estudio de esta Tesis. NC es un dominio relativamente no estructurado de unos 12 kDa, típicamente con dos motivos "dedos de zinc ⁽³²⁾" (CysX₂CysX₄HX₄Cys) altamente conservados, cada uno de los cuales une un ion Zn y son responsables de asociarse a ácidos nucleicos.

La Gag de BLV posee 392 residuos, y consta del domino MA (p15), CA (p24), y NC (p12). Contiene un domino tardío PPXY a nivel del dominio MA, involucrado en la unión a factores celulares que facilitan las últimas etapas de la liberación de la partícula⁽¹⁸²⁾. Un hecho interesante es que durante el proceso de maduración proteolítica, MA(p15) es escindido por la proteasa viral generando MA(p10), un péptido corto de siete residuos, y p4. Por tanto, si bien se considera que la Gag de BLV exhibe la organización en dominios mínima, en términos los productos generados durante la maduración su arquitectura es NH₂-MA(p10)-p4-CA(p24)-NC(p12)-COOH. Estos fragmentos, (sólo descritos en BLV), se sabe que poseen propiedades diferenciales de unión al ARN viral (ver más delante), pero a pesar de su potencial relevancia no se dispone de más información⁽¹⁹⁷⁾.

Unión de Gag a la membrana plasmática: El dominio MA

¿Cómo se asocia Gag a la membrana plasmática? El dominio MA es el responsable de orquestar el anclaje de Gag al alcanzar la membrana. El grupo miristoílo N-terminal de MA establece interacciones hidrofóbicas con los lípidos de la membrana que le permiten insertar hasta 10 de sus 14 carbonos en la mitad interna de la bicapa ⁽¹⁹⁸⁾. El dominio MA (myrMA) existe en un equilibrio la conformacion myr(s)MA (con el miristoílo (s)ecuestrado en una cavidad hidrofóbica), y myr(e)MA (donde está

(e)xpuesto al solvente) (Fig. 7A)⁽¹⁹³⁾. La transición entre ambas formas, denominada "*myristoyl switch*", permite regular de la accesibilidad del miristoílo, y por tanto la capacidad de Gag para unirse a la

membrana blanco en el lugar y momento correcto. myrMA es predominantemente un monómero en la conformación myr(s)MA, pero a concentraciones elevadas (\geq 400µM) trimeriza y el equilibrio se desplaza casi totalmente hacia myr(e)MA^(193, 199). Durante el ensamblado de la partícula Gag alcanza concentraciones de 2-14 mM ^(193, 200), favoreciendo el anclaje al potenciar la formación de myr(e)MA⁽²⁰¹⁾. La mayoría de las Gag de retrovirus de mamíferos están miristoiladas⁽¹⁹⁸⁾, pero la inserción del miristoílo no es suficiente *per se* para lograr un anclaje estable. De hecho, las Gag de RSV y EIAV no están miristoiladas⁽²⁰²⁻²⁰⁵⁾. Un corto segmento de residuos hidrofóbicos junto al sitio de miristilación, también contribuye a la unión interaccionando con los lípidos de la bicapa. Además, una región polibásica rica en lisinas, se une electrostáticamente a las cabezas polares de los fosfolípidos ácidos (negativos) de la cara interna de la membrana, como fosfatidil-serina (PS) y PI(4,5)P₂⁽²⁰⁶⁾.

Además de la asociación mediante interacción electrostática, la unión de MA a $PI(4,5)P_2$ dispara la inserción del miristoílo a la membrana mediante un mecanismo que recientemente ha comenzado a



Fig. 6. Arquitectura modular de las proteínas Gag. Se muestra esquemáticamente la organización de las Gag de los ortoretrovirus más representativos de los distintos géneros (excepto Epsilonretrovirus y Ortoretrovirus no clasificados): <u>HIV-1</u> (Human Immunodeficiency Virus type 1), Lentivirus; <u>RSV</u> (Rous Sarcoma Virus), Alpharetrovirus; <u>MPMV</u> (Mason-Pfizer Monkey Virus), Betaretrovirus; <u>MULV</u> (Murine Leukemia Virus), Gammaretrovirus; <u>BLV</u> (Bovine Leukemia Virus), Deltaretrovirus; <u>HTLV-1</u> (Human T-cell Leukemia Virus type 1); Deltaretrovirus. Los sitios de corte proteolítico se indican con triángulos negros. La M encuadrada N-terminal indica la presencia de miristilación.

elucidarse en profundidad. El $PI(4,5)P_2$ consiste de un esqueleto de glicerol que tiene unido un ácido graso saturado en la posición 1['], un ácido graso insaturado en la 2['], y un grupo fosfo-inositol en 3[']. Este último interacciona con la región básica del dominio MA, mientras que el ácido graso saturado en 1´ se une a la región hidrofóbica N-terminal adyacente⁽²⁰⁷⁾. Por su parte, el ácido graso insaturado en posición 2´ es reconocido por el motivo ³⁰KLKH³⁴ (en HIV-1) el cual es también parte de la región básica⁽²⁰⁷⁾. El secuestro del ácido graso 2' de PI(4,5)P2 gatilla alostéricamente la liberación del grupo miristoílo, desplazando la forma myr(s)MA hacia myr(e)MA. Así, el miristoílo queda accesible para insertarse en la membrana, presumiblemente sustituyendo el sitio previamente ocupado por el ácido graso 2' de $PI(4,5)P_2$ ^(9, 20, 23-26). Gag se ancla entonces a la bicapa mediante dos puentes lipídicos: su propio miristoílo y un ácido graso de PI(4,5)P₂^(1, 7, 190). Interesantemente, la composición lipídica de la membrana retroviral presenta diferencias importantes con la de la célula hospedero. En particular, el lado interno presenta un marcado incremento de PI(4,5)P₂ (usualmente minoritario), lo cual favorece tanto el direccionamiento de Gag al sitio correcto de ensamblaje, como la estabilidad de la unión a la membrana⁽²⁰⁸⁾. Un aumento en 2% de $PI(4,5)P_2$ en la membrana incrementa 10 veces su avidez por $Gag^{(209)}$. También existe una mayor abundancia de lípidos como esfingomielina (SM), fosfatidil-serina (PS), fosfatidil-etanolamina (PE), fosfatidil-colina (PC), colesterol, y glico-esfingolípicos⁽²¹⁰⁻²¹²⁾. Los ácidos grasos en 1´ del PS, PE, y PC también son capaces de interactuar con la región hidrofóbica N-terminal de MA, en forma análoga al ácido graso 1' de $PI(4,5)P_2^{(208)}$. Como resultado, el lado interno de la membrana expone una superficie con mayor densidad de carga negativa en los microdominios donde ocurre el ensamblado, favoreciendo la unión electrostática de Gag. El efecto del colesterol no se conoce, aunque parece favorecer la disponibilidad de PS para interaccionar con Gag.

Selección y empaquetamiento del genoma viral: El papel de NC y su vínculo con MA

La incorporación selectiva y precisa de las dos copias de ARN genómico, ampliamente minoritario respecto al exceso de ARNs celulares, es llevada a cabo por el dominio NC, que reconoce y une específicamente el ARN viral^(12, 213, 214). El ARN viral posee varias secuencias de acción en *cis* para su empaquetamiento, localizadas en la región 5'UTR (Fig. 7B *izq*.). La eficiencia del empaquetamiento depende principalmente de un elemento altamente conservado de aproximadamente 150 nt, el sitio psi (Ψ), el cual es reconocido por NC mediante un mecanismo poco conocido⁽²¹⁵⁾. La especificidad depende también de la dimerización del ARN viral, lo cual expone sitios de unión a NC de alta afinidad, previamente ocultos en la hebra monomérica (Fig. 7B *der*.). La dimerización ocurre mediante apareamiento de bases en el bucle de iniciación de la dimerización (DIS^a). NC también promueve la dimerización del ARN a través de su actividad chaperona de ácidos nucleicos^(216, 217). La actividad chaperona varía entre retrovirus, siendo elevada en lentivirus como HIV-1 y muy débil en deltaretrovirus como BLV/ HTLV-1.

Aunque la unión del ARN está típicamente asociada a NC, el dominio MA también cumple un papel relevante, siendo ambos dominios necesarios para seleccionar correctamente el genoma (Fig. 7C). La Gag de BLV es un caso interesante, ya que MA forma un complejo altamente específico con el extremo 5' del ARN genómico dimérico, mientras que NC mantiene interacciones inespecíficas con el ARN⁽²¹⁸⁾. MA también se une selectivamente al sitioΨ, por lo cual cumple un papel determinante en la selección específica del ARNg de BLV. La interacción simultánea de NC y MA al ARN induce una conformación curvada en Gag, otorgándole forma de ``U`' (Fig. 7C *inset*). ⁽²¹⁹⁻²²¹⁾. En esta conformación plegada sobre sí misma, el miristoflo permanece inaccesible el bolsillo hidrofóbico de MA, lo cual puede constituir un mecanismo adicional para impedir su exposición durante el tráfico hasta la membrana, evitando la unión inespecífica de Gag a otras membranas. Por otro lado, es interesante el hecho que el dominio MA de la Gag de BLV es escindido durante la maduración proteolítica, en p10, un péptido de siete residuos (7aa), y p4. Mientras que MA(p15) forma un complejo altamente específico con la señal de empaquetamiento de la región U5-5´ del ARN, MA(p10) mantiene su capacidad de unirse a ARN pero lo hace en forma completamente inespecífica. Por su parte, tanto p4 como 7aa no tienen afinidad apreciable por ARN. La función de estos fragmentos de MA con capacidad de unión alterada al ARN blanco no se conoce.



Fig. 7. Unión de Gag a la membrana plasmática. (A) "switch" entre las conformaciones [T] (tensa, monomérica) y [R] (relajada, trimérica), donde el miristoílo se encuentra secuestrado [s] y expuesto [e] respectivamente. (B) Izquierda; región 5´UTR del ARN genómico de HIV conteniendo las secuencias de empaquetamiento. Se indican: el sitio de iniciación de la dimerización (DIS), la señal de empaquetamiento Ψ (Psi), y el sitio donador de empalme alternativo principal. Derecha; Modelo de la selección y reclutamiento del ARNg. Los sitios de unión a NC se exponen luego de la dimerización del ARNg. El complejo Gag:ARNg se une a la membrana y la proximidad entre Gag adyacentes favorece su interacción. (C) El reclutamiento del ARNg y el tráfico del complejo con Gag hasta la membrana varía entre retrovirus. En RSV, Gag oligomeriza en el citosol en complejos hexaméricos Gag: ARNg que auto-ensamblan en la membrana plasmática. En HIV-1, Gag podría alternar entre conformaciones compactas y extendidas; se piensa que MA estaría unido a un ARN celular en lugar del ARNg viral. Al igual que en RSV, los complejos GAG:ARNg de HIV-1 forman oligómeros que posteriormente ensamblan en la membrana. En BLV el dominio MA se une al sitio Ψ mientras que NC interacciona con el ARNg de manera inespecífica. No se conocen más detalles del transporte hasta la membrana y el proceso de ensamblado, tanto en BLV como HTLV-1. En HTLV-1, además, tampoco es claro el papel de MA en la interacción con el ARNg y su encapsidación. El PI(4,5)P2 juega un papel clave en HIV-1 pero no en HTLV-1, mientras que se conoce muy poco en RSV y BLV. Inset; cambio conformacional en la Gag unida al ARN vía MA y NC (forma compacta), y unida la membrana plasmática (forma extendida). Tomado de ⁽¹¹⁻¹³⁾.

Al llegar a la membrana plasmática y comenzar la oligomerización, las moléculas de Gag interaccionan lateralmente, adoptando entonces una conformación extendida en la cual el dominio MA queda accesible para unirse a la membrana. Dicho de otra forma, la bicapa compite con el ARNg por la unión de MA. De esta forma, el dominio MA posiblemente adquiera una orientación óptima para unirse a la bicapa, guiada por interacciones electrostáticas entre la región básica y los lípidos negativamente cargados. Dado el contexto particular de los focos de ensamblado es lógico suponer entonces que sea este microambiente particular el que propicie la unión y ensamblado de los miles de moléculas de Gag que alcanzan y comienzan a auto-ensamblar en la membrana, asegurando la especificidad del proceso. Estos focos proveen un escenario en el cual todos los factores mencionados anteriormente pueden actuar en forma sinérgica, gracias a la alta concentración de Gag y el ambiente aportado por el microdominio lipídico. Aún resta comprender cómo y cuando se establecen estos microdominios.

Ensamblado y organización de la lattice inmadura: el papel del dominio CA.

Como se mencionó arriba, el ensamblado de Gag da lugar a la formación de una *lattice* inmadura en el interior del virión, con las moléculas adoptando una conformación extendida que determina una disposición radial (Fig. 8)^(219, 221). La unión a la bicapa vía MA y la interacción con el ARN central vía NC mantienen este alineamiento de Gag⁽²²²⁾. La estructura de la *lattice* se mantiene unida mediante interacciones homotípicas entre las aproximadamente 5000 moléculas de Gag que la conforman⁽⁵⁸⁾. El dominio MA, además de permanecer unido a la bicapa, mantiene interacciones homotípicas MA-MA que colaboran en la formación de la *lattice* a través de la formación de trímeros⁽²²³⁾. Sin embargo, la red de trímeros de MA no forma un entramado regular continuo, por lo cual colabora sólo parcialmente a la estabilidad y ordenamiento global de la estructura. El domino NC exhibe cierto grado de dimerización, pero que tampoco contribuye significativamente en términos de interacciones proteína-proteína^(224, 225). Sin embargo, NC cumple un papel importante formando un "puente" con el núcleo central de ARN⁽²²⁶⁻²²⁸⁾. Cabe notar que para BLV no existe ninguna información sobre la formación y/u organización de la *lattice* inmadura.

El dominio determinante de la formación de la *lattice* es CA. Las proteínas CA retrovirales consisten de dos dominios estructuralmente independientemente, unidos entre sí mediante un "*linker*" flexible^(6, 7, 229). En el contexto de la poliproteína Gag, el dominio CA mantiene su organización en dos dominios individuales, denominados N-terminal (NTD) y C-terminal (CTD), unidos a los dominios MA y NC respectivamente. Más adelante se expone en detalle la estructura de esta proteína, objeto de estudio de esta Tesis. En HIV-1, la *lattice* inmadura es estabilizada principalmente mediante interacciones laterales que involucran el dominio CA y SP1^(58, 113, 224, 230). Información detallada sobre los residuos involucrados en la formación y mantenimiento de la *lattice* así como su estructura a resolución moderada (esto es, sin detalle

atómico) comenzó a surgir muy recientemente. Actualmente no se dispone de una estructura a alta resolución (esto es, con detalle atómico) de la *lattice* inmadura, lo cual ha probado ser técnicamente muy desafiante⁽²³⁰⁻²³²⁾. Sin embargo, aunque se está lejos de resolver el complicado puzzle de interacciones que dan lugar al ensamblado de Gag, la combinación de enfoques complementarios (genéticos, bioquímicos, estructurales) ha revelado aspectos importantes del arreglo que adopta la proteína⁽⁸⁾. Mediante análisis de mutagénesis extensiva se han identificado un conjunto de residuos clave para la polimerización de la Gag de HIV-1, los cuales se ubican principalmente en el dominio CTD de CA (predominantemente en la región MHR^{*a*}, una región muy conservada en todas las CA retrovirales), así como en SP1⁽⁸⁰⁾. A su vez, notables estudios de crio-microscopía electrónica (crioEM)/crio-electrón tomografía (crioET) han permitido avanzar enormemente en nuestra comprensión de la estructura 3D de Gag. La resolución alcanzada es todavía moderada, siendo la más alta (esto es, de mayor detalle) de ~8Å⁽²³²⁾. Estas técnicas resultan completamente esenciales, debido a la naturaleza intrínsecamente compleja del ensamblado, que además presenta curvatura.

Mediante crioEM de viriones inmaduros de HIV-1 purificados se demostró por primera vez la estructura de la *lattice* radial de Gag, y reveló que consiste de una malla bidimensional (2D) curva, con una organización básica hexagonal^(222, 233). Este orden hexamérico sin embargo, existe únicamente a nivel local, ya que globalmente la *lattice* carece de una simetría estricta (como simetría icosahédrica que poseen otros virus). La arquitectura hexagonal proviene de la interacción entre los dominios CA (principalmente) y SP1, los cuales forman arreglos hexaméricos empaquetados en forma relativamente "cerrada". Interesantemente, la malla formada por Gag muestra de zonas con alto grado de ordenamiento, presumiblemente correspondientes a regiones de interacciones Gag-Gag muy ordenadas, así como zonas con evidente "desorden", donde la *lattice* aparece discontinua^(58, 230, 233). Estas regiones discontinuas incluso son extensas en algunos puntos de la partícula, y se considera que son resultado de la imposibilidad de Gag de formar una *lattice* estructuralmente homogénea completa sobre la superficie esférica. Así, la *lattice* tapiza una parte (~60%) de la cara interna de la membrana formando estructuras ordenadas, existiendo también sectores desordenados e incluso libres⁽²³³⁾.

Las reconstrucciones obtenidas hasta el momento, aunque de baja resolución, muestran que las moléculas de Gag se ordenan como hexámeros separados por 8nm^(58, 234). El dominio CA-NTD hexameriza formando un anillo que presenta un "agujero" central relativamente grande, mientras que los dominios CA-CTD se ubican por debajo. Los CTDs, a su vez, mantienen interacciones homodiméricas con los CTDs de hexámeros vecinos^(58, 230, 234-236). De esta forma, la *lattice* se mantiene mediante ejes de simetría local 6 (6-

^(a) Major homology region



Fig. 8. Estructura de la *lattice* **inmadura de Gag. (A)** A la *izquierda* se ve un corte central de una imagen obtenida mediante crioET de un virión de HIV-1 en proceso de brote. A la *derecha*, sobre la misma imagen tomográfica, se muestra una reconstrucción de las posiciones de los hexágonos de Gag en la *lattice* inmadura. Se indica en escala de color el grado de ordenamiento en el empaquetamiento de los hexámeros, desde un orden bajo (marrón) a un orden alto (amarillo). En (B) se muestra un modelo de la estructura de la Gag completa (la cual aún no ha sido resuelta), generado a partir de las estructuras atómicas de los diferentes dominios individuales que constituyen la poliproteína de HVI-1 [estas sí, ya disponibles: MA en rojo (1UPH.pdb), CA completa monomérica en verde –NTD- y azul –CTD- (3NTE.pdb), SP1 (α-hélice hipotética), NC en naranja (1MFS.pdb)]. **(C)** Esquema de la organización en dominios independientes de Gag. **(D)** Se muestra un esquema del arreglo radial de la *lattice* inmadura mantenida por Gag en la partícula viral inmediatamente luego de salir de la célula. El modelo [el mismo que en (C)] muestra a la Gag en una conformación extendida tal cual estaría dispuesta en la partícula. **(E)** Modelo mostrando un detalle de la organización hexamérica de Gag en la *lattice* inmadura, mantenida principalmente por los dominios CTD de CA. Se aprecia también los trímeros de MA anclados a la membrana plasmática. Tomado de^(7, 8).

fold) formado por hexamerización via NTDs, y 2 (2-fold) formado por los dímeros CTD-CTD. En otras palabras, puede visualizarse la malla como una red de hexámeros de NTDs, que se unen a los hexámeros vecinos vía dimerización de sus CTDs.

Las estructuras obtenidas por crioEM y crioET sugieren un mecanismo para lograr la curvatura de la *lattice* de Gag. Una malla 2D hexagonal perfecta es estructuralmente uniforme, y se extiende manteniendo una conformación plana. Lograr la curvatura de esta *lattice* requiere de i) la inserción de "defectos" pentaméricos, y/o ii) introducción de declinaciones entre hexámeros adyacentes. En otras palabras, se

30

puede curvar la *lattice* quebrando de alguna manera la uniformidad de su simetría. La existencia de "islas" de Gag junto con regiones vacías (o desordenadas) podría ser suficiente para mantener regiones uniformemente ensambladas a pesar de la restricción que impone la curvatura continúa⁽²³⁶⁻²³⁸⁾. De hecho, la forma relativamente cónica que exhiben los hexámeros de la Gag extendida podrían favorecer un ordenamiento curvo del ensamble (Fig. 8E). En concordancia con esto, se sabe que, *in vitro*, Gag ensambla en presencia de ARN formando ensambles esféricos. El análisis de las regiones ordenadas y desordenadas de Gag, tanto de partículas purificadas como de ensamblados *in vitro*, muestran sólo interconexiones hexaméricas manteniendo la *lattice*. La curvatura de la lattice inmadura proviene entonces de una malla incompleta de hexámeros, por lo que no se requiere de la incorporación de pentámeros⁽²³⁰⁾.

1.3 MADURACIÓN, ENSAMBLADO, Y ARQUITECTURA DE LA CÁPSIDE RETROVIRAL

La cápside viral: solución elegante y eficaz con recursos limitados

Los genomas virales poseen sólo unos pocos genes para cumplir la diversidad de funciones de su ciclo replicativo. Parte de la solución es el pleiotropismo de sus proteínas, y la capacidad de usurpar componentes de la célula hospedero. La protección del material genético impone otro problema, ya que la cápside, consta de varios cientos o miles de moléculas. Ante este dilema, los virus emplean una estrategia sumamente elegante (y exitosa): una única proteína (o unas pocas) capaz (capaces) de auto-ensamblar y formar una estructura supramolecular cerrada⁽¹⁸⁾.

Las cápsides se clasifican en tres tipos, según su simetría. Las *cápsides icosahédricas* (el prototipo) están presentes en muchos viriones esféricos, y pueden describirse en términos de los principios descritos en forma pionera por Caspar y Klug^(239, 240). El icosaedro es una estructura isométrica, con 12 vértices pentaméricos y 20 caras triangulares (ANEXO 2). Los virus icosaédricos más simples poseen 60 subunidades iguales, unidas mediante interacciones idénticas (denominadas equivalentes). También existen cápsides icosaédricas más grandes, gracias a un exquisito "truco" molecular: las subunidades de algunos virus son capaces de unirse mediante interacciones similares, aunque no idénticas. Estas interacciones, llamadas *quasi-equivalentes* responden a la teoría de la *quasi*-equivalencia deducida por Caspar y Klug. Así, pequeñas variaciones de un mismo "lenguaje" permiten expresar una enorme adaptabilidad estructural. Otros virus poseen cápsides helicoidales, con sus subunidades empaquetadas en un arreglo espiral que les otorga forma tubular. Varios miles de copias idénticas se ensamblan alrededor del ADN o ARN genómico viral, manteniendo interacciones *quasi*-equivalentes, de manera análoga a las cápsides icosaédricas.

Si bien la mayoría de cápsides poseen una de estas dos simetrías, algunas exhiben formas irregulares o complejas. La estructura de estas cápsides es globalmente asimétrica, no pudiendo describirse mediante los principios de simetría helicoidal o icosaédrica. Estas cápsides exhiben estructuras variables (morfotipos), incluso para un mismo virus, lo cual dificulta su caracterización. En esta categoría se encuentran los retrovirus, los poxvirus, y algunos virus extremófilos⁽²⁴¹⁻²⁴⁶⁾. Las cápsides complejas/irregulares poseen una o pocas subunidades, que también se unen mediante interacciones *quasi*-equivalentes. Algo no definido aún es si el número de interacciones quasi-equivalentes que pueden mantener las cápside irregulares es mayor que en las cápside simétricas. Esto se debe a que aún no se ha elucidado la estructura 3D de ninguna cápside irregular a resolución atómica; en el capítulo 3 se discute sobre este punto a partir de los resultados obtenidos.

Las cápsides son metaestables, esto es, son lo suficientemente estables para otorgar protección adecuada al genoma, pero su estructura también permite el desensamblado en el momento y lugar correcto^(3, 5, 58, 166, 183, 230, 235). Este pasaje puede ser gatillado por diferentes factores. En algunos virus sin envoltura el desensamblado se dispara por cambios conformacionales al unirse a un receptor⁽²⁴⁷⁻²⁵⁶⁾. En retrovirus parecen existir un conjunto de factores, aún escasamente comprendidos. En todo caso, es claro que los bloques de construcción se unen mediante interacciones suficientemente débiles para garantizar reversibilidad, mientras que el alto número de copias involucradas (1500 se estiman en HIV-1) adiciona un factor importante para lograr estabilidad: la suma de un gran número de interacciones débiles, en otras palabras, cooperatividad^(1, 17, 58, 213, 257-259).

Maduración de la *lattice* inmadura

Como discutimos, las partículas emergen de la célula hospedero como un virión inmaduro no infeccioso. Poco después, tiene lugar un profundo proceso de reorganización estructural, a través del cual la partícula adquirirá la estructura madura definitiva que le concederá capacidad infectiva (Fig. 9)^(154, 158, 260). Este proceso se denomina **maduración**, y es gatillado por la proteasa viral, la cual escinde a Gag en sus dominios independientes, resultando en la pérdida de la estructura inmadura (la *lattice* de Gag) y la reorganización del "*core*" de la partícula; los complejos Env en la membrana permanecen inalterados. El mecanismo por el cual la proteasa se activa luego del ensamblado de Gag y brote de la partícula no se conoce. Hasta ese momento, la proteasa permanece inactiva, en forma dimérica. Algunos datos sugieren que motivos estructurales generados luego del ensamblado de la *lattice* inmadura podrían gatillar la activación de Gag, presumiblemente promoviendo su auto-procesamiento a la forma activa^(261, 262). Durante la maduración, MA permanece asociada a la cara interna de la membrana vía su miristoílo e interacciones complementarias. NC, la "histona" viral, permanece unida al ARN genómico y colapsa hacia el centro de la partícula, formando un núcleo o *core* ribonucleoproteico. Este nuevo empaquetado genera un cambio conformacional del ARN que estabiliza el dímero, un elemento importante para mantener la nueva estructura del virión ⁽²⁶³⁻²⁶⁸⁾.



Fig. 9. Diferencias en la organización estructural del virión maduro e inmaduro de HVI-1. La organización de la partícula inmadura no infecciosa es notoriamente modificada En **(a)** se muestra un modelo de la estructura de la Ga completa (generada como se describe en la Fig. 8), resaltando la extensión de los distintos dominios y (con puntas de flecha) los sitios de corte proteolítico. En **(b)** y **(c)** se muestra un modelo esquemático de la organización del virión inmaduro y maduro respectivamente. En **(b)** se aprecia la lattice ordenada, conteniendo la Gag completa, y se resalta la presencia de zonas "desnudas". En **(c)** es evidente el cambio que genera la escisión de Gag, y se ve el recubrimiento parcial del lado interno de la membrana por MA, la formación de la cápside vía auto-ensamblado de CA, y el core central de NC unida al ARN genómico. En **(d)** y **(e)** se muestran micrografías electrónicas de partículas virales crio-vitrificadas, inmadura y madura respectivamente. En **(d)** se observa claramente la organización radia de la lattice inmadura de Gag, más ordenada en la región cercana a la membrana correspondiente a la zona de interacción Gag-Gag. Hacia el centro permanece el ARN predominantemente desordenado. En **(e)** se ve la cápside ya formada, evidenciando la presencia de un virión maduro infeccioso. Tomado de^(2, 5, 6).

La proteína de cápside

Una vez libre del contexto de Gag, la proteína CA auto-ensambla formando una estructura supramolecular hueca que encierra y protege el genoma: la cápside. Este evento es crucial, ya que marca el momento en que el virión adquiere capacidad infectiva. La comprensión del mecanismo molecular que subyace la maduración y formación de la cápside es un área de intensa investigación debido a su relevancia para el desarrollo de anti-retrovirales.

Las CA retrovirales son proteínas de 24-26 kDa con dos dominios independientes, amino-terminal (NTD, *N- terminal domain*) y carboxi-terminal (CTD, *C-terminal domain*), unidos por un

``linker`´ flexible de unos 4-6 residuos, sin estructura definida. Ambos dominios son globulares y están formados casi exclusivamente por alfa-hélices, siete en el domino NTD (α 1- α 7) y cuatro en el dominio

CTD (α 8- α 11)⁽²⁶⁹⁻²⁷²⁾. La mayoría de información estructural proviene de las estructuras resueltas por cristalografía de rayos Х 0 resonancia magnética nuclear (NRM^{a}) de los dominios individuales de varios retrovirus; no existe ningún dato para la CA de BLV (^(1, 5, 9, 10, 17, 191, 230, 271-277)v ⁽⁵⁾). Las revisado en únicas nativas completas estructuras disponibles son la forma monomérica de CA_{HTLV-1} (mediante NMR) y un dímero anti-paralelo de CA_{HIV-1} (mediante cristalografía de rayos X)^(9, 278). El corte proteolítico en Gag que separa los dominios MA y CA, genera una prolina (conservada) amino-terminal en CA (Pro1)^(275, 279). La Pro1 puede



Fig. 10. Efecto del procesamiento proteolítico de Gag en la unión MA-CA. (A) Se muestran para ls retrovirus RSV, HTLV-1, EIAV, y HIV-1, las estructuras obtenidas mediante difracción de rayos X o NMR de los dominios NTDs (cuatro superiores) así como cTDS (cuatro inferiores). (B) Se muestra (en cartoon) la estructura de los dos dominios de CA de HIV-1, NTD y CTD, formando parte de Gag (Inmadura) y luego del corte. Se puede apreciar como extremo amino-terminal del NTD se encuentra en una conformación extendida cuando CA forma parte de Gag (hacia el extremo amino se encuentra MA). En la CA libre madura, este segmento se pliega formando un β -hairpin mediante unión salina Pro1-ASp51. Tomado de ^(1, 2).

entonces formar un puente salino vía su grupo imino con un residuo aspartato (también conservado, Asp51 en HIV-1, Asp53 en BLV) en la hélice $\alpha 3^{(276)}$. Esto induce el plegamiento de los primeros 13 residuos formando una estructura conocida como β-hairpin (Fig. 10). Así, la maduración convierte esta región de CA previamente en una conformación extendida en Gag. En HIV-1 se sabe que la formación del β-hairpin ocurre no inmediatamente luego de la escisión MA-CA, sino hasta que también se escinde la unión CA/SP1^(5, 17, 280). La formación del β-hairpin parece ser común a todos los retrovirus^(2, 7). Tipicamente, las proteínas de cápside presentan baja homología: su identidad aminoacídica es de aproximadamente 20% entre diferentes géneros y 30-40% en CAs del mismo género. Una excepción saliente es un elemento altamente conservado tipo *sheet-turn-helix* situado al inicio del CTD, que presenta una alta identidad en todos los retrovirus y se denomina región de máxima homología (MHR, *major homology región*). Asimismo, dos cisteínas del CTD, localizadas fuera del MHR, también son altamente conservadas entre retrovirus ⁽²⁸¹⁾. Notablemente, a pesar de su baja identidad, la estructura terciaria de la proteína (así como su función) se encuentra conservada en todas las CA de los diferentes géneros de retrovirus.

¿Cómo se ensambla la cápside retroviral?

Las CA de lentivirus como HIV-1 existen como dímeros estables a concentraciones mayores a 10-20 μ M; de hecho CA_{HIV-1} cristaliza como dímero según se observó al resolverse su estructura 3D por difracción de rayos X⁽²⁸²⁾. Junto a la información estructural, estudios de mutagénesis dirigida (específicamente, *alanine scanning*) y de biofísica en solución, mapearon la región de interacción a un grupo de residuos hidrofóbicos del CTD. Esta interacción es relevante, ya que mutaciones puntuales en estos residuos bloquean el ensamblado *in vitro* y disminuyen drásticamente la infectividad ⁽²⁷⁶⁾. Se ha propuesto que el estado de oxidación de las cisteínas conservadas del CTD podría tener un papel en determinar el estado de oligomerización basal, esto es, en condiciones no polimerizantes. Específicamente, en CA_{HIV} y EIAV la formación de un puente disulfuro orienta las hélices 10 y 11 de forma tal que se expone la superficie hidrofóbica responsable de la dimerización. Más aún, el reemplazo de una o ambas cisteína(s) por serina en CA_{HIV-1} impide el ensamblado de la partícula viral y disminuye la infectividad ⁽²⁸³⁾. Al contrario de las CA lentivirales, las CA de HTLV-1 (la más homóloga a la CA de BLV) y RSV (virus de sarcoma Rous) existen en solución como monómeros, hasta una concentración de -al menos- 1 mM ⁽²⁷⁸⁾. De manera interesante, la estructura 3D obtenida por NMR de CA_{HTLV-1} reveló que ambas cisteínas se encuentran reducidas y la ausencia de un puente disulfuro imposibilita la exposición de esta superficie hidrofóbica.

Los estudios realizados usando proteínas CA recombinantes puras en ensayos in vitro han aportado un conjunto importante de información sobre la reacción de ensamblado in vitro y los factores que lo influyen, valiéndose del hecho que la proteína CA recombinante purificada auto-ensambla espontáneamente en condiciones adecuadas^(17, 35, 284-288). Los ensambles "tipo-cápside" formados in vitro recapitulan fielmente el arreglo que mantiene la proteína en la cápside auténtica, y son cruciales para disecar el efecto de diferentes parámetros sobre la reacción de auto-ensamblado, así como para análisis estructurales^(10, 226, 269, 284-286, 288). Las curvas de ensamblado in vitro evaluando la formación de polímeros de CA mediante medida de DO a 340 nm (turbidimetría) han resultado particularmente útiles. El análisis de estas curvas muestra que la reacción procede siguiendo una cinética sigmoidal, con una fase de retraso (lag) la cual es típica de reacciones de polimerización que requieren de la formación inicial de especies intermediarias que sirvan como núcleos para la nucleación. Sin embargo, hasta ahora el único intermediario identificado corresponde a la especie dimérica del monómero de CA_{RSV}, el cual per se es capaz de gatillar la reacción (287). Conocer los intermediarios de reacción es de gran interés no sólo para comprender la vía de ensamblado de la cápside sino también para diseñar drogas antirretrovirales que interfieran con su formación. La mayor parte de la información sobre las variables que gatillan el ensamblado procede de estudios con la CA_{HIV-1}, la cual autoensambla en presencia de concentraciones altas (mayor a 1M) de NaCl y muestra una fuerte dependencia del pH (ver por ej. (284)). El análisis por microscopía electrónica de transmisión de los productos de ensamblado muestra que en estas condiciones, la proteína auto-ensambla formando estructuras tubulares, de diámetro constante y longitud variable. Otros compuestos son capaces

de gatillar el ensamblado con grado variable de efectividad, como ser: otras sales, pH ácido, ácidos nucleicos, agentes de crowding, lípidos, aunque no se conoce cabalmente el efecto de otras variables fisiológicas como la temperatura, ya que la mayoría de experimentos implican el ensamblado a 25°C o 4°C (201, 282, 288, 289). Recientemente, Carter y cols. demostraron que los aniones divalentes, en particular fosfato, son inductores potentes del ensamblado de CA_{RSV} y en menor medida de CA_{HIV-1}, sugiriendo que el mecanismo de gatillado implique vías diferente⁽²⁸⁷⁾. Para CA_{RSV} la reacción es afectada, además, por el pH y la fuerza iónica, sugiriendo un proceso gobernado por control electrostático, posiblemente mediante el enmascaramiento de cargas. Sin embargo, el hecho que el efecto del fosfato no es reproducido por una fuerza iónica similar aportada por NaCl, sugiere que el mecanismo específico es más complejo. Una observación interesante es que cuando la reacción es inducida por fosfato dianiónico, las estructuras formadas pierden el hábito tubular y forman ensambles cerrados que son descritos como similares a cápsides nativas. Una hipótesis probable es que el fosfato simule de manera más efectiva las condiciones de carga y polianiones (como el propio ARN) del interior de la partícula viral. Los dominios individuales, NTD y CTD, de la CA_{HIV-1} son incapaces de oligomerizar en solución para formar hexámeros o pentámeros -bloques de construcción de la cápside- incluso en condiciones que promueven en ensamblado de la proteína completa. Asimismo, tampoco se ha detectado la formación de oligómeros discretos durante la polimerización de la CA_{HIV-1} completa.

Arquitectura de la cápside madura: múltiples enfoques para una estructura desafiante

A nivel estructural, las cápsides retrovirales son intrínsecamente pleiomórficas, y adoptan múltiples formas entre diferentes retrovirus (Fig. 11)^(32, 35). De hecho, el análisis de cápsides maduras de un mismo virus muestra que ni siquiera dos cápsides individuales podían superponerse (esto es, no hay dos idénticas)^(58, 290). La morfología más conocida es la cónica adoptada por HIV-1 y compartida por todos los lentivirus; los beta retrovirus presentan formas cilíndricas, y varios géneros (incluyendo el grupo BLV/HTLV) exhiben morfologías poliédricas/esféricas⁽²⁸⁶⁾. El pleiomorfismo de las cápsides retrovirales, a diferencia de las icosaédricas, impide elucidar su estructura completa intacta a resolución atómica mediante los abordajes disponibles (los cuales requieren simetría). Específicamente: i) el *linker* genera gran flexibilidad entre dominios); ii) los ensambles biológicamente relevantes (hexámero y pentámero) tienen baja estabilidad; y iii) existe una fuerte tendencia a oligomerizar formando ensambles muy heterogéneos.
Los primeros estudios de Ganser, Li, y Heymann, revelaron que la arquitectura de la cápside de HIV-1 puede describirse en términos de los principios geométricos de un cono fullereno^(291, 292). Estos modelos se basan en los conceptos arquitectónicos desarrollados por Buckminster Fuller, cuya construcción se basa en bloques de construcción hexaméricos V pentaméricos^(258, 284, 293, 294). El avance en la elucidación estructural de la cápside provino de múltiples abordajes. En particular, la resolución de las estructuras atómicas de los dominios individuales de las CA de varios retrovirus, y el hecho que la proteína recombinante autoensambla in vitro formando estructuras análogas a las cápsides auténticas^{(17, 73, 108, 177, 258, 284, 285, 289,}



Fig. 11. Morfologías representativas de cápsides retrovirales. (A) (*Arriba*) Se ven micrografías electrónicas de partículas virales maduras de Mo-MLV, MPMV, y HIV-1. (*Abajo*) Modelos de las estructuras de las cápside, todas organizadas según la geometría tipo fulereno (**B**) Micrografías electrónicas de BLV. (*Arriba*) un virión en proceso de brote. Se ve la zona electrón-densa correspondiente a la *lattice* inmadura. (*Abajo*) Una partícula viral libre, madura. Tomado de⁽⁵⁾.

²⁹⁵⁻²⁹⁸⁾. El estudio de estos ensamblados in vitro y de cápsides nativas purificadas (mediante crioEM y crioET) reveló que a pesar de la notoria asimetríasupramolecular de la cápside, la estructura consiste de una lattice formada por un bloque de construcción (capsómeros) básico hexagonal, que se cierra con capsómeros pentaméricos^(290, 299-301). La cápside consta de aproximadamente 1,500 monómeros de CA, ensamblados en 250 hexámeros y exactamente 12 pentámeros^(10, 302). La *lattice* es un arreglo hexagonal flexible que puede curvarse, lo que permite adaptarse a su variación continua de curvatura. Esto implica un cambio continuo en el ambiente local de interacciones que debe mantener el capsómero (particularmente el hexámero, mayoritario). Por tanto, el bloque de construcción es visto como un "continuo" de estructuras relacionadas (confórmeros). Esto permite generar rupturas locales (defectos) en la simetría de la lattice, lo cual se expresa en la asimetría global de la cápside. La visualización por TEM de partículas retrovirales aisladas muestra también la existencia de cápsides con múltiples capas, lo cual coincide con los ensambles *in vitro* que muestran paredes múltiples o tubos unidos lateralmente⁽²⁸⁶⁾. Otro hecho interesante es que un porcentaje de CA no se encuentra formando parte de la cápside, sino (aparentemente) unida a la membrana⁽⁵⁸⁾. La distribución específica de los pentámeros determina el morfotipo o geometría de la cápside: la forma cónica HIV-1 surge de la presencia de 7 pentámeros en el extremo más ancho, y 5 en el más angosto. En la forma tubular hay igual número de pentámeros en cada extremo, y en las formas poliédricas/esféricas la distribución es isométrica^(35, 235, 292, 293). Actualmente se acepta que la arguitectura de todas las cápsides retrovirales se basa en un principio común de lattice tipo fullereno, constituidas de capsómeros hexaméricos y pentaméricos^(1, 2, 7, 17, 158, 183, 235, 258, 259, 289). Los primeros datos estructurales de la lattice hexamérica a una resolución moderada (~9Å), provienen del trabajo de Ganser-Pornillos y cols., quienes resolvieron su estructura 3D a partir de cristales 2D de CA_{HIV-1} mediante crioEM⁽¹⁰⁾. Estos proceden de partículas esféricas de gran tamaño ensambladas in vitro, las cuales colapsan y aplastan al depositarse sobre grillas para su análisis mediante crioEM, comportándose como cristales 2D. Así, la estructura es representativa de la lattice hexamérica nativa. Valiéndose de este modelo de crio-EM de la red 2D de CA_{HIV-1}, Pornillos y cols. diseñaron un panel de mutantes, introduciendo cisteínas en residuos cercanos (≤ 6 Å según el modelo pseudo-atómico) en las interfaces NTD-NTD, potencialmente responsables de la hexamerización. De esta manera, se obtuvo un conjunto de mutantes capaces de ensamblar in vitro formando tubos idénticos que la proteína nativa, pero unidos covalentemente si las cisteínas se encuentran en posición favorable para interacionar. A su vez, se mutaron los dos residuos clave para la dimerización CTD-CTD entre hexámeros vecinos (W184A y M185A en la hélice α9, se describe más adelante), con el objetivo de evitar la polimerización (en principio, sin afectar la hexamerización)^{(275,} ³⁰³⁾. Estos mutantes de CA permitieron obtener tanto hexámeros como pentámeros, solubles y estables (Fig 12). Con algunos se obtuvieron cristales que difractaron adecuadamente y se resolvió su estructura 3D mediante difracción de rayos X. Este trabajo marcó un hito, ya que por primera vez fue posible visualizar con detalle atómico ambos capsómeros de la cápside retroviral, si bien provienen de proteínas modificadas incapaces de ensamblar. Tanto para el hexámero como para el pentámero, las estructuras resueltas a partir de diferentes mutantes (por tanto, diferentes cristales) fueron casi idénticas, sugiriendo que las modificaciones introducidas no perturbaron la conformación de cada bloque de construcción (Fig 12 y 13).

Se observaron tres interacciones que dan lugar a la formación del capsómero (hexámero o pentámero) y a su unión con capsómeros vecinos para formar una malla bidimensional. Brevemente, la hexamerización o pentamerización ocurre vía unión entre NTDs (interface NTD-NTD), y la unión a capsómeros vecinos ocurre mediante la dimerización entre sus CTDs (interface CTD-CTD); el capsómero se estabiliza además mediante la conexión entre el CTD de un protómero con el NTD del protómero vecino en el mismo bloque de construcción (ver por ej. ^(5, 260, 277, 292)). Conjuntamente, estos datos apoyan la descripción de la cápside retroviral como una estructura construida a partir de los principios matemáticos de la geometría de los fullerenos.

La interfaz NTD-NTD. El capsómero hexamérico exhibe un anillo central de NTDs bastante rígido, rodeado por un "cinturón" de CTDs que aparece relativamente móvil (Fig 14). El anillo central de NTDs es simétrico, formado por interacción de las tres primeras hélices α (α 1, α 2, y α 3) de cada protómero, dando lugar a un "barril de 18 hélices" central. La α 1 se ubica en la posición más central del anillo hexamérico, y define el poro o hueco central del capsómero. El anillo se mantiene mediante interacciones entre las hélices α 1/ α 3 de un protómero con las hélices α 1/ α 2 del protómero vecino. Así cada protómero aporta un *barril de 3 hélices* que interacciona con los dos protómeros que lo flanquean. Un estudio reciente sugirió que la formación del β -harpin durante la maduración proteolítica es necesaria para la formación de la hélice α 1, crucial para la hexamerización. Concretamente, la comparación de las estructuras resueltas mediante NMR



Fig. 12. *Estructuras cristalográficas del capsómero pentamérico y hexamérico de HIV-1*. En *(a)* y *(d)* se muestran las estructuras (vistas desde arriba, con las hélices representadas como cilindros) del hexámero y pentámero de HVI-1. El anillo de NTDs se muestra en naranja y el anillo de CTDs en azul. La posición de las cisteínas introducidas para el *cross-linking* de los protómeros se indica con esferas amarillas. En *(b)* y *(e)* se muestra una visión equivalente, representados en *cartoon*. En *(c)* y *(f)* se muestran los capsómeros rotados 90°, vistos de costado, y representados en cartoon. Esta organización es muy similar a la observada en el pentaméro, revelando la naturaleza *quasi*-equivalente de las interacciones en ambos capsómeros. En el recuadro (A) se ve una vista lateral del hexámero estabilizado mediante puentes disulfuro. Se resalta cada protómero en una color distinto; los dominios NTDs se muestran en todo claro, y los CTDs en tono oscuro. Se indican al costado los anillos formados por los NTDs y CTDs. En el recuadro (B) se ve el mismo hexámero, representado como superficie, indicándose la posición de las hélices $\alpha 1/\alpha 2/\alpha 3$. Nótese la posición centra del la $\alpha 1$.Tomado de ⁽¹⁷⁾.

de las estructuras de la CA de EIAV (un lentivirus) nativa (o sea, con su β -harpin correctamente formado) y una versión conteniendo una cola de seis histidinas en el extremo amino (en la cual el β -harpin no puede formarse), reveló que en ausencia del β -harpin, la hélice α 1 está mayoritariamente desplegada y potencialmente incapaz de polimerizar⁽³⁰⁴⁾

La organización cuaternaria del pentámero es muy similar a la del hexámero. En el pentámero, las hélices $\alpha 1/\alpha 2/\alpha 3$ de cada protómero forman la superficie de asociación con los protómeros adyacentes formadon un *barril de 15 hélices*^(1, 10, 277). La asociación entre NTDs es la principal responsable de la hexamerización. La formación del pentámero requiere de un nuevo ángulo cada *barril de 3 hélices*, pero los contactos involucrados son esencialmente idénticos, resaltando la quasi-equivalencia de ambos

capsómeros, según los principios de Caspar y Klug⁽²⁴⁰⁾. Esto permite la transición hexámero-pentámero mediante sólo una pequeña rotación de los NTDs, específicamente a nivel del barril de 3 hélices que conforma la superficie de unión (Fig. 13).

La interfaz NTD-CTD. En el capsómero (tanto en el hexámero como en el pentámero), el CTD de un protómero contacta el NTD del protómero advacente^(10, 289, 305). Está formada por interacción entre la a4 del NTD de un protómero y la α8 del CTD del protómero advacente. Algunos contactos adicionales se establecen entre la α 7 del NTD y la $\alpha 8/\alpha 11$ del CTD. El punto de contacto consiste de algunos puentes de hidrógeno (1-3 residuos) que conforman una interacción "helix-capping" (puentes de hidrógeno entre residuos localizados en los extremos de las hélices que interactúan). Estos contactos actúan como un punto sobre el cual pivotan ambos



Fig. 13. Quasi-equivalencia de los contactos en protómeros del capsómero retroviral. En (a) y (b) se muestran las estructuras (vistas desde arriba) del hexámero y pentámero de HVI-1, resaltando cada protómero en diferente color. Se indican las hélices de la tríada de contacto $\alpha 1/\alpha 3$ de un protómero y $\alpha 2'$ del protómero adyacente. (a**b)** Es evidente el cambio de ángulo entre protómeros adyacentes en el pentámero (72°) y el hexámero (60°). La transición hexámeropentámero se logra mediante un movimiento muy sutil de los NTDs. Esto se observa perfectamente en (c), donde se muestran superpuestas la tríada del hexámero y la triada H1/H3 (en azul claro) y H2[′] (naranja claro) del pentámero. En (d) se muestran superpuestas las tríadas H1/H2/H3 de dos NTDs adyacentes, para el pentámero y el hexámero. La posición aproximada de cada NTDs se indica por la superficie azul y naranja. La flecha indica la rotación que ocurre durante la transición hexámero a pentámero (mismo código de colores). Tomado de ⁽¹⁷⁾.

dominios, a la vez que limitan la movilidad del CTD respecto al NTD. Así, el anillo móvil de CTDs se ancla sobre el anillo rígido de NTDs, mientras que la interface NTD-CTD permite la rotación de cuerpo rígido entre ambos anillos.

La interfaz CTD-CTD. El CTD es conformacionalmente flexible, y puede dimerizar adoptando una variedad de conformaciones^(271, 306-308). La dimerización tiene lugar principalmente a través del empaquetamiento simétrico vía las hélices α 9 de cada CTD, específicamente mediante los residuos Trp184 y Met185. Algunas interacciones hidrofóbicas entre la hélice 3₁₀ de un CTD, y la hélice α 9 del otro CTD^(15, 88, 272, 276, 309). Varias estructuras cristalinas han sido resueltas y revelan ligeras variaciones en la conformación del CTD así como de la interfaz de dimerización. En particular, la estructura del dímero de CTD resuelto por NMR y difracción de rayos X revela diferencias en la estructura terciaria (particularmente a nivel de la hélice de dimerización α 9) así como en la estructura cuaternaria (específicamente, en el ángulo de entrecruzamiento de las hélices α 9 en la díada CTD-CTD. A su vez, la estructura del CTD monomérico no es idéntica a la del CTD en el dímero. Estos resultados indican que el CTD puede adoptar diferentes conformaciones así como distintos ángulo relativos respecto al eje de simetría 2 que relaciona ambos CTD en el dímero.

Globalmente, las interacciones proteína-proteína que mantienen los capsómeros y la *lattice* unida son relativamente débiles y posiblemente transitorias. La interacción NTD-NTD es muy baja, ya que no se han observado (ni aislado) hexámeros de CA ni de NTD recombinantes purificados⁽²⁹⁵⁾. En el anillo de NTDs (la superficie de interacción más grande) los contactos constan de sólo un pequeño núcleo hidrofóbico rodeado de contactos polares débiles, en su mayoría "puenteados" moléculas de agua, y escasos contactos hidrofóbicos. Globalmente, la baja afinidad de las interacciones CA-CA individuales, balanceadas con la avidez de un gran número de interacciones, proporciona a la estructura la estabilidad adecuada para cumplir su función de ensamblado/desensamblado en el momento y lugar correcto.

El elusivo mecanismo de generación de curvatura en la cápside

La curvatura de la cápside retroviral varía continuamente en toda su superficie. Esto implica, como se mencionó anteriormente, la necesidad de una ruptura constante de la simetría a nivel local. Cerca de los pentámeros, la declinación aumenta, mientras que la malla hexagonal (la mayor parte de la cápside) debe adaptar un amplio rango de declinaciones entre hexámeros adyacentes. Esto se observa claramente evaluando el cambio en el ángulo que relaciona hexámeros adyacentes. Este ángulo, denominado "de

mordida" (*bite angle*) entre hexámeros vecinos varía desde ~180° en las regiones más planas de la cápside, hasta ~135° entre dos hexámeros vecinos conectados a un pentámero. Son varias las fuentes que permiten lograr esta plasticidad estructural. Por un lado, la flexibilidad intrínseca de la proteína, a través del *linker* y la variabilidad conformacional de cada dominio independiente. Esta flexibilidad también confiere a las CA retrovirales capacidad para multimerizar en diferentes condiciones físico-químicas. Por otro lado, la rotación que ambos dominios pueden mantener entre sí en el capsómero a través de la interfaz NTD-CTD. Este punto de anclaje y rotación entre ambos dominos es un elemento crucial para canalizar la curvatura entre hexámeros vecinos. Los puentes de hidrógeno *"helix-capping"* que unen ambos dominios NTD-CTD otorgan estabilidad a la unión, y al mismo tiempo permiten la flexibilidad rotacional requerida. Finalmente, la unión de la díada CTD-CTD es también flexible. Ambos dominios pueden rotar entre sí alrededor del eje 2 que los relaciona, el cual es aproximadamente perpendicular al plano del hexámero. Así cada capsómero



Fig. 14. La *lattice* hexagonal 2D (plana) de CA de HIV-1. (*A*) Vista superior de la estructura 3D cristalográfica de una capa bidimensional de hexámeros de CA_{HIV-1}, la cual recapitula la *lattice* hexagonal madura de la cápside auténtica. El anillo de NTDs se muestra en naranja y el anillo de CTDs en azul. Esta estructura se obtuvo a partir de cristales de un mutante de cisteínas que cristalizó fortuitamente formando planos de hexámeros unidos entre sí mediante dimerización de sus CTDs. Se aprecia como la malla se mantiene unida únicamente vía contactos diméricos CTD-CTD entre hexámeros vecinos. Como se mencionó en el texto, la misma organización se observó en la estructura obtenida a partir de cristales 2D, reforzando la idea que en la malla hexamérica plana, los capsómeros sólo se unen mediante interfaces diméricas (ejes 2) CTD-CTD. En (*B*) se muestra la estructura del dímero CTD-CTD obtenida a partir de dos cristales distintos. En (*C*) se muestra en detalle los contactos polares y puentes de agua que median la interacción entre dos protómeros adyacentes (uno en azul, y uno en naranja). En amarillo se indican los posibles puentes de hidrógeno. Se resalta la interacción NTD-NTD' y la interacción CTD-NTD'. En (*D*) se muestra los puentes de hidrógeno "hélix-capping" que conforman la interfaz NTD-CTD. Para simplificar, se muestan (representadas en cilindros) las hélices α7/α 8/α11 del CTD, y las hélices α3/α4/α7 del NTD' vecino. Adaptado de ^(1, 2, 9, 10).

puede adoptar rotaciones relativamente independientes de su vecino, pero los contactos entre capsómeros permiten transmitir esta información local a mayores distancias. Globalmente, se considera que las interfaces NTD-CTD y CTD-CTD son la clave para generar la curvatura, siendo los movimientos y adaptaciones locales los determinantes de la morfología y asimetría global de la cápside⁽¹⁷⁾.

La resolución de la estructura de tubos ensamblados in vitro de CA de HIV-1 mediante crioEM ha aportado elementos indispensables para visualizar (y comprender) como las rotaciones de cuerpo rígido entre ambos dominios logran curvar la malla hexamérica. La primera reconstrucción de crioEM de estos tubos, compuestos únicamente de hexámeros, reveló que el contacto entre los capsómeros adyacentes en una superficie curva involucra tanto contactos diméricos CTD-CTD (ejes 2) como triméricos CTD-CTD (ejes 3) (Fig. 15)⁽¹⁵⁾. Esto contrasta con la organización de la *lattice* hexamérica plana que mencionamos anteriormente, en donde los hexámeros vecinos se unen sólo mediante contactos diméricos CTD-CTD. Unos pocos residuos cargados por fuera del parche hidrofóbico también contribuyen a la unión. La interfaz trimérica entre CTDs parece ser exclusiva de la lattice curva, e involucra principalmente interacciones entre residuos de la cara hidrofóbica de la hélice α10 anfipática del CTD. Por tanto, se piensa actualmente que los contactos entre CTD en los ejes 2 constituyen la unión básica de la malla, mientras que los contactos entre CTDs en los ejes 3 colaboran aportando estabilidad adicional cuando la lattice se curva. En este escenario, es probable que la ausencia/presencia de ejes pseudo-3 entre CTDs de tres hexámeros vecinos varíe (responda) a la curvatura variable de la lattice en la cápside nativa. La transición hexámeropentámero parece depender (al menos en HIV-1) del control electrostático mediado por una arginina altamente conservada (Arg18 en HIV-1) en retrovirus, incluyendo el grupo BLV/HTLV^(17, 277). Esta arginina se localiza en el comienzo de la hélice al, posicionándose hacia el centro mismo del anillo de NTDs, tanto en el hexámero como en el pentámero (Fig. 15). La mutación Arg18Ala altera la morfología del ensamblado, favoreciendo la formación de esferas y conos además de los típicos cilindros. El cambio de Arg18 por un residuo hidrofóbico voluminoso genera una predominancia de esferas, indicando que la estabilización del anillo de NTD está vinculada a la curvatura de la *lattice*^(10, 302). La mayor abundancia relativa de esferas indica un favorecimiento por la introducción de pentámeros, por lo que se piensa que el residuo Arg18 conservado juega un papel crucial en la transición^(17, 179, 302). En el hexámero, el anillo central mantiene 6 argininas cercanas, cuya carga positiva es balanceada por un anillo de carga negativa de aspartatos, y que se localiza inmediatamente por debajo, también en la hélice α 1. Este residuo (Asp21 en HIV-1) está conservado en retrovirus, y también se posiciona hacia el centro del anillo de NTDs. La formación de un pentámero implica un aumento en la repulsión electrostática, debido al acercamiento de las argininas en el anillo pentamérico. Es lógico suponer que el contexto local del ensamblado (balance de fuerzas/cooperatividad) determine la manera en que el control electrostático favorece la introducción de pentámeros para soportar la curvatura. Si bien se han logrado avances notables en nuestra comprensión estructural de la cápside, existen todavía aspectos fundamentales sin resolver.

En particular, aún no se ha logrado obtener una estructura a resolución atómica del arreglo hexamérico de la proteína CA madura intacta. Por otro lado, aún no se conocen cabalmente las bases



Fig. 15. Contactos entre hexámeros en la lattice curva. Las interacciones entre capsómeros hexaméricos en la *lattice* curva involucran contactos CTD diméricos y triméricos. Los datos mostrados en **(a-b)** fueron los primeros que demostraron la existencia de ejes 3 en la *lattice* hexagonal curva, si bien a una resolución relativamente baja resolución (>9Å). (a) Vista superior del mapa de densidad (superficie) -determinada mediante crioEM- de un nanotubo ensamblado in vitro de CA_{HIV-1} [similar al que se ve en (c)], donde se muestra un acercamiento del contacto entre tres hexámeros vecinos. En óvalos rojos se indican las interfaces diméricas CTD locales (ejes 2). En triángulos rojos se indican las interfaces triméricas CTD locales (ejes 3). En hexágonos rojos se indican los anillos hexaméricos NTD locales (ejes 6). Se aprecia claramente la presencia de ejes 2 y 3 de CTDs vecinos manteniendo la lattice. La presencia de contractos triméricos no deja los "huecos" que sí son visibles en la lattice hexagonal plana. En (b) se ve en detalle la interfaz trimérica de CTDs, resaltándose los residuos involucrados en el contacto. El ajuste ("docking") de las estructuras 3D de alta resolución disponibles de los NTD y CTD en el mapa de crioEM permitió generar este modelo pseudo-atómico de la lattice hexamérica curva. Muy recientemente, se logró generar un nuevo modelo pseudo-atómico a resolución moderada (8Å), (la mejor resolución alcanzada al momento), reconstruido a partir de nanotubos de CA_{HIV-1} (c). En la reconstrucción (d) se puede observar la presencia de contactos diméricos y triméricos (indicados con flechas amarillas). Nuevamente se aprecia como la presencia de ejes 2 y 3 genera un empaquetamiento muy "cerrado" de la malla hexamérica. El acercamiento en (e) muestra el modelo pseudo-atómico (representado como cartoon: NTDs en azul, CTDs en naranja) de mayor resolución, superpuesto con el mapa de densidad de la crioEM (en gris). Se aprecian tres hexámeros vecinos, contactando a través de la dimerización de sus CTDs y la presencia de una interfaz donde convergen tres CTDs (el eje 3). Los anillos de NTDs no contactan entre sí. (f) Un detalle de la interfaz hidrofóbica trimérica, resaltando la interacción en el eje 3 vía el núcleo hidrofóbico formado por las tres hélices 10 α anfipáticas, y el anillo externo de contactos electrostáticos. Tomado de ^(15, 16).

biofísicas del mecanismo de auto-ensamblado y tampoco se ha logrado obtener información estructural a resolución atómica de ensambles biológicamente relevantes de proteínas de cápside completas no modificadas⁽³¹⁰⁾. Finalmente cabe destacar el hecho que, la importancia del ensamblado en la infectividad ha generado un enorme interés en su mecanismo, convirtiendo a las CA retrovirales en un atractivo blanco terapéutico^(229, 311). Al presente, se han diseñado varios péptidos que hacen blanco en distintas regiones de CA_{HIV-1}. Estos péptidos y sus derivados han mostrado ser eficaces para inhibir el ensamblado o bloquear el desensamblado de la cápside en ensayos *in vitro*; también disminuyen drásticamente la infectividad del virus en ensayos *in vitro* con cultivos celulares. Varios de estos compuestos ya han sido patentados y se evalúa su actividad como anti-retrovirales.

Sobre esta Tesis: ánimo y objetivos

La meta amplia de esta Tesis -a largo plazo en nuestro laboratorio- es sentar las bases para el estudio de la biología estructural de Deltaretrovirus, particularmente BLV, con foco en la comprensión de las bases biofísicas y estructurales que gobiernan el ensamblado de la partícula viral. La formación de la cápside de BLV es un terreno inexplorado, y de hecho no se conoce casi nada para HTLV-1, su homólogo más cercano. Esta tesis comenzó con un escenario marcadamente fundacional, abordando la pregunta más básica: ¿cómo se gatilla el proceso de auto-ensamblado? Además del interés en BLV, el estudio del mecanismo de ensamblado de la cápside deltaretroviral, al momento desconocido, era interesante per se, ya que podría aportar elementos novedosos, no revelados en otros modelos. En este sentido, a pesar del notorio avance (particularmente en los últimos cinco años) en la comprensión de la cápside retroviral, quedan aún varias preguntas fundamentales no respondidas a nivel biofísico y estructural, tanto para BLV como para retrovirus en general. En particular: ¿cómo se modula el auto-ensamblado? ¿Cómo se favorece la generación de pentámeros que permitan formar estructuras cerradas, similares a las cápsides nativas? ¿Cómo se correlaciona el ensamblado in vitro con un contexto fisiológicamente relevante, esto es, similar al interior de la partícula viral? ¿Cómo es la estructura 3D del capsómero hexamérico nativo, el bloque básico de construcción? ¿Cuáles son los contactos nativos que median la interacción proteína-proteína que une los capsómeros y permiten el crecimiento (polimerización) de la lattice madura? ¿Dónde subyace la flexibilidad del capsómero, que permite a los ensamblados adoptar distintas morfologías?

Nuestra hipótesis fue que la formación de la cápside de BLV debe ocurrir mediante principios similares a los de otros retrovirus, esto es, basada en una arquitectura tipo fullereno. Sin embargo, suponemos que la proteína debe exhibir diferencias intrínsecas que se reflejan en sus propiedades de autoensamblado y/o estructurales de sus bloques de construcción, que últimamente determinan las características morfológicas propias de la cápside de deltaretrovirus (como ser su morfología esférica). Consideramos que el estudio de la biología estructural del ensamblado de la cápside del BLV aportará información que será muy valiosa para retrovirus en general, pudiendo aportar elementos novedosos aún no revelados en estudios con otras CA retrovirales.

Objetivo General y Objetivos Específicos

<u>El objetivo general de esta Tesis fue caracterizar, por primera vez en este retrovirus, las propiedades</u> <u>biofísicas de auto-ensamblado y elucidar la estructura tridimensional (cristalina) a alta resolución de la</u> <u>proteína CA del BLV</u>. Nos propusimos caracterizar los determinantes moleculares del proceso de autoensamblado de la cápside del BLV, enfocándonos en: i) sus propiedades biofísicas (cinética, termodinámica) de la polimerización de CA *in vitro*; ii) morfología ultra-estructural de las estructuras supra-moleculares auto-ensambladas *in vitro*, y iii) estructura 3D a resolución atómica de CA de BLV. Nos propusimos correlacionar estos conocimientos para comprender cómo la proteína es capaz de autoasociarse, de forma específica y ordenada, para dar lugara la cápside, y cómo la estructura 3D la proteína almacena esta información.

Nuestros objetivos específicos fueron:

1) *Identificar y caracterizar factores fisicoquímicos que gatillan y/o modulan el auto-ensamblado de* CA_{BLV} . Alcanzar este objetivo nos permitirá comenzar a comprender de qué manera el ambiente fisicoquímico modula la formación de la cápside, particularmente bajo las condiciones biológicas en la cuales ocurre la formación de la cápside (específicamente: el interior de la partícula viral).

2) Elucidar la estructura tridimensional de CA_{BLV} mediante difracción de rayos X de cristal único. Una comprensión precisa de las interacciones proteína-proteína que dan lugar a la estructura cuaternaria de la cápside retroviral requiere de estructuras a alta resolución (atómica) de ensambles biológicamente relevantes de la proteína CA, que mimeticen las interacciones de la cápside auténtica.

De acuerdo a los objetivos planteados, las preguntas más relevantes que buscamos responder son:

¿Qué factores gatillan y/o modulan que la proteína auto-ensamble en forma espontánea?

¿Qué tipo de estructuras pueden formarse mediante auto-ensamblado *in vitro* y cuál es su organización ultraestructural?

¿Cómo es la estructura 3D de CA_{BLV} y/o sus dominios individuales?

 $¿Qúe regiones/residuos de CA_{BLV}$ median las interacciones proteína-proteína que determinan el mecanismo de auto-ensamblado?

CAPÍTULO 2

Auto-ensamblado de CA_{BLV}

Previa al capítulo:

"Self-assembly properties of Bovine Leukemia Virus capsid protein: implications for deltaretroviral assembly"

Se resume (es español) el contexto del trabajo, los principales hallazgos, y sus implicancias.

En este capítulo se presentan los resultados obtenidos sobre las propiedades de ensamblado *in vitro* de la proteína de cápside del BLV. Este trabajo está actualmente culminado y los resultados fueron capitalizados en un manuscrito que será enviado para considerar su publicación en una revista internacional con referato. Concretamente, el manuscrito se titula *"Self-assembly Properties of Deltaretrovirus Capsid Protein"* Gonzalo Obal, Federico Carrión, Xiaokang Xhang, Lorena Tomé, Gonzalo Moratorio, Sergio Bianchi, Jean Lepault, Otto Pritsch; será enviado a la revista Retrovirology. Previamente, los resultados iniciales obtenidos, presentados en una edición especial de la misma revista: Meeting Abstract, *"In vitro characterization of Bovine Leukemia Virus capsid Protein Self-assembly"* (2011) Retrovirology. <u>Gonzalo Obal</u>, Jean Lepault, Federico Carrión, Lorena Tomé, Gonzalo Moratorio, Gonzalo Rama, Sergio Bianchi, Jotto Pritsch.

Como se mencionó en la introducción, se han estudiado las propiedades de auto-ensamblado de una variedad de proteínas CA retrovirales, siendo la más ampliamente caracterizada la CA de HIV-1. En estos estudios se ha trabajado con las proteínas recombinantes purificadas, las cuales tienen la capacidad de autoensamblar *in vitro* al ser expuestas a condiciones fisicoquímicas favorables, como ser fuerza iónica, pH, etc. La polimerización ocurre en forma espontánea, sin la necesidad de la presencia de otros factores celulares, tales como otras proteínas virales o la presencia de chaperonas celulares. Los ensambles formados *in vitro* tienen además la propiedad de representar fielmente las estructuras nativas, esto es, exhibiendo la misma organización de las cápsides auténticas. La formación de polímeros de gran tamaño (hasta varios ciento de nanómetros) abre una puerta para estudiar del proceso en forma relativamente sencilla, mediante el uso de un ensayo turbidimétrico. En este ensayo, la polimerización es seguida mediante dispersión de la luz a una longitud de onda adecuada (usualmente 340nm o superior). A esta longitud de onda la proteína en su estado oligomérico basal no dispersa la luz hasta que la polimerización comienza: a partir de ese momento, comienzan a formarse especies de mayor tamaño que sí dispersan la luz. A medida progresa la reacción, aumenta la dispersión (turbidez) proporcionalmente (en condiciones favorables optimizadas previamente) a la velocidad y extensión del auto-ensamblado. De esta manera, haciendo uso de la turbidimetría, es posible estudiar el efecto de un gran número de variables sobre la reacción de forma relativamente high-throughput. Esto es particularmente beneficioso, ya que la turbidimetría consume una gran cantidad de proteína (orden de varios miligramos por ensayo). A pesar que las CA retrovirales ensamblan in vitro se conoce un rango relativamente acotado de condiciones. En particular, la abrumadora mayoría de datos provienen del ensamblado de la CA de HIV-1 a alta fuerza iónica (1-3M) aportada por una sal monovalente (casi exclusivamente NaCl, y en casos excepcionales KCl). Otras variables (pH, cationes y aniones, aditivos) son analizadas en forma más o menos sistemática siempre en presencia de alta fuerza iónica. Este escenario también se repite en trabajos donde se estudiaban las CA de otros retrovirus. Globalmente, la información disponible apuntaba únicamente al requerimiento de alta fuerza iónica para lograr gatillar el auto-ensamblado. Este escenario implicaba dos problemas potenciales. Por un lado, obviamente la presencia de una alta actividad salina en la reacción hacía difícil (al menos a priori) establecer cuál era el impacto de otras variables, muchas veces aparentemente enmascaradas por la sal. Esto sin embargo resultaba inevitable, ya que ninguno otro factor aparecía como inductor adecuado. Por otro lado, el requerimiento aparentemente absoluto de alta concentración de sal era difícil de conciliar con un contexto fisiológicamente relevante.

Dado que la cápside se forma en el interior del virión, es (por lo menos) lógico considerar que las condiciones prevalentes en citosol son más cercanas a la de la partícula. Por tanto, la alta fuerza iónica monovalente era interesante en cuanto a que inducia ensamblado, pero no era (en principio) representativa de las condiciones intra-virión. Por otro lado resultaba intrigante que ninguna otra condición pudiera soportar el ensamblado *in vitro*. ¡En particular las condiciones de fuerza iónica y pH fisiológicas resultaban pésimos inductores! Asimismo, los ensambles formados *in vitro* consisten casi exclusivamente de tubos, que como hemos mencionado, constan únicamente de hexámeros. Por tanto, la mayoría de condiciones estudiadas no aportaban los factores requeridos para inducir la formación de los pentámeros capaces de formar estructuras cerradas similares a las cápsides. El hallazgo de que la presencia de altas concentraciones de aniones divalentes (tales como fosfato y sulfato, ~0.5M) era un fuerte promotor del auto-ensamblado de la CA de RSV era interesante, ya que sugería un potencial papel del ARN viral, vía su carga negativa. Más aún, de estos iones daban lugar a ensambles esféricos, indicando la formación de pentámeros. Sin embargo, la CA de HIV-1 únicamente formó nanotubos idénticos a los obtenidos con NaCl. Más aún, en presencia de agentes de *`crowding´* macromolecular (los cuales mimetizan la elevada

concentración de macromoléculas (>400mg/mL) del interior del virión) también se forman exclusivamente tubos a alta fuerza iónica.

En esta Tesis comenzamos el estudio del auto-ensamblado de la cápside del BLV sin antecedentes específicos en literatura, o propios. Por tanto, primeramente pusimos a punto un protocolo para expresión y purificación de la proteína CA nativa en Escherichia coli (ver "CA_{BLV} is monomeric in solution"). Como mencionamos antes, la proteína auténtica posee un β-hairpin N-terminal típico cuyo papel en la polimerización no se conoce cabalmente. Por tanto produjimos la CA recombinante completa nativa (sin etiquetas que faciliten su enriquecimiento), y optimizamos un protocolo donde obtenemos ~100mg de proteína/L de cultivo. Estudiamos entonces el comportamiento de la proteína en solución, a efectos de establecer su estado de oligomerización basal, encontrando que existe como monómero. Posteriormente, estudiamos sistemáticamente un amplio rango de variables fisicoquímicas, y logramos definir cuáles son determinantes del auto-ensamblado. En particular, encontramos que la reacción depende principalmente de la temperatura y tiene un óptimo a la temperatura del hospedero bovino (38°C) y pH fisiológico (ver "CA_{BLV} self-assembly is a strongly temperatura-dependent process, compatible with a nucleation-driven mechanism"). Sin embargo, en una variedad de condiciones, la fuerza iónica fisiológica resultó un potente inhibidor del ensamblado, incluso en condiciones muy favorables de temperatura y concentración de proteína (ver "Modulation of CA_{BLV} assembly by solution variables: effect of ionic strength, pH, and salts"). Esto nos enfocó en nuestro interés de establecer el correlato entre el ensamblado in vitro y el contexto fisiológico, aspecto no resuelto en el campo. En este camino, evaluamos distintos contextos orientados a emular factores presumiblemente presentes en el interior del virión. Los resultados que presentamos describen este proceso. Culminamos este trabajo definiendo que es principalmente la fuerza iónica monovalente el principal inhibidor de la polimeración de CA_{BLV}, incluso en el rango fisiológico. Sin embargo, distintos factores nativos tales como la presencia de ARN (también dianiones) y agentes de crowding resultaron (para nuestro interés y sorpresa!) potentes promotores del ensamblado, incluso capaces de contrarestar la inhición salina (ver "Macromolecular crowding or RNA are essential for triggering capsid self-assembly in a cytosolic-like medium and regulates assembly morphology"). En base a este resultado. pudimos demostrar que la polimerización no ocurre en un tampón ``cytosolic-like''; sin embargo la reacción se gatilla en presencia de ARN y distintos tipos de crowders, completándose en pocos minutos (interesantemente, coincidente con el tiempo de maduración de las partículas una vez brotan de la célula) cuando la CA se encuentra a una concentración similar a la del virión. Más aún, encontramos que en presencia de estos factores, la proteína modifica su hábito de ensamblado, pasando de formar mallas bidimensionales a generar partículas esféricas de tamaño similar al de cápsides auténticas.

En resumen, en esta parte del trabajo de Tesis pudimos definir (por primera vez hasta donde conocemos) cuáles son los factores que gatillan y modulan el ensamblado de la CA de BLV, y que es

lógico suponer tengan un impacto predominante en el contexto del virión. Consideramos que los resultados obtenidos en esta parte de la tesis constituirán un aporte valioso para comprender cómo se forma la cápside del BLV. En términos más amplios, este conocimiento tendrá impacto en la biología estructural del ensamblado deretrovirus, y aportará información relevante para el diseño de drogas anti-retrovirales que hacen blanco sobre la formación de la cápside.

Índice de Figuras Capítulo 2

Figure 1. Production of recombinant CA _{BLV}	73
Figure 2. Size-exclusion chromatography-MALLS analysis of CA _{BLV}	74
Figure 3. Analysis of higher-order self-assembly of CA _{BLV}	75
Figure 4. Electron micrographs of CA _{BLV} assembly	76
Figure 5. Effect of different ions on higher-order self-assembly of CA _{BLV}	77
Figure 6. Characterization of CA _{BLV} self-assembly induced by phosphate	78
Figure 7. Effect of macromolecular crowding on CA _{BLV} self-assembly reaction	79
Figure 8. Effect of macromolecular crowding on CA _{BLV} self-assembly reaction	80
Figure 9. Effect of RNA on CA _{BLV} self-assembly reaction	81
rigure 9. Effect of KIVA on CABLy sen-assenioly reaction	

Self-assembly Properties of Deltaretrovirus Capsid Protein

("draft" del artículo en preparación, a ser enviado Retrovirology)

Gonzalo Obal 1,2*, Federico Carrión 1, Xiaokang Xhang 3, Lorena Tomé 1,2, Gonzalo Moratorio 4, Sergio Bianchi 1,5 Jean Lepault 6, Otto Pritsch 1.2§

1 Institut Pasteur de Montevideo, Unit of Protein Biophysics, 2020 Mataojo, 11400 Montevideo, Uruguay

2 Departmento de Inmunobiología, Facultad de Medicina, Universidad de la República, Avda. General Flores 2125, 11800 Montevideo, Uruguay.

3 Institut Pasteur, Unité de Virologie Structurale, Département de Virologie and CNRS Unité Mixte de Recherche 3569, 25 Rue du Dr Roux, 75724 Paris Cedex 15, France.

4 Laboratorio de Virología Molecular, Centro Investigaciones Nucleares, Facultad de Ciencias, Universidad de la República, Iguá 4225, Montevideo 11400, Uruguay

5 Departamento de Fisiopatología, Hospital de Clínicas, Universidad de la República, Montevideo,

6 Centre de Recherche de Gif, Laboratoire de Virologie Moléculaire et Structurale, CNRS (UPR 3296),

Avenue de la Terrasse, 31198, Gif sur Yvette Cedex, France.

§Corresponding author

Abstract

After budding from the host cell, retroviral virions undergo a major structural rearrangement (maturation) that ultimately leads to the formation of a supramolecular *core* known as capsid, which surrounds and protects the viral ARN as well as its replicative enzymes. Successful morphogenesis of an infectious virus relies on proper formation of the capsid, and failure to form a structurally correct core is fatal to virus replication. Capsids are formed by self-assembly of thousands of copies of the capsid (CA) protein, organized as a basic hexameric *lattice* that closes by incorporating 12 pentamers. Purified recombinant CA proteins can spontaneously assemble *in vitro*, and studies on CAs from several retroviruses have provided fundamental insights on the reaction properties. A key knowledge gap in our understanding that arises from these experiments is the difficulty to reconcile the *in vitro* data with the physicochemical context found by CA inside the virion, presumably closest to that of cytosol. Particularly, assembly at near-physiological conditions is poorly efficient for most CA proteins tested *in vitro*, showing no tendency to form capsid-like particles.

Here we study the *in vitro* self-assembly properties of the recombinant CA protein of Bovine Leukemia virus, an oncogenic *Deltaretrovirus* (HTLV/BVL group) that chronically infects B cells of bovines. Using MALLS, we determined that CA_{BLV} exist in solution as a monomer, at concentrations up to 1mM, which is close to the expected concentration inside the particle. Interestingly, self-assembly in strongly dependent on temperature, with CA remaining as an stable monomer below 20°C and increasing its polymerization rate concomitantly with temperature up to an optimum around 38°C (indeed the bovine host normal body temperature). Assembly is also concentration-dependent and exhibits a sigmoidal progress, suggesting that the polymerization mechanism requires nucleating species. Polymerization shows an optimum close to pH 7-8 but, contrary to CA_{HIV-1} , assembly is favoured at low ionic strength and is strongly inhibited near to, and above physiological values. Surprisingly, we find that a cytosolic-like medium (CLM) strongly inhibited the initiation of assembly, even at concentrations as high as 1mM CA_{BLV} . Thus, we test the hypothesis that specific intra-viron conditions could be determinant in triggering polymerization under near-physiological conditions. Notably, the addition of both, crowding agents or RNA, at a concentration resembling that of the intraviral particle bypass this inhibition and polymerization became strongly stimulated. Moreover, under this specific regime, the predominant assembly habit shifted toward formation of spherical particles, which size and shape closely resembles that of authentic capsids.

Our results are consistent with a model whereby the physicochemical environment provided inside the virion particle is indeed the key determinant for initiation of retroviral capsid assembly and modulation of the capsid structural morphology. These findings contribute to better understand the determinants of CA assembly in a physiologically relevant scenario.

Introduction

Bovine Leukemia Virus (BLV) and Human T-cell Leukemia Virus (HTLV) are structurally and functionally related oncogenic deltaretroviruses within the subfamily *Orthoretroviridae*, inducing inflammatory and/or lymphoproliferative diseases ^(4, 43). While HTLV-1 targets human T-cells, BLV infects B-cells of cattle, causing a life-long infection that generate a major animal health problem worldwide due to important economic losses ⁽³¹²⁾.

Like in other retroviruses, assembly of BLV/HTLV virions is driven by Gag, a polyprotein precursor composed of three major domains: MA (matrix), CA (capsid), and NC (nucleocapsid). After particle budding, the virus-encoded protease cleaves Gag and releases the individual domains: N-terminally myristoylated MA remains anchored to the viral envelope, NC condenses with the viral RNA, and CA spontaneously self-assemble forming a closed structure, named capsid, which surrounds and protects the ribonucleoprotein complex. This dramatic structural rearrangement known as maturation is essential for infectivity, and thus constitutes an attractive target for novel antiretroviral strategies ^(229, 273, 281, 313).

X-ray and NMR structures of several retroviral CA proteins showed that, despite their low sequence identity, they share a common 3D fold, with two independently-folded α -helical domains [named N-terminal domain (NTD) and C-terminal domain (CTD)] connected by a short flexible linker^(5, 108, 285). The basal –unassembled- state of CA_{HIV-1} (*Lentivirus*) is a dimer, whereas those of EIAV (*Lentivirus*), HTLV-1,

RSV (*Alpharetrovirus*), and JSRV (*Betaretrovirus*), exist as monomers in solution^{(314, 315), 11(316)}. The actual figure describes the mature capsid using the principles of fullerene geometry, as a lattice composed of a net of hexamers formed by homotypic NTD-NTD and heterotypic NTD-CTD interactions, held together via dimerization of neighbouring CTDs^(5, 58, 73, 108, 260, 277, 317, 318). CA can also assemble pentamers, which allows introducing curvature in the hexagonal lattice and finally forming a closed hollow structure. As a consequence of the specific distribution of pentamers in the lattice, the mature capsid is intrinsically pleiomorphic and adopt different shapes across genera, such as polyhedral spheres in deltaretroviruses (HTLV/BLV) and alpharetroviruses (e.g RSV), cones in lentivirus (e.g HIV), and cylinders in betaretroviruses (e.g MPMV) ^{(5, 318),(313), (319)}.

Retroviral CA proteins are capable of assembling in the absence of cellular factors, and most of our understanding comes from *in vitro* studies using purified recombinant protein in well-defined solution conditions. This set up have provided an invaluable tool for dissecting the effect of different parameters on self-assembly. Kinetic studies with the CA proteins of HIV-1 and RSV -by far the paradigms of assembly studies- showed that polymerization proceed through a rate-limited nucleation-driven process, which is supposed to be a common pathway to all retrovirus^(259, 284, 285, 287-289, 295). This mechanism exhibits a sigmoidal kinetic profile, with a first slow/disfavoured step (lag phase) in which nucleating species are formed, and imposing a critical protein concentration below which polymerization will not occur⁽³²⁰⁾. These nuclei act as templates, seeding the rapid addition of subunits (exponential growth phase) and accretion of the polymer. Consumption of non-assembled protein continues until equilibrium with higherorder oligomers is reached, causing a deceleration and a subsequent steady state (stationary phase). CA_{HIV}. 1 assembly in vitro depends greatly upon pH and requires to proceed a very high ionic strength (1 to 2.5 M NaCl) (108, 279, 284, 285, 295, 297). Under most conditions studied, the protein assembles almost exclusively forming hollow tubular arrays (nanotubes), indicating a marked preference for hexameric associations ^{(5,} ³¹⁸⁾. Divalent anions, such as phosphate and sulphate, can also promote oligomerization of RSV and HIV CA proteins $^{(287)}$. Notably, dianions induce CA_{RSV} to assemble into spherically-closed particles resembling

authentic capsids, indicating that they are sufficient to boost formation of pentamers as well as hexamers, a hypothesis confirmed by structural studies⁽²⁷⁷⁾. Interestingly, CA_{HIV} only form tubes similar to those seen under high-salt conditions, indicating a predominant hexameric array. The differential impact of solution conditions in modulating the final architecture of CA proteins of different genera possibly mirrors retrovirus-specific determinants of the pathway, even though they all arrive to broadly similar morphological outcomes.

Clearly, the conditions typically used for *in vitro* assembly do not reflect the physiological context inside the viral particle, despite they somehow create an environment that recapitulate the structures of native capsids^(2, 5, 58, 108). The intrinsic tendency to assemble of CA proteins previously characterized is insignificant under close to physiologic conditions, suggesting the absence of intra-particle regulating conditions. Recently, some lines of evidence have highlighted the potential impact of physiological factors on *in vivo* assembly. For instance, a lipid mixture mimicking the composition of the viral envelope of HIV-1 greatly accelerates CA_{HIV-1} polymerization⁽³²¹⁾. Macromolecular crowding agents (interacting-inert macromolecules which surrogate the highly-concentrated intracellular/intra-particle conditions) also stimulate polymerization of CA_{HIV-1}⁽²⁹⁷⁾. Importantly, addition of crowding agents at neutral pH achieves polymerization of the protein at the physiological-like ionic strength of 150 mM NaCl instead of 2.25 M. Those findings shifted the conceptual framework towards the biological implications of the potential synergy of intravirion-specific conditions/components required for efficiently triggering polymerization^{(259,} ²⁸⁷⁾. In this line of reasoning, during assembly, millimolar concentrations of CA proteins are expected to be exposed to the highly polyanionic surface of the two viral RNA molecules, envelope constituents, and other compounds. Furthermore, self-assembly should occur in a relatively low-salt/neutral pH environment derived from the cytosol, and at the host normal body temperature.

Despite noteworthy advances in the structural biology of retroviral CA proteins and their assemblies, many questions remain unresolved. The assembly pathway has not been defined, as the inherent complexity of the reaction has thwarted to unravel the mechanistic details. Also lacking is a

comprehensive understanding of how physicochemical factors dictate the mechanism by which CA building blocks are assembled together, as well as modulating the preference for hexamer/pentamer incorporation. On the other hand, definition of the role of the intravirion conditions, still poorly characterized, is of crucial relevance for understanding retroviral capsid formation.

In this work we report the characterization of the *in vitro* self-assembly properties of the mature form of Bovine Leukemia Virus capsid protein, in an effort to identify factors/conditions required for triggering assembly, with particular emphasis in near-physiological ones. Here we describe: i) the analysis of CA_{BLV} solution behaviour aimed at determine the basal (unassembled) state of the protein, ii) a systematic study of the *in vitro* assembly reaction as a function of different physicochemical parameters, and iii) the analysis by electron microscopy of the ultrastructural morphology of *in vitro* assembly products. Particularly, we found that CA_{BLV} has the capacity to spontaneously self-assemble under the sole requirement of the physiological physicochemical conditions (temperature, pH, and ionic strength) provided by the host cell. Our results provide insights towards understanding BLV capsid assembly, information that can contribute to other retroviruses, particularly HTLV. Additionally, they pose CA_{BLV} as an interesting model for testing capsid assembly under the influence of intravirion compounds and/or conditions.

Results

 CA_{BLV} is monomeric in solution. As no previous studies have yet been described in literature concerning the BLV capsid protein, we first devised and optimized a strategy to obtain the recombinant mature form of BLV capsid protein, via expression of the full-length native sequence with no tag in *E. coli*. Our two-step protocol involving anion-exchange chromatography followed by gel permeation (see materials and methods) yielded large amounts of recombinant CA_{BLV} protein (~100 mg/L), of purity higher than 99% as determined by densitometry of Coomassie-stained SDS-PAGE gels (Fig. 1 left).

We first address the elucidation of the nature of the smallest association state of CA_{BLV} . The purified protein in Storage Buffer (20 mM HEPES [pH 7.4], 50 mM NaCl) passed through the column as a

single peak, eluting at the same position over the range of concentrations used (0.25-1 mM), with no evidence of self-association (Fig.1 center). Comparison with standards of known molecular weight (MW) and size (Stokes radius, R_H) yielded an apparent mass of 34 kDa, and an estimated R_H of 2.5 nm, consistent with the expected size for the CA_{BLV} monomer (Fig. 1 center, top panel).

The hydrodynamic behaviour was also characterized by dynamic light scattering (DLS). DLS analysis showed that the protein remains monodisperse in solution, with a main population centered at a hydrodynamic radius (R_H) of 2.6 nm, from which a MW of 32 kDa can be estimated (Fig. 1 right, bottom panel), in excellent agreement with the results obtained using small-zone SEC. The apparent MW obtained by DLS and SEC is higher (about 33-40% respectively) than expected for the monomer. Since both techniques yield a hydrodynamic dimension, the larger MW suggest that CA_{BLV} posses an anisotropic shape. Overall, the results obtained by analytical SEC and DLS are self-consistent, and established that CA_{BLV} protein exists as a monomer in solution at concentrations up to at least 1mM. The higher than expected hydrodynamic size is also compatible with the fact that the protein consist of two individual NTD and CTD domains, which are flexibly tethered.

To further characterize the solution behaviour of the unassembled state, size exclusion chromatography coupled to multi-angle laser light scattering (SEC-MALLS) was used to obtain a shape-independent determination of the mass of CA_{BLV} . MALLS analysis unambiguously confirmed the monomeric state of the protein up to 1mM, with a Mw = 24,2 ± 0.1 kDa (Fig. 2).

 CA_{BLV} self-assembly is a strongly temperature-dependent process, compatible with a nucleationdriven mechanism. Retroviral CA proteins are generally induced to assemble under different nonphysiological conditions, typically involving high ionic strength (1M NaCl or greater), at near-neutral or mildly acidic pH, ^(108, 279, 285, 295). Thus, we choose to initiate the study of CA_{BLV} assembly properties in a solution buffered at the physiological-like pH = 7.4 (storage buffer: 20mM HEPES pH7.4, 50mM NaCl). During initial experiments, it became apparent to us that CA_{BLV} polymerization depends greatly on temperature. To investigate the role of temperature and further other parameters on the self-association process, we used a turbidimetric assay following light scattering at 340 nm. Turbidimetry has proved to be a very useful methodology to characterize the factors influencing triggering and kinetics of retroviral capsid assembly. The effect of temperature on CA_{BLV} assembly is shown in Fig. 3a. The protein do not exhibit noticeable polymerization while kept at low temperature (shown in the graph is the turbidimetric data at 4°C), and SEC analysis indicated it remains in its unassembled monomeric state (not shown). When exposed to higher temperatures (range 4 to 38°C), the protein spontaneously self-assemble, with kinetic parameters being a function of the temperature. Specifically, both the assembly rate (turbidity slope) and the extent of polymerization (steady state value) increase concomitantly with temperature. Notably, the temperature-dependence analysis showed that the efficiency of the reaction reaches an optimal regime at the bovine host physiological temperature (38°C), exhibiting the higher assembly rate and steady state value. Above 38°C (40-42°C), polymerization rate starts to decrease, with higher temperatures leading to protein denaturation (data not show).

In vitro studies with other retroviral CA proteins indicated a rate-limited mechanism for assembly, involving nucleation intermediates. For a self-assembly process to be categorized as a nucleated cooperative polymerization, a necessary condition is the requirement of a minimal protein concentration (Cc, critical concentration) to proceed⁽³²⁰⁾. Thus, we evaluated the assembly rate of CA_{BLV} as a function of protein concentration at 38°C (Fig. 3b). The rate of assembly notoriously increases with increasing concentration, and the resulting maximum turbidity (DO at plateau) is proportional to the initial protein concentration, from which a Cc of approximately 160 μ M can be estimated. Altogether, the assembly characteristics of CA_{BLV} -specifically: sigmoidal kinetics with a noticeable lag time and the requirement of a Cc to proceed-suggests that polymerization takes place via a reversible nucleation-driven mechanism.

We also wondered if any species of discrete size -possibly constituting nucleating intermediatescould be detected. Thus, we followed CA_{BLV} polymerization by DLS, as this technique is highly sensitive to the presence of even small amounts of oligomeric species, and using a protein concentration (200µM) giving a sufficiently slow polymerization rate (data not shown). Only a bimodal distribution of species with physical volume/dimensions matching unassembled monomeric protein co-existing with large size assemblies ($R_{\rm H} > 100$ nm, compatible with the presence of small tubes) of increasing sizes was observed, and no discrete oligomers were detected. This suggests that intermediate (nucleating) species most probably exist in very low amounts and/or are short-lived, so their formation/consumption is exceedingly fast to be detected by DLS.

Modulation of CA_{BLV} assembly by solution variables: effect of ionic strength, pH, and salts. We next examine the effect of physicochemical parameters on CA_{BLV} self-assembly. We firstly considered the fact that retroviral CA proteins studied so far, require a high ionic strength (given by monovalent salts, typically NaCl) to initiate polymerization. Hence, we wanted to determine what role solution ionic strength play on the polymerization of CA_{BLV} . The results obtained showed that increasing the solution ionic strength from 50-200 mM NaCl at pH 7.4 notoriously diminishes the tendency of the protein to self-associate (Fig 3c). Specifically, higher ionic strengths concomitantly increases the lag time, diminishes the assembly rate, and shorten the extent of reaction, as evidenced by the progressive diminishing of the maximum turbidity. Monovalent salt concentrations above 200 mM NaCl almost completely arrest assembly.

We then evaluated the results of inducing the assembly reaction at different solution pHs while keeping constant both ionic strength and protein concentration (Fig. 3d). The assembly of CA_{BLV} remain arrested in mildly-acidic solution (pH 4.0 - 5.0) but drastically increases in the range pH 6.0 – 8.0. At more basic pHs (above pH 8.0 – 10.0) the assembly rate notoriously diminishes. CA_{BLV} assembly in a low salt buffer exhibits an optimum in the physiological range of pH 7.0 – 8.0.

One very interesting fact was that the thermal-induced polymerization (range 4-38°C) was found to be reversible upon cooling, with preformed assemblies completely dissociating to monomers at 4°C (as evaluated by SEC and DLS, data not shown). This process is also reproducible, as switching back the same CA_{BLV} solution to higher temperatures, gave turbidity profiles with superimposable curves. Considering the biologically relevant conditions of the experimental setup, we also wanted to determine the assembly rate of CA_{BLV} at a concentration more similar to that probably existing inside the retroviral particle, where is expected to reach higher concentrations than those used in *in vitro* studies. The actual stoichiometry of CA (or Gag) in BLV retroviral particles is unknown, but it has been reported that in HIV-1 virions (average ~145nm diameter) the core is formed from 1,000-1,500 molecules of CA ⁽⁵⁸⁾. Assuming a similar number of CA molecules for a BLV viral particle (diameter ranging between 60-120 nm^(4, 59)), we take as an approximation a $[CA_{BLV}] = 2$ mM. At this high concentration, CA_{BLV} self-assembly after a shorter lag, leading to a rapid rise in turbidity, and completely polymerizing in approximately 15 minutes (Fig 3e).

Analysis of *in vitro* assembly products by negative stain MET shows that CA_{BLV} polymerizes under close to physiological conditions forming planar 2D sheets, contrasting the behaviour to HIV-1 CA which forms exclusively nanotute (Fig. 4). Those 2D sheets are compatible with the sole presence of hexamers, forming a planar structure via lateral CTD-CTD contact between hexameric capsomers, as mentioned in the Introduction. As clearly obvsereved in the micrographs, these 2D lattices are notoriously flexible, which seems compatible with the flexibility required to form authentic capsids.

We next performed a turbidimetry-based screening for salts potentially affecting (promoting/inhibiting/modulating) the assembly of CA_{BLV}, using a high ionic strength (1.5 M salt), as other retroviral CA proteins are induced to multimerize in that condition. We also were intrigued by the way that ions can modulate the reaction, thus we used salts with differently-charged ions at an equivalent final ionic strength. The rationale is that charge-screening effects on assembly should be a function of solution ionic strength, irrespective of ions net charges. On the contrary, if ion-specific effects exist, the relative contribution of salts should depend of ion identity. As expected from the previously observed ionic strength dependence of the reaction, the presence of high monovalent salt concentrations (both Na⁺/K⁺ salts, above 250 mM) completely abolished the reaction (Fig 5). Interestingly, when added at a concentration adjusted to give the same ionic strength, divalent anion salts (PO_4^{2-} and SO_4^{2-}) were found to be remarkably more effective at promoting polymerization than monovalent salts. Moreover, self-assembly promoted in the presence of dianions was more effective than temperature-induced assembly, as the reaction requires a lower protein concentration to proceed at a lower temperature. Dianion-triggered assembly reaction also exhibits a sigmoidal kinetics, with a lag phase and a turbidity plateau.

Phosphate-driven self-assembly of CA_{BLV}. The above results, together with the fact that divalent anions – most notoriously, phosphate dianion - were shown to promote the assembly of HIV-1 and RSV capsid proteins *in vitro*, prompted us to evaluate the role of phosphate on CA_{BLV} assembly in more detail. CA_{BLV} polymerizes very rapidly in the presence of phosphate, and the effect is markedly concentration-dependent as increasing phosphate concentration at pH 7.4 (from 325 to 500 mM) drastically increases assembly rate, maximum DO reached, and decreases the lag phase (Fig. 6a). In the presence of a constant amount of phosphate at pH 7.4, increasing protein concentration caused a progressive shortening of lag time and accelerates polymerization (Fig. 6b). Thus, both phosphate and protein concentration are synergistic in promoting assembly.

In a later experiment, we checked if the assembly-promoting capacity of phosphate dianions can compete with the strong inhibitory effect of monovalent ions present in solution as background salt. Thus we started assembly by jump-dilution, adding simultaneously both NaPO₄ and NaCl, helding constant both phosphate and protein concentrations. As seen in Fig. 6c, even in the presence of a strongly-promoting concentration of phosphate, self-association is negatively influenced by increasing monovalent salt ionic strength. Thus increasing sodium chloride concentration inhibits polymerization, counteracting the effect of the dianion.

We also assessed the impact of pH on phosphate-induced assembly. As expected considering the promoting effect of phosphate dianion, assembly rate increases when the pH of NaPO₄ solution becomes more basic, as the quotient $HPO_4^{2^-}/H_2PO_4^{-1}$ logarithmically augments with pH (Fig. 6d). In order to discriminate the specific effect of pH on dianion-induced assembly, we performed the experiment in solution of different pHs and using $SO_4^{2^-}$ (Na₂SO₄) as a trigger. The rational is that the pKa for protonation of $SO_4^{2^-}$ is approximately 1.9, making the fully deprotonated dianion the predominant specie over the pH range used. Thus, protein dissolved in buffer at different pH was induced to assembly by addition of 500 mM sodium sulphate. The turbidimetric profile of sulphate-triggered assembly show a clear dependence of pH, exhibiting a maximum polymerization rate at pH 7.0 and decreasing towards both acidic and basic pH

values (Fig. 6e). This indicates that SO_4^{2-} strongly promotes assembly, with the polymerization efficiency roughly following the pH-dependence observed in storage bufferin the absence of the divalent anion (shown in Fig. 3d).

We finally analyzed by TEM the *in vitro* assembly products induced by phosphate and observed that CA_{BLV} polymerized into roughly spherical, rounded structures, with a diameter of around 100nm, roughly compatible with an average BLV native capsid Thus, phosphate modifies CA_{BLV} assembly habit from 2D sheets (possessing only hexamers) towards spherical capsid-like structures (possessing hexamers as well as pentamers, which are needed for closing the lattice into a spherical structure.

Macromolecular crowding or RNA are essential for triggering capsid self-assembly in a cytosolic-like medium and regulates assembly morphology

Our interest in the physiological relevance of the impact of physicochemical conditions governing CA_{BLV} assembly lead us to the to test if typical intra-virion conditions can be playing a role in triggering and/or modultating the reaction in the authentic particle. We first analysed the capacity of the protein to polimerize in a buffer most closely mimicking the ionic composition of the cytosol. So we tested the assembly reaction in CLM (cytosolic like medium [140 mM KCl, 20 mM NaCl, 2 mM MgCl₂, 1 mM EGTA, 20 mM Pipes, pH7.0], which is extensively use in literature to emulate intracellular conditions. Surprisingly, assembly was completely arrested, with no noticeable polimerization even at the favourable conditions used (Fig 7 a, none). This finding strongly suggested to us that it will be indeed some intravirion factor/conditions that support capsid assembly in the particle. One remarkable feature of cells is the very high concentrations of macromolecules occurring *in vivo*, accounting as much as 50-400mg/mL, which leads to a 20-40% of the total volume being occupied by macromolecules. The same holds of the interior of viruses, which are as much macromolecularly crowded as the cytosol of the cells they infect. Dextran70, Ficoll70, and PEG are extensively used as model crowding agents in macromolecular crowding studies as diverse as protein folding, stability, binding, and polymerisation. Those crowder agents are extensively used to mimic *in vitro* the cell-like crowded environment generated in vivo from a

heterogeneous mixture of macromolecules. So, we tested is the addition of different crowding agents are capable to stimulate assembly. To our surprise, we notice that all crowders indeed bypass the inhibitory effect of the CLM buffer, which we most logically assumed is caused by its ionic strength (Fig 7 a). But mimicking crowding not only allows self-assembly under otherwise arresting conditions, but it also alters the assembly behaviour towards formation of spherical particle nicely resemblim authentic capsids (Fig. 7 b, and gallery below). Under crowding regime (for all tested crowders), the reaction kinetics exhibits its prototypical behaviour, with a signmoidal progress curve and a concentration dependent lag phase (Fig. 8). The assembly lag time decreases concomitantly with increasing concentration of crowder agent.

Finally, encouraged by these findinds we tested if RNA can also exhibit an effect similar to that of crowders. Indeed, RNA is present at a very high concentration in the viral particle (\geq 18 mg/mL) so it is logically to consider that it also behaves as a crowder agent. On the other hand, the net negative charge exposed by the two nucleic acid strands could also promote assembly in a way similar to that shown by dianions. Interestingly, when CABLV was incubated at a concentration similar to that found inside the viral particle (1 mM, which indeed can be and underestimated value), along with RNA matching its intravirion concentration, we notice that again assembly was potenciated. Indeed, after a excessively long reaction considering the high CA concentration, the protein remained arrested in CLM alone, but completely assembled in one hour in the presence of RNA.

Discusion

In this study, we present a systematic analysis of the *in vitro* self-assembly properties of the capsid protein of Bovine Leukemia Virus. To our knowledge, this is the first work addressing the influence of different physicochemical parameters on the self-association capacity of the capsid protein of a Deltaretrovirus, with a particular emphasis on physiological-like conditions.

In order to establish the unassembled solution state of the protein, we used two hydrodynamic biophysical tools, small-zone SEC and DLS. Both analyses confirm that CA_{BLV} exists as a stable monomer in a solution buffered at the physiological-like pH 7.4 and low ionic strength. The hydrodynamic behaviour

of CA_{BLV} is highly anisotropic, suggesting that the protein has very likely an elongated shape, which is consistent with the prototypical fold of other retroviral CA proteins, including CA_{HTLV-1}. Moreover, as was also shown for CA_{HTLV-1}⁽²⁷⁸⁾, CA_{BLV} remains monomeric at concentrations up to 1mM, with no evidence of monomer-dimer equilibrium⁽²⁷⁸⁾. This basal oligomeric state is differs with that of lentiviral CA proteins, which exist as dimers in the low micromolar range. Thus, CA_{BLV} belongs to the group of Orthoretroviral capsid proteins exhibiting a monomeric basal –unassembled- state, along with CA proteins of HTLV-1⁽²⁷⁸⁾, $RSV^{(286, 287)}$, $EIAV^{(314, 315)}$, and $JSRV^{(316)}$. Our data clearly show that CA_{BLV} is able to self-assemble *in vitro* under near-physiological temperature, pH, and ionic strength conditions, at relatively low protein concentrations. Indeed, the assembly rate is strongly favoured at protein concentrations in the millimolar range. Recent determination from microscopy studies have indicated that the concentration of free CA inside the retroviral particle is in excess of about 3 mM. Assuming a similar concentration for the BLV viral particle, it is very likely that BLV can polymerize rapidly once free after Gag processing. This finding was remarkable and surprising as other retroviral CA protein studied so far require non-physiological solution conditions to multimerize, such as high ionic strength and/or mildly acidic pH. This fact is even more relevant considering that the intrinsic tendency of CA_{BLV} to self-associate shows an optimal in the physiological range of solution conditions tested. The kinetics of assembly follows a sigmoidal behaviour, with a clear lag phase, and reaches a steady-state. The assembly rate was highly concentration-dependent and clearly posses a minimum critical concentration (Cc) to initiate polymerization. Taken together, these characteristics are typical of a well-behaved rate-limiting step in assembly and suggest that CA_{BLV} assemble through a nucleation-driven pathway. Evidence for a nucleation-dependent self-association has been also found for CA proteins of RSV and HIV-1 in *in vitro* assays^(287, 295, 321). The observed behaviour of CA_{BLV} contribute further support to the notion that retroviruses of different genera employ common principles to form their capsids via a de novo nucleation-limited mechanism. Further studies are needed for elucidating the nature of the nucleation complex.

A salient feature of CA_{BLV} assembly process is its strong temperature-dependence. Indeed, while maintained in a neutral-pH low-salt buffer, the protein can be induced to polymerize by simple rising

temperature towards the bovine's normal temperature (38°C). On the other hand, the assemblies formed in vitro completely dissociate into monomers when the solution is cooled down. Thus, temperature is a main factor controlling the equilibrium between monomeric building blocks and higher-order assemblies. Moreover, the fast and quantitative reversibility of the equilibrium suggest that interactions governing CA_{BLV} self-association must involve individually weak association energies. Many biological selfassembling systems are greatly enhanced by temperature, generally as a consequence of hydrophobic, hydrophilic⁽³²²⁾, or hydrogen-bonding mechanisms. Unfortunately, a molecular explanation of the origin of temperature effect on CA_{BLV} self-association will require a structural understanding of the protein interfaces involved, actually unknown. In any case, based on the marked temperature sensitivity of CA_{BLV} assembly it is tempting to speculate the existence of a high kinetic barrier for CA_{BLV} polymerization, which disfavours the initial nucleation step. Temperature will thus positively affect the overall kinetics/affinity of the reaction simply by favouring subunit encounter/docking. For CA_{HIV-1} it has also been noted that incubation of the protein at 37°C increase the assembly efficiency, although experiments were performed using 1M NaCl as the triggering stimulus, making difficult to realize a direct parallelism with CA_{BLV}^{(259,} ³²³⁻³²⁵⁾. However, it is possible that temperature could promote surpassing an energetic barrier of retroviral CA assembly, more likely to nucleation. In any case, it is interesting that CA_{BLV} assembles efficiently under near-physiological pH and ionic strength, exhibiting an optimum at the host body temperature and diminishing towards lower or higher temperatures.

The marked salt-mediated effects on CA_{BLV} assembly suggest that electrostatic interactions may also play a major role in modulating protein-protein interactions. Structural and mutagenesis studies with HIV and RSV CA proteins have revealed that electrostatic repulsions must be overcome for selfassociation to occur, which help explaining why both proteins are effectively induced to polymerize by high-salt solutions (\geq 1M NaCl). In the case of CA_{BLV} , this electrostatic control appears to be more complex than "simple" charge screening mediated by salt ions. On one hand, increasing ionic strength given by monovalents salts (both Na⁺/K⁺ with Cl⁻ as counterion) have a marked tendency to inhibit polymerization. Beyond the physiological-like I = 140 mM at pH = 7.4, polymerization is sharply abolished, and concentrations in the molar range results in complete quenching of CA_{BLV} assembly. This behaviour is quite different to that of other retroviral CA proteins studied so far. We speculate that this negative effect of monovalent salts could be attributed to screening/shielding of charges on the CA_{BLV} surface resulting in a reduction of electrostatic attractions required for protein assembly.

On the other hand, and in principle counter intuitively, the presence of divalent anions at high concentrations, particularly phosphate, has a remarkably opposed outcome, strongly inducing protein selfassociation. Phosphate-triggered assembly of CA_{BLV} also follows a sigmoidal multiphasic reaction with lag, growth, and stationary phases, again suggesting the requirement of nucleating species. The promoting effect of phosphate (as well as other dianions) has also been observed for the retroviral CA proteins of RSV and HIV^(287, 288). In the case of CA_{RSV} protein, both divalent and monovalent salts -separately or combinedcan effectively trigger assembly, and it has been proposed that both type of ions can shield charge repulsions, even synergistically. In this case, the contribution of charge screening and shielding to assembly promotion will be determined by the total ionic strength, irrespective of the ions charge numbers. However, the fact that under equivalent ionic strength the capacity to promote CA_{BLV} assembly is determined by the nature of salt ions (specifically, dianions stimulate assembly while monoions block the reaction), suggest the existence of dianion-specific effects on CA_{BLV} self-assembly. Our interpretation of this result is that the bivalent negatively-charged geometry of dianions rather than solely charge could possibly act by "bridging" interacting protein interfaces. This argument helps conciliating the opposite effect of monovalent anions, which posses a marked capacity to inhibit polymerization, even in the presence of high concentration of phosphate. It is also in agreement with an electrostatic control of CA_{BLV} polymerization, as suggested by the sharp effect of pH on the reaction in all conditions tested.

Hexamers appears to be the most favoured array of CA proteins when assemble *in vitro*, while introduction of pentamers is less preferred, although their incorporation is an unavoidable requisite for closing the capsid. The fact that CA_{BLV} assembles into tubular structures in a typical buffer solution, suggest that some kind of regulation provided by intra-virion environment is missing when polymerizing free in solution. In this line of reasoning an interesting result is that phosphate dianions modify the higher-

order assembly habit of CA_{BLV} from a tubular organization to roughly spherical structures that resemble native capsids. This remarkable efficiency to stimulate assembly is similar to the behaviour described by Purdy *et al* for $CA_{RSV}^{(287)}$. Although the effect of phosphate ions is not completely understood in structural terms, cryo-electron microscopic analysis of CA_{RSV} *in vitro* assemblies suggest that interaction of CA-CA interfaces involve overcoming electrostatic repulsions, which are stronger in pentameric associations, supporting the hypothesis of neutralization of opposing charges⁽²⁷⁷⁾. Despite no structural information is yet available for CA_{BLV} , a similar explanation could be envisaged, as it share with CA_{RSV} the tendency to rapidly polymerize in the presence of divalent ions at neutral/near-neutral pH, whereas monovalent salts alone failed to trigger assembly ⁽²⁷⁷⁾. Another key issue is that although phosphate potentiates assembly of BLV, RSV, and HIV-1 CA proteins, only CABLV and CA_{RSV} modify their assembly habit towards capsidlike structures, whereas CA_{HIV-1} invariably forms tubes. This reinforces the notion that different CA proteins can respond differentially the same triggering stimulus.

The strong stimulating potential of dianions on RSV and HIV-1 assembly has been hypothesized to mimic the effect of polyanions inside the viral particle, such as negatively-charged lipid on the viral membrane and/or the two genomic RNA molecules. Overall, the ability of phosphate dianions in accelerating assembly suggests that their effect *in vitro* may related to the conditions inside retroviral particle, where CA protein is in close proximity to the phosphate backbone of the RNA strands. It likely that the phosphate groups of the backbone provides a highly-charged negative surface which potentiates initiation of assembly, as observed in solution with dianions. The functional importance of CA-phosphate interactions is suggested by the notorious enhancement of the assembly rate induced when an oligonucleotide is added to the assembly reaction. The biological significance of the present results can be supported considering the conditions of the cellular and viral particle context. In this scenario, it is tempting to speculate that the phosphate groups of the phosphodiester backbone of retroviral RNA may contribute to self-assembly triggering as well as redirecting capsid assembly from cylinders to capsid-like spherical structures. The data presented here support this proposal *in vitro*, on the view that RNA exhibit the remarkable capability of bypassing the inhibitory effect of CLM under a near-physiological

concentration regime. Macromolecular crowding per se is also sufficient to promote assembly as well as modulating the assembly behaviour towards capsid-like structures. It will be interesting to corroborate if RNA can also show this effect. Indeed is tempting to speculate that it shall be so, considering that phosphate dianionsgenerates closed particles, but it is still speculative.

In conclusion, the results obtained in this work have important biological implications, as they demonstrated that CA_{BLV} protein is able *per se* to self-assemble under physiological-like solution conditions into functionally relevant higher-order structures. It is possible that some viral particle components such as genomic RNA, can possibly act as facilitators of assembly. In any case, our results imply that for CA_{BLV} protein the basic process of assembly can occur spontaneously *in vitro* under physiological-like conditions, and quite possibly *"in vivo"* as well. Considered together, the results presented here allow proposing a scenario in which CA_{BLV} self-assembly can spontaneously occur solely under the physicochemical conditions imposed by the host environment.

Figures



Figure 1- Production of recombinant CA_{BLV}. (left) SDS-PAGE (15% polyacrylamide) gel analysis of different stages of the purification of CA_{BLV}. TNI: total non-induced extract; TI, total induced extract; Q-sepharose: eluate: enriched CA_{BLV} remains in the flowthrough of anionic exchange; SEC-purified CA: pooled fractions eluted from preparative SEC, containing finally polished recombinant protein; M: molecular mass marker. (center) Definition of the solution state of CA_{BLV}. (center) small-zone SEC analysis of CA_{BLV} solution behaviour. The elution profiles corresponding to different concentrations (in mM) of the protein are shown. The positions of molecular weight standards (numbers corresponding to molecular mass in kDa) and column total volume are indicated with short lines. (right, top) plot of square root of –[log partition coeficient (Kav)] as a function of Stokes radius for SEC protein standards. The estimated Stokes radius for CA_{BLV} (2.6 nm, monomer) is indicated with an arrow. (right, bottom) Dynamic light-scattering analysis of CA_{BLV} in solution at 37 μ M. Shown is the size distribution expressed by % Intensity versus the hydrodynamic radius (R_H, in nanometers). The peak is centered at a R_H = 2.6 nm.



Figure 2 – Size-exclusion chromatography-MALLS analysis of CA_{BLV}. Elution profile

followed by absorbance @280nm (left y axis) on Superdex 200 10/300 of increasing concentrations of CA_{BLV} (250, 500, 750, 1000µM) in storage buffer (20mM HEPES, 50mM NaCl, pH 7.4). Superimposed is shown the absolute molecular weight distribution (right y axis) determined from normalized light scattering @ 90degrees. The analysis was performed at 20°C.



Figure 3 – Analysis of higher-order self-assembly of CA_{BLV}. (a) Effect of temperature on self-assembly kinetics of CA_{BLV} at 330 μ M in storage buffer (SB, 20mM HEPES, 50mM NaCl, pH 7.4). (b) Concentration dependence of CA_{BLV} assembly kinetics in the 125-370 μ M range in SB. (c) Effect of ionic strength on assembly kinetics. CA_{BLV} was kept at a concentration of 330 μ M in storage buffer, and the final ionic strength (in mM) was adjusted with concentrated NaCl (in SB). (d) Influence of pH on assembly kinetics. CA_{BLV} at 370 μ M was assembled in solutions buffered at different pHs (detailed in M&M) (e) CA_{BLV} at 2000 μ M assembled in SB. All experiments were performed a 38°C.


Figure 4. Electron micrographs of CA_{BLV} assembly. Shown is a gallery of negatively stained

electron micrographs of CA_{BLV} (500uM) assembled in SB at 38°C.



Figure 5 – Effect of different ions on higher-order self-assembly of CA_{BLV}. Turbidimetric

profile of CA_{BLV} at 100 μ M induced to assemble at 30°C in SB by addition of varying concentration of distinct salts, at an adjusted concentration giving a final Ionic Strength of 1,5M. Reactions were performed in SB at 30°C.



Figure 6 – Characterization of CA_{BLV} self-assembly induced by phosphate. (a) Effect of

phosphate concentration. Turbidimetric profile of CA_{BLV} at 100 µM induced to assemble at 30°C in SB by addition of varying concentration of sodium phosphate at pH 7.4. (b) Different concentrations of CA_{BLV} (in the range 55-100 µM) were induced to assemble at 30°C in SB by addition of 500 mM sodium phosphate at pH 7.4. (c) Effect of increasing ionic strength (given by addition of NaCl, with the final concentration indicated in mM) on the assembly rate at 30°C in SB of 80 µM CA_{BLV} induced by addition of 500 mM sodium phosphate at pH 7.4. (d) Assembly at 30°C of 70 µM CA_{BLV} triggered by addition of 500 mM sodium phosphate at different pH values at 30°C in SB (ver M&M). (e) Effect of pH on sulphate-triggered assembly. Solutions of 80 µM CA_{BLV} at different pH values were induced to assemble at 30°C with 500 mM Na_2SO_4 . (f) Electron microscopy analysis of the morphology of CA_{BLV} assemblies formed *in vitro* in the 50 mM sodium phosphate at pH 7.4.



Figure 7. Effect of macromolecular crowding on CA_{BLV} self-assembly reaction. (a) Comparison of the effect of different crowders (at the same concentration of 15 % w/v) on the self-assembly of CA at 500µM in CLM. All reactions were performed at 38°C. (b) Assembly products of CA from (a) in F70 (left) and D70 (right) visualized by negative stained TEM. Belown is shown a gallery of micrographs obtained from assembly products generated in the presence of F70 (upper row) and D70 (lower row).



Figure 8. Effect of macromolecular crowding on CA_{BLV} self-assembly reaction. (a-f) Assembly of CA at a concentration of 400 μ M in CLM buffer, in the presence of varying concentrations (indicated in % weight/volume) of different macromolecular crowding agents; (a) BSA; (b) Ficoll-70; (c) Dextran-70; (d) PEG1K; (e) BSA. None: CLM without any added crowder. All reactions were performed at 38°C.



Figure 9. Effect of RNA on CA_{BLV} self-assembly reaction. Assembly of CA at 1000 μ M in

CLM without (no RNA) or with (+RNA) added (18mg/mLt RNA)to the assembly reaction. All

Methods

Expression and purification of recombinant CA_{BLV}.

cDNA coding the full-length native sequence of Bovine Leukemia Virus capsid protein was obtained from genomic DNA of the cell line FLK (fetal lamb kidney), persistently infected with BLV (FLK-BLV). The amplified DNA was subcloned into the pET16a expression vector (Novagen) in frame with the initial methione derived from the vector, but without any added tag, and the construct was verified by DNA sequencing. CA_{BLV} protein was expressed in *Escherichia coli* strain BL21(DE3)pLysS cells induced with 1 mM isopropyl-β-D-thiogalactopyranoside at optical density 0.7 and further incubated at 30°C for 4 hours. Harvested cells were resuspended in 20 mM Tris [pH6.0], 50mM NaCl, supplemented with complete EDTA-free protease inhibitor (Roche), lysed via DNase/RNase treatment followed by sonication, and then centrifuged at 20,000g for 1 hr. The supernatant was loaded onto a Q-Sepharose Fast Flow column (Amersham) equilibrated with 20mM Tris [pH6.0], and the flow-through containing the enriched protein was recovered. Final polishing of the protein was performed by gel filtration chromatography using a Superdex 75 column (Amersham) equilibrated in standard Storage Buffer [SB, 20 mM HEPES pH 7.4, 50 mM NaCl]. Purified CA_{BLV} protein was concentrated to a final stock concentration of *circa* 20 mg/mL, supplemented with 0.02% NaN₃. Protein homogeneity was determined to be >99% as evaluated by SDS-PAGE and densitometric analysis. Verification that purified protein posses the authentic mature sequence was done by MALDI-TOF mass spectrometric analysis as described below. Purified protein preparations were kept until analysis at 4°C, remaining stable for several weeks. Before experiments were performed, protein quality was checked by dynamic light scattering and analytical size exclusion chromatography. Protein concentration was determined by measuring the absorbance at 280 nm and using calculated values of the extinction coefficients.

Mass spectrometry analysis.

Mass spectrometry measurements were performed in a 4800 MALDI TOF-TOF Analyser (Applied Biosystems). For molecular mass determination CA_{BLV}, the recombinant protein was desalted with a C4 reverse-phase micro-column (OMIX Pippete tips, Varian) and eluted directly on a sinapinic acid matrix (10mg/mL in acetonitrile-H₂O 60%, 0.1% trifluoroacetic acid). The protein was analyzed using positive linear mode, and the spectra was calibrated with Mix3 (Applied Biosystems) as an external standard. For peptide mass fingerprinting of CA_{BLV}, digestion was carried out by trypsin treatment (Sequencing-grade Promega) overnight at 37°C. Peptides were extracted from the gels using 60% acetonitrile in 0.2% TFA, concentrated by vacuum drying and desalted using C18 reverse phase micro-columns (OMIX Pippete tips, Varian). Peptide elution from micro-column was performed directly into the mass spectrometer sample plate with 3µL of matrix solution (α -cyano-4-hydroxycinnamic acid in 60% aqueous macetonitrile containing 0.2% TFA). Mass spectra of digestion mixtures were acquired in a 4800 MALDI-TOF/TOF instrument (Applied Biosystems) in reflector mode and were externally calibrated using a mixture of peptide standards (Applied Biosystems). Collision-induced dissociation MS/MS experiments of selected peptides were performed. Proteins were identified by NCBInr database searching with peptide m/z values using the MASCOT program and using the following search parameters: monoisotopic mass tolerance, 0.05 Da; fragment mass tolerance, 0.2 Da; methionine oxidation, as possible modifications and one missed tryptic cleavage allowed. Digestion and MS/MS analysis confirmed removal of the vector-derived initial methione, and the presence of proline in the first N-terminal position, as occurs in the authentic mature capsid sequence.

Absolute Mass determination by SEC-couple static light scattering

All SEC-MALLS analysis were performed at 20°C on a Superdex 200 10/300 GL analytical grade column (GE Healthcare Bioscience) using an Äkta purifier coupled to a DAWN[®] HELEOS[®] multi-angle light scattering detector, and Optilab[®] T-rEXTM on-line differential refractometer. 50µL of purified CA_{BLV} at different concentrations were loaded in the SEC column pre-equilibrated with Storage Buffer and run at 0.4 mL/min.

Small-zone analytical size exclusion chromatography.

Small-zone-SEC was performed at 20°C in a Superdex 75 10/300 column (GE Healthcare) equilibrated in Storage Buffer using an ÅKTA Purifier FPLC System (GE Healthcare). A volume of 100 µL of protein sample at different concentrations ranging from 6 mg/mL (\cong 0.25 mM) to 24 mg/mL (\equiv 1 mM) were injected onto the column with a 500 µL loop and eluted with Storage Buffer at a flow rate of 0.5 ml/min. Elution was followed spectrophotometrically at a wavelength of 280 nm. Column was calibrated with protein standards of known molecular weight and (Mw) Stokes radius (R₈) (low molecular weight gel filtration kit, GE Healthcare-Amersham): Bovine serum albumin (67kDa/34.8nm), Ovalbumin (43kDa/30.5nm), Chymotrypsinogen A (25kDa/20.9nm), Ribonuclease A (13.7/16.4nm). Column void (Vo) and total (Vt) volume were determined using Blue dextran 2000 and NaCl respectively. Apparent molecular weight (Mr) calculation was performed by plotting the partition coefficient (Kav) *versus* log(Mr). Kav was calculated from the formula $K_{av} = (V_e V_0)/(V_t V_0)$, where Ve is the elution volume, V₀ is the void volume, and V_t is the total volume. From this calibration graph, the molecular weights of the eluted proteins were calculated, and a plot of $\sqrt{-\log(K_{av})}$ *versus* Rs was used to estimate the Stokes radius of CA_{BLV} protein.

Dynamic and static light scattering experiments.

Size measurement by DLS of CA_{BLV} protein dissolved in Storage Buffer were performed at a concentration of 37 μ M. Intensities of scattered light and correlation times were measured using a Zetasizer NanoS (Malvern, UK) in batch mode in a 50 μ L cell, maintained at 18 °C. Diffusion light scattering fluctuations of the scattering intensity due to Brownian motion were recorded and an Autocorrelation function was derived which allows calculation of diffusion coefficient. Assuming proteins as spheres, hydrodynamic radius was estimated from diffusion coefficient using Einstein-Stokes equation.

Kinetic *in vitro* **assembly assays.** For any experiment, stocks of CA_{BLV} solution at an appropriate concentration were held on ice until dilution. Preliminary experiments were conducted to establish the most appropriate wavelength, which was chosen as $\lambda = 340$ nm. At this λ , a linear turbidity (optical density, DO)

working range from 0 to 2,5 optical units is achieved for the experimental conditions described. Reactions were performed in UV-transparent 96-well microplates (Corning) using a final volume of 100 μ L. Readings were obtained using a Multiskan FC microplate reader (Thermo Scientific) provided with a temperature-controlled incubator. During measurements, reactions were mixed by intermittent shaking between readings to avoid disturbances due to precipitation of assembly products. The critical concentration (Cc) was obtained from plot of the steady-state values of turbidity (DO at plateau) as function of total protein concentration. The value for critical concentration was determined by extrapolation of plateau value at DO = 0, and the corresponding intercept in the concentration coordinate was taken as the Cc⁽³²⁰⁾.

Cytosol-like medium was prepared as previously described (CLM: 140 mM KCl, 20 mM NaCl, 2 mM MgCl2, 1 mM EGTA, 20 mM PIPES, pH 7.0)^(326, 327).

For evaluating the effect of pH on assembly, CA_{BLV} buffer-exchanged by dialysis against the appropriate buffer (20 mM citric acid pH 4.0; 20 mM MES pH 5.0; 20 mM MES pH 6.0; 20 mM HEPES pH 7.0; 20 mM HEPES pH 8.0; 20 mM Tris pH 9.0; 20 mM CHES pH 10.0) containing 50 mM NaCl, to minimize mismatch errors. For evaluating the effect of differently-charged salts, final salt concentrations were adjusted to give comparable ionic strengths, calculated as $I = \frac{1}{2}(\Sigma c_i z_i^2)$, where c_i is the molar concentration of ion i and z_i is the charge number of ion i. Assembly was induced by jump-dilution by addition of salt to give the appropriate final concentration. In the case of phosphate-triggered reactions, CA_{BLV} in Storage Buffer was induced to assembly by jump-dilution, adding the appropriate volume of sodium phosphate. The effect of ionic strength given by monovalent salt on phosphate-triggered reaction was evaluated by addition of a concentrated NaCl solution in Storage Buffer to the assembly reaction. Solutions of sodium sulphate at different pHs were prepared using the buffers described above. Inositol phosphate (Sigma) and oligo-dT₁₂ (IDT) were prepared as concentrated stock solutions in Storage Buffer.

Electron microscopy analysis of assembly products.

In vitro self-assembly products of recombinant purified CA protein were taken from polymerization reaction with a pipette. Sampleswere then adsorbed ontoairglow discharged carbon-coated grids and stained with uranil acetate to the sample pH (7.0-8.0). Images were recorded in an electron microscope (model CM12; Philips) operated at 80kV, at different nominal magnifications.

Competing interests

The authors declare no competing financial interests.

Authors' contributions

GO conceived the study, designed and performed the experiments, analyzed the data, written and edited the manuscript. JL design and performed all electron microscopy studies, analyzed data, and contributes critically in drafting the manuscript. FC contributed to protein expression optimization and performed recombinant protein purification. FC, LT, GM, GR, and SB contribute reagents/tools and helped to draft the manuscript. OP conceived the study, designed the experiments, written and edited the manuscript. All authors read and approved the final manuscript

Acknowledgements

We are particularly grateful with Dr. Félix Rey (Unité de Virologie Structurale) for helpful discussions and critical reading of the manuscript and also for facilitating the use of the infrastructure of the his lab. This work made use of the Plate-forme de Microscopie Ultrastructurale (Imagopole, Insitut Pasteur) were the EM analisis were performed. Our gratitude is extended to the Madelón Portela, Carlos, Batthyany, and Rosario Durán (Unidad de Bioquímica y Proteómica Analítica, Institut Pasteur Montevideo) for mass spectrometry analysis. Funding for this research was provided by CNRS grant

References

1.Rodriguez, S.M., et al.: Preventive and therapeutic strategies for bovine leukemia virus: lessons for HTLV. Viruses. 3(7): p. 1210-48.

2.Gillet, N., et al.: Mechanisms of leukemogenesis induced by bovine leukemia virus: prospects for novel antiretroviral therapies in human. Retrovirology, 2007. 4: p. 18. 3.Rodriguez, S.M., et al.: Preventive and therapeutic strategies for bovine leukemia virus: lessons for HTLV. Viruses, 2011. 3(7): p. 1210-48.

4.Muriaux, D., J.L. Darlix, and A. Cimarelli: Targeting the assembly of the human immunodeficiency virus type I. Curr Pharm Des, 2004. 10(30): p. 3725-39.

5.Neira, J.L.: The capsid protein of human immunodeficiency virus: designing inhibitors of capsid assembly. FEBS J, 2009. 276(21): p. 6110-7.

6.Neira, J.L.: The capsid protein of human immunodeficiency virus: molecular recognition and design of antiviral agents. FEBS J, 2009. 276(21): p. 6097.

7.Vogt, V.M., Retroviral virions and genomes, in Retroviruses, J.M. Coffin, Hughes, H. E., Varmus, H. E., Editor. 1997, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York. p. 27-70.

8.Ganser-Pornillos, B.K., M. Yeager, and W.I. Sundquist: The structural biology of HIV assembly. Curr Opin Struct Biol, 2008. 18(2): p. 203-17.

9.Gross, I., H. Hohenberg, and H.G. Krausslich: In vitro assembly properties of purified bacterially expressed capsid proteins of human immunodeficiency virus. Eur J Biochem, 1997. 249(2): p. 592-600.

10.Li, S., et al.: Image reconstructions of helical assemblies of the HIV-1 CA protein. Nature, 2000. 407(6802): p. 409-13.

11.Jin, Z., et al.: Model for lentivirus capsid core assembly based on crystal dimers of EIAV p26. J Mol Biol, 1999.286(1): p. 83-93.

12.Birkett, A.J., et al.: Cloning, expression, purification, and characterization of the major core protein (p26) from equine infectious anemia virus. Biochim Biophys Acta, 1997. 1339(1): p. 62-72.

13.Mortuza, G.B., et al.: Structure of the capsid amino-terminal domain from the betaretrovirus, Jaagsiekte sheep retrovirus. J Mol Biol, 2009. 386(4): p. 1179-92.

14.Briggs, J.A., et al.: The stoichiometry of Gag protein in HIV-1. Nat Struct Mol Biol, 2004. 11(7): p. 672-5.

15.Cardone, G., et al.: Visualization of a missing link in retrovirus capsid assembly. Nature, 2009. 457: p. 694-699.

16.Hyun, J.K., et al.: Proton-driven assembly of the Rous Sarcoma virus capsid protein results in the formation of icosahedral particles. J Biol Chem. 285(20): p. 15056-64.

17.Jones, I.M. and Y. Morikawa: The molecular basis of HIV capsid assembly. Reviews in medical virology, 1998.8: p. 87-95.

18.Lanman, J., et al.: Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange. Nat Struct Mol Biol, 2004. 11(7): p. 676-7.

19. Yeager, M.: Design of in Vitro Symmetric Complexes and Analysis by Hybrid Methods Reveal Mechanisms of HIV Capsid Assembly. J Mol Biol, 2011. 410(4): p. 534-52.

20.Weiland, F. and S. Ueberschar: Ultrastructural comparison of bovine leukemia virus (BLV) with C-type particles of other species. Arch Virol, 1976. 52(1-2): p. 187-90.

21.Purdy, J.G., et al.: Retroviral capsid assembly: a role for the CA dimer in initiation. Journal of Molecular Biology, 2009. 386: p. 300-314.

22.Lanman, J., et al.: Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry. J Mol Biol, 2003. 325(4): p. 759-72.

23.Purdy, J.G., et al.: Critical role of conserved hydrophobic residues within the major homology region in mature retroviral capsid assembly. J Virol, 2008. 82(12): p. 5951-61.

24.Lanman, J., et al.: Kinetic analysis of the role of intersubunit interactions in human immunodeficiency virus type 1 capsid protein assembly in vitro. J Virol, 2002. 76(14): p. 6900-8.

25.Ehrlich, L.S., et al.: Assembly of recombinant human immunodeficiency virus type 1 capsidprotein in vitro. Journal of Virology, 1992. 66: p. 4874-4883.

26.Barklis, E., et al.: Characterization of the in vitro HIV-1 capsid assembly pathway. Journal of Molecular Biology, 2009. 387: p. 375-389.

27.Andreu, J.M. and S.N. Timasheff: The measurement of cooperative protein self-assembly by turbidity and other techniques. Methods Enzymol, 1986. 130: p. 47-59.

28.von Schwedler, U.K., et al.: Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. EMBO J, 1998. 17(6): p. 1555-68.

29.del Alamo, M., G. Rivas, and M.G. Mateu: Effect of macromolecular crowding agents on human immunodeficiency virus type 1 capsid protein assembly in vitro. J Virol, 2005. 79(22): p. 14271-81.

30.Barrera, F.N., et al.: Envelope lipids regulate the in vitro assembly of the HIV-1 capsid. Biophys J, 2008. 94(2): p. L8-10.

31.Ghysdael, J., R. Kettmann, and A. Burny: Translation of bovine leukemia virus virion RNAs in heterologous protein-synthesizing systems. J Virol, 1979. 29(3): p. 1087-98.

85

32.Khorasanizadeh, S., R. Campos-Olivas, and M.F. Summers: Solution structure of the capsid protein from the human T-cell leukemia virus type-I. J Mol Biol, 1999. 291(2): p. 491-505.

33.Kingston, R.L., et al.: Structure and self-association of the Rous sarcoma virus capsid protein. Structure, 2000. 8: p. 617-628.

34.Leikin, S., D.C. Rau, and P. V.A: Temperature-favoured assembly of collagen is driven by hydrophilic not hydrophobic interactions. Nat Struct Biol, 1995. 2(3): p. 205-10.

35.Alfadhli, A., et al.: Analysis of human immunodeficiency virus type 1 Gag dimerization-induced assembly. J Virol, 2005. 79(23): p. 14498-506.

36.Huseby, D., et al.: Assembly of human immunodeficiency virus precursor gag proteins. J Biol Chem, 2005. 280(18): p. 17664-70.

37.Morikawa, Y., T. Goto, and F. Momose: Human immunodeficiency virus type 1 Gag assenbly through assembly intermediates. Journal of Biological Chemistry, 2009. 279: p. 31964-31972.

38.Swatton, J.E. and C.W. Taylor: Fast biphasic regulation of type 3 inositol trisphosphate receptors by cytosolic calcium. J Biol Chem, 2002. 277(20): p. 17571-9.

39.Rossi, A.M. and C.W. Taylor: Analysis of protein-ligand interactions by fluorescence polarization. Nat Protoc. 6(3): p. 365-87.

CAPÍTULO 3

Estructura de CA_{BLV}

Previa al capítulo:

"Crystal Structure of Native Retroviral Hexameric Lattice Reveals High Conformational Plasticity"

Se resume (es español) el contexto del trabajo, los principales hallazgos, y sus implicancias.

En este se capítulo se describe los resultados del eje estructural de esta Tesis, en el cual se hizo foco en la caracterización de la estructura cristalina de la proteína CA de BLV. Este trabajo está actualmente culminado y los resultados fueron capitalizados en un manuscrito que será enviado para considerar su publicación en una revista internacional con referato. Concretamente, el manuscrito se titula "Crystal Structure of Native Retroviral Hexameric Lattice Reveals High Conformational Plasticity", Gonzalo Obal, Felipe Trajtenberg, Federico Carrión, Lorena Tomé, Nicole Larrieux, Otto Pritsch, Alejandro Buschiazzo; será enviado a la revista Science. Al comenzar, enfrentamos una situación similar a la descrita para el auto-ensamblado de CABLV: ninguna información previa disponible. En este sentido, este eje era motivador ya que de tener éxito, podríamos aportar los primeros datos de la estructura 3D de una proteína relevante de un virus con importancia sanitaria en nuestro País. Reconocimos otro problema inmediatamente. La literatura sí era muy amplia para otros retrovirus, como siempre con HIV-1 a la vanguardia debido al enorme interés en el desarrollar drogas contra nuevos blancos, y dado el papel clave de la cápside, ésta había atraído un enorme interés en los últimos años. Una gran cantidad de estructuras a alta resolución, determinadas mediante cristalografía de rayos X y/o NMR habían aportado una visión global de la estructura 3D de los dominios individuales NTD y CTD de la CA de varios retrovirus, abarcando diferentes géneros. La figura de la CA como una proteína con dos dominios plegados independientemente, básicamente α -helicoidales, unidos por un *linker* flexible, estaba bien establecida. De los antecedentes específicos surgía un concepto ya ampliamente aceptado para ese momento: la imposibilidad de cristalizar la proteína completa nativa, debido a su gran flexibilidad intrínseca y su tendencia a oligomerizar en productos muy heterogéneos. De hecho, varios grupos importantes habían llegado a esta conclusión luego de más de diez años de intentos.

Nosotros decidimos en esta Tesis comenzar con la estrategia más lógica, producir e intentar la cristalización de los dominios individuales de la CABLV. Luego de una larga serie de intentos durante más de dos años, logramos obtener cristales útiles para difractar, y finalmente pudimos resolver la estructura 3D (a resolución atómica) de los dos dominios individuales de la CA de BLV (se detalla en la Metodología y Resultados). Estos resultados nos permitieron visualizar por primera vez la estructura 3D de la CA del BLV, la cual de hecho mostró las características propias de las CA retrovirales. Infortunadamente, ninguno cristalizó formando un oligómero biológicamente relevante. El NTD cristalizó como monómero y el CTD como un dímero no relevante, por lo cual no fue inicialmente posible extraer información sobre los contactos proteína-proteína potencialmente involucrados en el ensamblado. En paralelo, tuvimos éxito en obtener cristales y posteriormente resolver la estructura 3D de una versión del NTD incapaz de formar el β-harpin N-terminal, debido a la presencia de un residuo de leucina en el extremo amino. Este residuo extra (de hecho, el último residuo del domino MA de la Gag de BLV, deja a la prolina en posición 2, por lo cual es incapaz de formar un puente salino con el aspartato conservado, y como resultado los primeros 12 residuos (correspondientes al β -harpin) permanecen desplegados. Esta estructura resultó virtualmente idéntica a la del NTD nativo (esto es, con β -harpin), diferenciándose únicamente en que la ausencia de un β -harpin plegado. Notablemente, la hélice $\alpha 1$ se observó completamente plegada, de idéntica forma que el NTD nativo. Esto fue un hallazgo interesante, ya que se había descrito que la ausencia del β-harpin (situación análoga a la de CA inmadura, formando parte de Gag) determinaba que el desplegado de la α1 clave para la oligomerización. Sin embargo nuestros estudios aportaron la primer evidencia estructural a alta resolución de que la estructura terciaria del core de α -hélices el NTD permanece intacto.

Como se describió en el capítulo anterior, la CA de BLV tiende a formar mallas 2D. Este hecho, junto con el abundante conocimiento que acumulamos sobre las propiedades del auto-ensamblado de la proteína, nos impulsó a intentar cristalizar la proteína nativa completa. Asimismo, contábamos con la información estructural de los dominios independientes que resolvimos previamente (lo cual de hecho, como se describe en la Metodología y Resultados, resultó de crucial importancia). Nos enfocamos entonces en intentar cristalizar la proteína CA completa nativa. En paralelo, uno de los principales grupos en biología estructural resolvió (mediante una serie de ingeniosas mutaciones y *cross-linkings* de cisteínas) la estructura de los capsómeros hexamérico y pentamérico de la CA de HIV-1. Sin embargo, quedaban aún muchas e importantes preguntas. Por un lado, las cisteínas destinadas a estabilizar los capsómeros mediante cross-*linking* entre los 6 o 5 protómeros respectivamente, podrían haber afectado la flexibilidad intrínseca del bloque de construcción. Por otro, y muy particularmente, se introdujeron mutaciones en el CTD con el objetivo de inhibir la polimerización, por lo cual aún restaba definir cuáles eran los contactos nativos que median la asociación entre capsómeros y por tanto el auto-ensamblado. Esto es de fundamental relevancia para comprender cabalmente los determinantes moleculares de la formación de la cápside.

Luego de varios meses de intentos, conseguimos cristalizar la proteína completa nativa, lo cual fue un hito importante. Los siguientes dos años nos enfocamos en optimizar estos cristales hasta que logramos obtener un primer cristal que si bien difractó, la resolución obtenida (>4Å) era muy baja para visualizar detalles atómicos. Sin embargo, nos permitió identificar en el empaquetamiento del cristal "capas " o mallas 2D compuestas de una unidad

asimétrica hexamérica (ver Figs. 1 y 2). Considerando que la CA_{BLV} ensambla formando ´´2D sheets´´, planas, las cuales sabemos por datos en otras CA retrovirales que constan únicamente de hexámeros, estos resultados nos sugirieron la posibilidad de que el cristal consista de planos o capas apiladas de hexámeros. De ser así, cada capa de hexámeros sería una representación fiel de la lattice hexamérica madura de CA. Durante el último año y medio de esta Tesis, invertimos gran parte del tiempo en optimizar estos cristales, hasta que finalmente logramos obtener cristales (de hecho, 1!) que difractó adecuadamente, y posteriormente (usando la información estructural de los dominios independientes) pudimos resolver la estructura 3D de CA a resolución atómica (2.75 Å) (ver Figs. 1 y 2).

En resumen, en esta parte del trabajo de Tesis mostramos los resultados que nos permitieron obtener las estructuras cristalinas a resolución atómica de los dominios individuales de la CA de BLV, incluyendo una versión del NTD incapaz de formar el β -harpin. Esto nos permitió definir por primera vez la estructura de los dominios de una CA deltaretroviral, sino que nos permitió determinar que la estructura terciaria del dominio es idéntica en su versión inmadura como madura, a excepción del β -harpin (ver Fig. Sup.2). Por otro lado, y de hecho el resultado más relevante, tuvimos la suerte de lograr elucidar la estructura de la CA_{BLV} intacta (lo cual no se consideraba posible), formando un ensamble biológicamente relevante: el capsómero hexamérico, el cual de hecho es el bloque de construcción principal de la cápside retroviral. Encontramos en el cristal que los hexámeros se mantienen unidos a sus vecinos a través de contactos laterales vía sus CTDs, formando una malla bidimensional hexagonal, la cual es completamente representativa de la lattice auténtica (ver Figs. S4 y S5). Así, la estructura constituye una fuente valiosa ya que permite visualizar los contactos nativos que mantienen unida la *lattice*. Esta es la primera vez que se logra visualizar a resolución atómica, la malla bidimensional de capsómeros hexaméricos de la proteína CA retroviral nativa, intacta. Consideramos que este logro es un avance mayor dentro la retrovirología en general, más allá de BLV.

Índice de Figuras Capítulo 3

Figure 1. Structure of the BLV hexameric	97
Figure 2. Crystal structure of BLV CA	98
Figure 3. Deviation from strict 6-fold symmetry within the CA hexameric capsomer	99
Figure 4. Close-up of the two independent trimeric interfaces stabilizing the	
BLV CA lattice, as observed in the crystal asymmetric unit	100
Table S1. Data collection and refinement statistics	106
Figure S1. Crystal structures of BLV CA and derived sub-domains	109
Figure S2. BLV CA compared with other retroviruses	111
Figure S3. BLV CA compared with HTLV-1	112
Figure S4. Crystal packing of BLV CA	113
Figure S5. Deviation from strict 6-fold symmetry within and among CA hexameric capsomers	114
Figure S6. Quasi-2-fold axes with CTD-CTD inter-hexamer contact interfaces	115

Crystal Structure of Native Retroviral Hexameric Lattice Reveals High Conformational Plasticity

("draft" del artículo a ser enviado Retrovirology)

G. Obal^{1,2†}, F. Trajtenberg^{3†}, F. Carrión¹, L. Tomé^{1,2}, N. Larrieux³ O. Pritsch^{1,2*}, A. Buschiazzo^{3,*}

¹ Institut Pasteur de Montevideo, Unit of Protein Biophysics, Mataojo 2020, 11400, Montevideo, Uruguay.
² Dep de Immunobiología, Universidad de la República, Av. General Flores 2125, 11800, Montevideo, Uruguay.

³Institut Pasteur de Montevideo, Unit of Protein Crystallography, Mataojo 2020, 11400, Montevideo, Uruguay.

Institut Pasteur, Dept of Structural Biology & Chemistry, 25, rue du Dr Roux, 75015, Paris, France.

*Correspondence to: pritsch@pasteur.edu.uy, alebus@pasteur.edu.uy.

†These authors made equal contributions to this work.

Abstract:

In retrovirus, capsid assembly is essential for infectivity, as its diverse functions rely on its exceptional fullerene-like structure. Despite key significance, we lack atomic-level structural information of capsomers assembled in a native hexagonal packing environment. Here, we report the 2.75 Å crystal structure of the native capsid protein of oncoretrovirus Bovine Leukemia Virus, forming a two-dimensional hexameric array. The planar lattice show tightly packed hexamersfrom strict 6-fold symmetry. Both inner NTD and outer CTD substantially flexible, allowing for slight warping of capsomers. CTD quasi-2-fold and quasi-3-fold interfaces, variable declinations between surrounding. This firstview of anintact mature latticeathigh-resolution the plasticity of the building blockand adjustment of inter-capsomer interactions. As fullerene architecture is conserved among retrovirus, these insights may well be generalized.

Main Text:

Infectious an electron-dense "core" particle contained within a lipid envelope. This core assembles after proteolytic maturation of the Gag precursor polyprotein, which drives budding of immature particles out of the plasma membrane of infected cells^(328, 329). The viral protease (PR) becomes active within the budded particle, cleaving Gag into the mature matrix (MA), capsid (CA) and nucleocapsid (NC) proteins. Gag proteolysis is essential for the infectivity of all retroviruses, leads to disassembly of the Gag latticevirionsCA assembly into the core particle, enclos the genomic RNA in a ribonucleoprotein (RNP) complex with NC together with the reverse transciptase (RT) and integrase (INT) enzymes. must disassemble to allow transport of the resulting proviral DNA as a pre-integration complex across the nuclear pores for integration into the host genome. The core particle is thus an essential structure in the retrovirus cycle, and unveiling its intrinsic molecular features is key for a full understanding of retrovirus biology.

Retrovirus cores are quite pleomorphic, displaying a spherical, conical or tubular aspect, depending on the retrovirus⁽²⁹³⁾. Moreover, core particles of different sizes are observed in the same preparation of a given retrovirus ⁽³⁰¹⁾, complicating their structural characterization by traditional methods. Nevertheless, and in spite of poor sequence conservation of the CA building block, they are believed to display the same basic fullerene array of hexameric CA capsomers with 12 interspersed CA pentamers to allow particle closure^(108, 258). Furthermore, CA displays the same tertiary structure in all retroviruses, composed of two independently domains, N-terminal and C-terminal (NTD and CTD), connected by a flexible linker⁽³²⁹⁾. The crystal structures of the isolated domains have been obtained for a number of retroviruses^(88, 275, 330, 331), but the structure of the intact CA molecule has been challenging because of its intrinsic flexibility^(278, 302, 309, 332).

A major landmark in understanding the organization of retrovirus core particles has been the identification of cysteine mutations in the NTD of HIV-1 CA that led to disulfide stabilization of either hexamers or pentamers. These cysteine mutants allowed crystallization of the full-length CA, providing high-resolution snapshots of the organization of the two types of capsomers ^(302, 333). A caveat was that, in order to obtain these crystals, the authors had to introduce further mutations (W184A and M185A in HIV-1 CA) to maintain the capsomers in solution by destabilizing lateral contacts between CTD domains, which normally lead to lattice formation. In spite of mutations, the results obtained from these studies, in combination with tubular assemblies of native HIV-1 CA at sub-nanometer resolution ⁽³³⁴⁾ have provided very important insights into the architecture of the core particle. In particular, they confirmed the organization of a central 5-fold or 6-fold symmetric and rigid ring of NTDs (depending on the capsomer), surrounded by an outer "belt" of relatively mobile CTDs. The CTD pack against the NTD of the adjacent

subunit in the ring, so that the oligomer displays essentially no intra-protomer NTD-CTD contacts other than the covalent link. CA assembly is formed by CTD-CTD dimeric contacts between hexamers, whereas additional trimeric interfaces are to play key atubular ^(15, 302, 334)Working with CA from the bovine leukemia virus (BLV) – a retrovirus very close to the human T-cell leukemia virus in the delta-retrovirus genus^(4, 312)we identified conditions in which the native CA crystallizes with an hexamer in the asymmetric unitA. Although these crystals were twinned and diffracted to about only 2.7 Å resolution, the crystallization in parallel of the individual BLV NTD and CTD provided high resolution diffraction data (1.7Å, see Table and f), which guided the refinement of the intact BLV CA crystal structure to 2.75Å resolution in spite of the twinning. (335)ed The crystal structure of the isolated domains NTD and CTD of BLV are the most similar to those of HTLV, with root mean square deviation (rmsd) of 1.9Å(for 98 Ca atoms) and 1.2Å (for 69 Ca atoms), respectively. The NTD of BLV and HTLV CA lacks helix a6 (Fig. S2), and has a shorter $\alpha 4-\alpha 5$ loop(Fig. 1B), lacking the cyclophilin binding site that is present in HIV-1 CA. To facilitate the comparative analysis, we use the HIV numbering for the secondary structure elements (SSE). Superposition of BLV and HIVNTDs results in 2.9Å rmsd for 60 equivalent Ca atoms. The CTDs are closer structurally, with an overall rmsd of 1.3Å for 38 Ca atoms. Concerning the tertiary organization of BLV CA, the linker between NTD and CTD is longer, along with 7, with respect to the HIV counterpart.asymmetricA^(15, 88, 302, 336)neighboringing.However, the first 4 residues are clearly seen in all protomers showing the key DThe same motif is well defined density and confidently modeled in , adopting a similar orientation to the ones observed in hexamers, the 3^(88, 270, 337)We also solved the structure of a NTD variant (no-beta) having the Pro1-Asp53 salt bridge disrupted by inserting one additional N-terminal residue. Both NTD structures are almost identical (rmsd 0.6Å for 106 Ca atoms), including helix $\alpha 1$, indicating that proteolytic β-hairpin refolding does not affect its overall structure.very The CA crystals belong to the hexagonal space group P6₅, and the CA hexamer in the asymmetric unit displays pseudo C_6 symmetry, organized as the HIV-1 capsomers in the tubular structures described previously^(15, 334). It makes planar layers that are related by the 65 screw axis to form the crystal, with the quasi 6-fold (Q6) molecular axis shifted with respect to the crystal axes, and tilted by about 7 degrees (24). Comparison with the closest hexamer obtained by applying strict 6-fold symmetry to a CA protomer and preserving the center of mass yields a rmsd of 2.8 Å (for 1136 Cα superposed atoms, fig. 3). Moreover, if the rings of NTDs and CTDs are symmetrized separately, they give rmsd's of 2.4 Å and 1.9 Å, for respectively 732 and 420 Ca atoms, relative to their counterparts in the asymmetric hexamer. The deviation from C_6 symmetry reflects a highly plastic CA hexamer, and gives rise to quasi-equivalent protein contact interfaces, both within and among neighboring capsomers (fig. S5A). The loose NTD-CTD connection within each protomer (fig. S5B), as well as a variable pattern of NTD-NTD contacts within the capsomer (Fig. 3), explain the rupture of symmetry. The description is made easier by considering the deviation with respect to a symmetric hexagonal lattice, in which the capsomers would be related laterally by strict 2-fold and 3-fold axes that are 93

parallel to the central molecular 6-fold, which would coincide with the crystallographic axis (in such a case, the space group would be P6 with a c axis of about a sixth of the c axis of the P6₅ space group; see Table).

In the BLV CA lattice, the deviation from C₆ symmetry of the tilted CA hexamers gives rise to three independent quasi-2-fold (Q2) and two quasi 3-fold (Q3) lateral contacts2A. Remarkably, the CTD "dimers" related by the 3 independent Q2 axes are very similar to each other, resulting in a CTD dimer that packing of helices 9S6Trple at an angle of roughly 45 degrees2, as observed in the crystal structures of the isolated HIV-1 CTD dimer. Trp133/Ile168 ressembles the role of the Trp184/Met185 pair of HIV-1 (Fig. S5), suggesting a functional conservation of this bulky hydrophobic residue. In contrast, the two independent Q3 contacts are quite different to each other4, although they bring into interaction the same surfaces observed in the HIV CA tubes ⁽³³⁴⁾. Dimeric and trimeric CTD interfaces leads to the snugly fit of the BLV CA hexamers observed in the crystal packing (Fig. 2A). By packing at different heights with respect to the quasi-3-fold axes (Fig. 4), these helices eventually rescue the overall planar 2D lattice while maintaining a tiled packing.

Working now with a non-engineered protein allowed us to uncover the natural plasticity of a retroviral capsid assembly unit. The asymmetric CA hexamer provide a snapshot of the of the flexible capsomer and the type of quasi-equivalent 2-fold and 3-fold contacts that lead, in this particular case, to a planar lattice of pseudo C_6 symmetry. The captured structure predicts that, when assembled in solution, the capsomer adopt multiple isoenergetic conformations. The NTD ring now appears as a key source of flexibility, contrasting the rigid NTD core paradigm. This global plasticity have not been seen in structures of cross-linked HIV-1 CA hexamer⁽³⁰²⁾, raising the question of whether disulphide stabilization of its NTD ring could have imposed constraints that hindered genuine capsomer flexibility. Comparative structural studies will be valuable for identifying similar plasticity to BLV CA in other retroviral CA proteins.

The hexagonal lattice of authentic capsids is and variably curved, so capsomers must adapt to different, with local breakings from perfect C_6 symmetry. The plasticity observed in our structure provides aninsight of how the lattice can adapt to different packing environments. In the crystal, the whole capsomer is seen twisted and warped in response to assembly; both properties suggest a mean for distorting a perfectly flat lattice. Interestingly, this deformation adopted to fit in the plane, the Q2 contacts, suggesting a main role in maintaining the assembly scaffold. The also that CTD-CTD trimer interfaces are key for stabilization and adaptability of planar hexameric sheet. Therefore, the sole presence of 3-fold contacts is not a mandatory determinant of curvature, as suggested by previous results ⁽³³⁴⁾. Q3 interfaces are instead remarkably more malleable and alter the declination between contacting hexamers, providing clues on how to curve the lattice. Supporting our observation, from the recent HIV-1 tubular assembly atomic model it is

possible to predict a up to 6 different Q3 axis are required, depending on the curvature degree, for cylindrical assembly. Hexamer distortion as well as communication between, through interacting surfaces, offers a network for efficient long-range propagation of rearrangements. Our up an important gap in our structural understanding of capsid assembly, with important consequences for guiding inhibitor development.



Figure 1.Structure of the BLV CA hexameric capsomer **(A)**Two orthogonal views of the asymmetric unit of BLV CA.The capsomer is depicted as cylindrical cartoon and smoothed surface, where each protomer is shown in different colors. NTD ring is outlined for clarity and in the bottom view NTD and CTD layers are highlighted. The molecular dimensions of the capsomer are indicated.**(B)** Cartoon representation of the BLV protein, coloured with a blue-to-red ramp (N- to C-terminal sense), highlighting the difference in secondary structure elements with respect to HIV-1 CA, in pale blue beneath.





Figure 2. Crystal structure of BLV CA **(A)** Solvent accessible surface of seven interacting hexameric capsomers. The central capsomer is highlighted with darker tones, and the two colours distinguish NTD (green) and CTD (blue). Zoom-in views show details of inter-hexameric interfaces at selected quasi-3-fold (left inset) and quasi-2-fold (right inset) axes. In the insets, BLV protomers are coloured (selected interacting residues showed in sticks), with HIV-1 CA (pdb 3H47) superposed in grey. Trimer interface (left inset): dashed arrows highlight the larger separation observed in the HIV-1 structure, explaining the absence of trimer interactions. Dimer interface (right inset): the dashed arrow highlights the rotational shift between BLV and HIV-1. The α 8- α 9 loop and the first turn of α 9 are absent in the HIV-1 structure (dotted line). **(B)** Tiled organisation of successive capsomers in the crystal. The 2D lattice plane is horizontal in this view. The quasi-6-fold axis is shown as a black stick on the left-most hexamer. The dashed lines (perpendicular to the quasi-6-fold axes) connect identical positions of different protomers within the hexamers, highlighting the ~10% vertical shift in certain adjacent interfaces.





Figure 3. Deviation from strict 6-fold symmetry within the CA hexameric capsomer. Cartoon representation showing two perpendicular views of the NTD ring (the CTD has been omitted). Helices are shown as cylinders for greater clarity. In pink, a strict 6-fold symmetric hexamer of NTDs was built with the NTD from protomer A. The exact 6-fold axis is shown as a stick with the same color in the center of the hexamer. The NTD of protomer A was separately superposed successively onto the positions of the other five protomers, according to the crystal structure, resulting in the blue hexamer (this procedure eliminates contribution of coordinate deviations coming from flexible, independently refined protomers, so that rmsd figures only reflect the deviation from strict C_6 point symmetry). Rms pairwise deviations result in diverse figures, ranging from 2.1 Å to 3.5 Å (except for protomer A, which expectedly gives 0 Å rmsd). Note that. The asymmetry of the NTD hexamer ring is also highlighted by drawing the rotation axes that relate each pair of protomer positions. Shown as thinner blue sticks toward the center of the capsomer, they show crossing angles 10 degrees. of up to

Figure 4.



Fig. 4. Close-up of the two independent trimeric interfaces stabilizing the BLV CA lattice, as observed in the crystal asymmetric unit. **(A)** Two orthogonal orientation views of the interface between protomers B/D/F of three neighbor hexamers. **(B)** The A/C/E interface is shown. Different colors distinguish the chains. On the left panels, side-views with respect to the quasi-3-fold axes (depicted as purple sticks) allow to highlight the positions of R188 C α atoms (dashed lines), revealing variable height shifts among the interacting protomers. R188 and I184 (on the same side of α 10) are shown as sticks for reference. On the right, solvent accessible surface representations show the close contacts among the three interacting protomers.

References and Notes:

Acknowledgments:

The structures are available in the PDB under accession codes 4PH0, 4PH1, 4PH2 and 4PH3. We acknowledge funding from CNRS (France, program LIA 316), ANII (Uruguay), CeBEM (www.cebem.org.ar) and FOCEM (MERCOSUR Structural Convergence Fund, COF 03/11). We thank Andrey Lebedev for useful programs to analyze NCS, and Beatriz Guimaraes at beamline Proxima 1 (synchrotron Soleil) for assistance with full-length CA data collection.

Supplementary Materials:

Materials and Methods

Table S1

Figures S1-S5

Supplementary Materials:

Materials and Methods:

Cloning, expression, and purification of proteins

BLV capsid coding sequences were obtained *via* restriction-free cloning⁽³³⁸⁾ from proviral DNA of the cell line FLK-BLV (GenBank: EF600696.1|:956-1600).

Full-length CA from BLV (encoding residues Pro 1 to Leu 215) was expressed from pET22b expression vector (Novagen) containing no extra (non-native) residues except for the vector-derived N-terminal methionine. BLV CA was expressed in BL21(DE3)pLysS *E. coli* cells (Stratagene) after induction with 1 mM IPTG for 4 hours at 30 °C. Harvested cells were resuspended in 20 mM Tris.HCl pH6.0, 50 mM NaCl, 0.1 mM DNAse, protease inhibitor cocktail (complete EDTA-free, Roche) and then lysed with sonication. The soluble fraction was obtained by centrifugation at 20,000 g for 1 h, and loaded onto a DEAE-Sepharose Fast Flow column (Amersham) equilibrated with 50 mM Tris.HCl pH 6.0. The flow-through was subjected to size exclusion chromatography using a Superdex 75 column (GE Healthcare) equilibrated in storage buffer (20 mM HEPES pH 7.4, 50 mM NaCl). Purified BLV CA protein was concentrated by ultrafiltration to 20 mg/mL and stored for further usage. The removal of the initial methionine was verified by mass spectrometry after in-gel digestion, with a 4800 MALDI TOF/TOF instrument (Applied Biosystems).

The N-terminal domain (NTD) of BLV capsid possessing the mature N-terminal β -hairpin, consists of residues Pro1-Pro125 followed by a carboxyl-terminal hexa-histidine tag. This construct was expressed with a pET28a (Novagen) plasmid in BL21(DE3)pLysS *E. coli* cells. Cells were grown and induced as described above for full-length CA. Harvested cells were resuspended in 20 mM sodium phosphate pH 7.4, 20 mM imidazole, 500 mM NaCl. Cell lysis was achieved by sonication in the presence of 0.1 mg/mL DNAse and protease inhibitors (complete EDTA-free, Roche), and centrifuged at 20,000 g for 1 h. Purification was performed in two steps, with a HisTrap Chelating (Ni2+) HP column (GE Healthcare), followed by size exclusion chromatography with a Superdex 75 column (GE Healthcare), equilibrated in storage buffer, concentrated by ultrafiltration to 20 mg/mL and stored at 4 °C.

The C-terminal domain (CTD) construct, comprising residues Trp125-Leu215, was cloned into pET22b with an N-terminal His6 tag and expressed in BL21(DE3)pLysS *E. coli* cells. Cells were grown, induced, and lysed as described above for the NTD. The supernatant was loaded onto a HiTrap Chelating (Ni2+) HP column (GE Healthcare) pre-equilibrated with binding buffer (50 mM sodium phosphate pH 7.4, 500 mM NaCl), cleaved within the column overnight at 4 °C by addition of thrombin, and then eluted isocratically with binding buffer. Thrombin was removed with a HiTrap Benzamidine FF column (GE Healthcare) connected in series. The protein was finally purified by size-exclusion chromatography on a Superdex 75 column in storage buffer, concentrated by ultrafiltration to 40 mg/mL and kept at 4 °C.

To generate the NTD in which the β -hairpin structure is unfolded, different constructs comprising residues Pro1-Pro125 of BLV CA were designed, possessing different N-terminal extensions at residue

Pro1 derived from the matrix-capsid (MA-CA) junction of the immature Gag polyprotein sequence (including the last 1-4 amino acids of MA). The constructs were cloned into pET28a with an N-terminal His6 tag and a thrombin cleavage site for expression in *E. coli* and further purification, as described above for the CTD. Serendipitously, we identified that one such construct, containing an extra leucine residue (corresponding to the last C-terminal residue of MA) immediately upstream of Pro1, yields a protein that suffers spontaneous cleavage at the vector-derived sequence remaining after thrombin treatment (GSHMAS \downarrow L, arrow indicating the proteolytic site). This cleavage occurs within a few hours after purification and the resulting protein (residues MA[Leu110]-CA[Pro1-Pro125]) remains stable during weeks, as confirmed by mass spectrometry of in-gel digestion analyses. Consequently, we used this cleaved protein for crystallization after a final size exclusion chromatography step (Superdex 75 [GE]) in storage buffer. The purified protein was concentrated to 50 mg/mL in storage buffer and kept at 4 °C.

Protein Crystallization

Proteins were crystallized using the hanging-drop vapour-diffusion method at 20 °C. Crystals of native full-length BLV CA were grown by mixing 3 μ L of protein at 10 mg/mL in storage buffer, with an equal volume of crystallization solution (20 mM Tris pH 7.0, 130 mM NaCl, 25% w/v PEG 4000). The protein formed plate-like crystals of varying size and thickness appearing in 1-2 days, typically growing to full size in approximately 1 week. Cryoprotection was achieved in two steps: i) reservoir solution was replaced with mother liquor containing 40% w/v PEG 4000, and 1 μ L of this solution was carefully added into the drop containing the crystals, and incubated overnight at 20 °C; ii) selected crystals were then taken out from the drop and kept for a few seconds in 40% w/v PEG 4000 crystallization solution, and finally frozen in liquid nitrogen for mounting and data collection.

Crystals of NTD (mature form with β -hairpin) at 30 mg/mL in storage buffer were grown in 0.2 M ammonium acetate, 1.0 M ammonium sulphate (2 μ L each). The protein formed hexagonal crystals that grow to full size within a week. Crystals were then harvested and cryoprotected by soaking directly in mother liquor containing 25 % v/v glycerol for 2 min.

Crystals of NTD-nobeta (unfolded β -hairpin) were grown by mixing 2 μ L of protein at 40 mg/mL in storage buffer, with an equal volume of 100 mM MES pH 5.0, 1.6 M ammonium sulphate. Full-size crystals were obtained approximately within 2 weeks and soaked for 1 min in reservoir solution containing 1.0 M sodium iodide and 25 % v/v glycerol as cryoprotectant.

For the CTD, a 20 mg/mL protein solution in storage buffer was mixed with an equal volume (2 μ L each) of 100 mM Tris-HCl pH 8.5, 1.25 M ammonium sulphate. Two morphologically different crystals, clusters of thin elongated plates and bar-shaped crystals were produced after one week, reaching full size in approximately two weeks. Single 3D bar-shaped crystals were harvested and cryoprotected by soaking in mother liquor supplemented with 25 % v/v glycerol.

Diffraction and Data Processing

X ray diffraction data sets (Supplementary Table 1) were processed using XDS⁽³³⁹⁾ and Scala or Aimless⁽³³⁵⁾. NTD, NTD-nobeta and CTD data sets were collected at 1.5418 Å wavelength, 100 K with a home source (Micromax007-HF, Varimax_HF multilayered mirrors and Mar345 image plate detector). Full-length CAdata were collected at 0.97857 Å wavelength, 100K, at beamline Proxima 1 (Soleil synchrotron) with a Dectris Pilatus3 (6M) hybrid pixel detector.

Structure Determination and Refinement Procedures

The structure of NTD-nobeta was phased using single anomalous diffraction after quick soaking the crystals with Nal. The iodides were identified by direct methods using the anomalous difference amplitudes⁽³⁴⁰⁾, and the substructure refined with SHARP⁽³⁴¹⁾. Protein phasing resulted in an overall phasing power on anomalous difference structure factors of 1.62 (dropping below 1 at 3.0 Å resolution) and 0.392 figure of merit on acentric reflections. Density modification using histogram matching and solvent flipping⁽³⁴²⁾ with 58.4% solvent content, allowed for straightforward definition of the correct heavy atom substructure handedness, and a readily interpretable experimental electron density map. Refinement was performed with Buster⁽³⁴³⁾, iterated with manual rebuilding/validation with Coot⁽³⁴⁴⁾.

The crystal structure of NTD was solved by molecular replacement using NTD-nobeta as search probe with Phaser⁽³⁴⁵⁾. Further refinement and manual rebuilding/validation iterations were performed with Phenix⁽³⁴⁶⁾ and Coot⁽³⁴⁴⁾.

The CTD structure was also solved by molecular replacement, but no available model from the PDB allowed for a correct solution to be identified. The search probes were generated by comparative modelling with the Rosetta suite⁽³⁴⁷⁾, using PDB ID 3H47 as a template. In this way, 10,000 models were obtained, scored and clustered. The top 10 models of each of the best 20 clusters were selected. Regions with greater structural variability were trimmed. An automated script was used to feed each one of these 200 models into Phaser⁽³⁴⁵⁾ to perform MR and the results were ranked according to log likelihood gain. The top models corresponded to right solutions, and used for subsequent stages of refinement and manual rebuilding using Buster⁽³⁴³⁾ and Coot⁽³⁴⁴⁾.

The full-length CA structure was solved by placing 6 copies of each of the high-resolution NTD and CTD structures, in the asymmetric unit using Phaser⁽³⁴⁵⁾. BLV CA crystallized in space group P6₅, with hemihedral twinning. Twinning was detected by analyzing intensity statistics, and different programs ⁽³⁴⁸⁻³⁵⁰⁾ gave consistent results. Refinement was thus performed using a least squares target with twin law 'k, h, -l' as implemented in Phenix⁽³⁴⁶⁾. A large number of CA crystals were used to collect X ray diffraction data, ultimately choosing the one that showed least twinning fraction (refined to 0.42). Refinement with Phenix, iteration with loop building using Rosetta and manual rebuilding/validation with Coot allowed us to refine to crystallographic R values of 18.0% (22.1% R_{free}). The high-resolution structures of BLV NTD and CTD were used as references in the generation of external restraints used throughout refinement.

Simulated annealing in torsional space was performed in the first refinement cycles, to diminish model bias after molecular replacement. Tests were performed by independently excluding different fragments from the molecular replacement models, and confirming the appearance of significant peaks (>5 σ) in the corresponding regions of sigmaA-weighted, $2mF_{obs}$ -DF_{calc} and mF_{obs} -DF_{calc} Fourier maps. As well, several regions and/or side chains that had been trimmed off from the starting molecular replacement solution models, appeared clearly defined in the electron density maps from the very beginning of the refinement procedure. Non-crystallographic symmetry restraints were used throughout refinement.

Structure Analysis

Several tools from the CCP4⁽³⁵¹⁾ and Phenix⁽³⁴⁸⁾ packages were used to manipulate coordinates (rotations, translations), superpose models, calculate Patterson self rotation functions, analyze residue contacts and other general crystallographic procedures. Pymol (v 1.6.9.0, http://pymol.org) was used for visualization, figure production and also for some calculations: the 'rms cur' command to calculate rmsd between atom selections without performing fitting; the 'pair_fit' command to superpose user-defined sets of segments by pairwise fitting; the scripts 'mark center' and 'axes', to draw NCS axes; the plugins 'SuperSym' and 'Supercell' for extended crystallographic symmetry analyses (see http://pymolwiki.org/index.php/ for details on these commands, scripts and plugins).

Non-crystallographic axes were defined by first obtaining a model with averaged atomic coordinates of the involved molecules, which are forming part of dimers (applying one 180° rotation), trimers (applying two successive 120° rotations) or hexamers (5 60° rotations). Such averaged models were then used to calculate the position of the NCS axis by least squares minimization and singular value decomposition (Kabsch algorithm⁽³⁵²⁾, following a Procrustes approach).

Table S1. Data collection and refinement statistics.

	CA	СТD	NTD	NTD-nobeta
Data collection				
Space group	P6 ₅ [¶]	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions				
a, b, c (Å)	94.3, 94.3, 257.2	47.7, 52.9, 90.9	40.7, 56.0, 103.1	65.6, 66.2, 81.3
α, β, γ (°)	90, 90, 120	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	47.14 - 2.75 (2.92 - 2.75) *	45.70 - 2.46 (2.59 – 2.46)	51.53 - 1.45 (1.55 - 1.45)	51.37 - 2.44 (2.45 – 2.44)
Unique reflections measured	33567 (5374)	8737 (1194)	40594 (3621)	13762 (127)
R _{meas}	0.07 (1.07)	0.13 (0.44)	0.06 (0.25)	0.08 (0.29)
l / σl	22.4 (2.05)	10.2 (4.1)	12.0 (5.1)	39.7 (11.1)
Completeness (%)	99.9 (99.5)	98.9 (96.3)	91.4 (90.7)	100 (98.4)
Redundancy	10.5 (10.3)	3.5 (3.5)	2.3 (2.3)	14.2 (14.4)
Refinement				
Resolution (Å)	47.15 - 2.75	35.42 - 2.46	27.74 - 1.45	51.37 - 2.44
Number of reflectionsused	31821 (1676)	7834 (867)	37092 (2304)	12666 (1037)
(number in the free set)				

R _{work} / R _{free}	0.180/0.221	0.199 / 0.253	0.160 / 0.184	0.206 / 0.245
Number of atoms				
Protein	9078	1723	2047 [†]	1817 **
Ligand/ion	0	0	12 (glycerol) / 10 (SO ₄ ²⁻)	6 (glycerol) / 55 (l ⁻) ⁺⁺
Water	0	92	372 [†]	82
<i>B</i> -factors (Å ²)				
Wilson plot	82.2	36.1	7.6	49.5
Protein	102.0 [§]	33.8 [§]	10.6 [§]	43.0
Ligand/ion	-	-	29.0	63.7
Water	-	28.4	23.4	42.9
R.m.s deviations				
Bond lengths (Å)	0.01	0.01	0.01	0.01
Bond angles (°)	1.4	1.2	1.0	1.1
Number of residues in Ramachandran plot regions [‡] (favored/outliers)	1129 / 1	210/1	258 / 0	224 / 0
PDB ID	4PH0	4PH1	4PH2	4PH3

¶ Hemihedral twinning was present according to law k, h, -l (final refined twinning fraction = 0.43); refinement was performed with a lsq_twin amplitudesbased target function ⁽³⁴⁶⁾ *Values in parentheses are for highest-resolution shell.

+ 142 (out of the 1976) protein atoms, corresponding to 10 residues, were modeled with 2 alternate conformations; 5 (out of the 367) water atoms were modeled with 2 alternate conformations

⁺⁺ 56 (out of the 1789) protein atoms, corresponding to 4 residues, were modeled with 2 alternate conformations; 6 (out of the 49) iodide atoms were modeled with 2 alternate conformations

§ Including TLS contribution.

 \ddagger Calculated with Molprobity $^{\scriptscriptstyle (353)}$




Fig. S1. Crystal structures of BLV CA and derived sub-domains. The left panels correspond to cartoon representations of the contents of the asymmetric unit of each structure as labelled, coloured according to a ramp blue-to-red (N-terminal to C-terminal direction of the polypeptide chain). Waters (red) and other ligands (coloured by atom) are also depicted. The right panels are wall-eye stereo figures of representative regions of each of the corresponding structures on the left, showing a chicken-wire representation of the sigmaA-weighted electron density maps $(2mF_{obs}-DF_{calc})$ contoured at 1σ levels (except for NTD, at 1.5σ), as obtained after full refinement. The final refined model is shown as stick representation, coloured by atom, fitted within the density.

Figure S2.



Figure S3. BLV CA compared with other retroviruses. Structure-based sequence alignment of mature CA proteins of BLV, HIV-1, and HTLV-1. The position of secondary structure elements in BLV-CA is shown above the alignment (β -strands in orange, α -helices in blue, 3_{10} helix in yellow). The extension of structurally equivalent regions for each CA protein is highlighted in boxes. Identical residues are coloured in bold red fonts. The position of α 6 (only present in HIV-1 CA) is indicated with a transparent helix and lighter colours. The NTD-CTD hinge loop is depicted in green.

Figure S3.



Fig. S3. BLV CA compared with HTLV-1. Cartoon representation of BLV CA (in green), superimposed onto one of the models of the NMR ensembles of either the NTD of HTLV-1 CA (pdb 1G03) in slate blue, or yet the full length CA of this Deltaretroviridae (pdb 1QRJ) in transparent orange. Note the absence of β -hairpin in the full-length HTLV-1 construct (since part of the C-terminus of the matrix protein had been maintained). Also note the stark difference in the β -hairpin orientations between BLV and HTLV-1, highlighted with a dashed arrow.



Fig. S4. Crystal packing of BLV CA. **(A)** Crystal packing in space group P6₅, showing the unit cell in red with the unique axis in vertical orientation (unit cell axis c, as labelled). A "column" of hexameric units is depicted onto the right of the figure, the bottom three represented in ribbons, and the rest on top in 'all atom' lines. Colours are chosen to differentiate the six different CA hexamers present per unit cell (one per asymmetric unit). From the bottom, one particular atom has been labelled with the name of the protomer chain to which it belongs, highlighting the sinusoidal translation line that connects one protomer of the first hexamer, to the neighbour protomer in the hexamer of the layer on top of it, and so on. **(B)** At the level of the unit cell origin, marked with a dashed line that connects with panel (a), a 90° rotation is applied on the orientation view, according to a horizontal axis on the plane of the paper. Seven interacting hexamers are shown, all coloured by atom except for the one that belongs to the column of panel (a), kept in the original cyan tone for reference. The bottom inset zooms in a single hexamer in the asymmetric unit, with cell axes a and b as labelled.

The simplest crystal built by making lateral contacts to make a 2D lattice and then stacking them would have each CA hexamer maintaining its symmetric assembly such that each protomer in the crystal makes exactly the same contacts. But the packing contacts in the dimension perpendicular to the layers may not be favourable for such a symmetric arrangement, so that the individual hexamers adapt to make stable packing interactions at the expense of symmetry, in a further example of the functional malleability of the CA capsomer.



Fig. S5. Deviation from strict 6-fold symmetry within and among CA hexameric capsomers. **(A)** Similar representation as in main text Fig. 1B, but rotated 90 degrees according to a horizontal axis in the plane of the paper. The crystal unit cell is shown, with the c axis perpendicular to the plane of the paper. The quasi-6-fold axis is shown as a violet stick on the left-most hexamer. Also the two independent quasi-3-fold axes (yellow) and the three quasi-2-fold axes (green), are depicted as sticks. Note the differences in the orientations of all these non-crystallographic symmetry axes, reflecting the intrinsic asymmetry of the hexameric unit. **(B)** Ribbon representation depicting variations of intra-hexamer NTD-CTD interactions, among the six protomers, overlaid in different colours with NTDs superposed. Key interacting residues in sticks. One interacting NTD (labelled NTD') of a neighbour protomer is shown in grey. The six protomers display a spectrum of conformations, with pairwise root mean square deviations (rmsd) ranging from 0.7Å to 3.1Å, after aligning the whole polypeptide. Variable NTD-CTD relative orientations account for most of this variation, rmsd figures dropping to 0.5-0.6Å by separately fitting the NTDs or the CTDs (also showing that both domains are equally flexible).





Fig. S6. Quasi-2-fold axes with CTD-CTD inter-hexamer contact interfaces. Lateral (left panels) and bottom (right panels) views of the dimeric inter-hexamer interfaces, according to the three independent quasi-2-fold axes. Different tones of red and blue distinguish the CTDs of the two protomers involved in each interface, chains A and D on the top, B/E in the middle, and C/F in the bottom rows. Note the overall similar organisation among all, with key interacting amino acids W133 (from the 3_{10} helix at the beginning of the CTD) and I168 (from helix α 9) highlighted with sticks and transparent spheres, as reference points. The hydrophobic packing ensured by W184 and M185 in HIV-1, is functionally conserved in BLV, carried out respectively by W133 (albeit coming from a different main-chain position, at the beginning of the 3_{10} helix) and I168 (from α 9). CTD helices α 10 and α 11 are omitted for clarity.

Referencias

- 1. J. A. Briggs, H. G. Krausslich, The molecular architecture of HIV. *Journal of molecular biology***410**, 491 (Jul 22, 2011).
- 2. B. K. Ganser-Pornillos, M. Yeager, O. Pornillos, Assembly and architecture of HIV. Advances in experimental medicine and biology726, 441 (2012).
- 3. B. K. Ganser-Pornillos, U. K. von Schwedler, K. M. Stray, C. Aiken, W. I. Sundquist, Assembly properties of the human immunodeficiency virus type 1 CA protein. *Journal of virology***78**, 2545 (Mar, 2004).
- 4. J. Benjamin, B. K. Ganser-Pornillos, W. F. Tivol, W. I. Sundquist, G. J. Jensen, Three-dimensional structure of HIV-1 virus-like particles by electron cryotomography. *Journal of molecular biology***346**, 577 (Feb 18, 2005).
- 5. S. Li, C. P. Hill, W. I. Sundquist, J. T. Finch, Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature***407**, 409 (Sep 21, 2000).
- 6. B. K. Ganser, S. Li, V. Y. Klishko, J. T. Finch, W. I. Sundquist, Assembly and analysis of conical models for the HIV-1 core. *Science*283, 80 (Jan 1, 1999).
- 7. R. K. Gitti *et al.*, Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science***273**, 231 (Jul 12, 1996).
- 8. T. R. Gamble *et al.*, Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science***278**, 849 (Oct 31, 1997).
- 9. C. C. Cornilescu, F. Bouamr, X. Yao, C. Carter, N. Tjandra, Structural analysis of the N-terminal domain of the human T-cell leukemia virus capsid protein. *Journal of molecular biology***306**, 783 (Mar 2, 2001).
- 10. G. B. Mortuza *et al.*, High-resolution structure of a retroviral capsid hexameric amino-terminal domain. *Nature***431**, 481 (Sep 23, 2004).
- 11. S. Khorasanizadeh, R. Campos-Olivas, M. F. Summers, Solution structure of the capsid protein from the human T-cell leukemia virus type-I. *Journal of molecular biology***291**, 491 (Aug 13, 1999).
- 12. R. Campos-Olivas, J. L. Newman, M. F. Summers, Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses. *Journal of molecular biology***296**, 633 (Feb 18, 2000).
- 13. O. Pornillos *et al.*, X-ray structures of the hexameric building block of the HIV capsid. *Cell***137**, 1282 (Jun 26, 2009).
- 14. S. Du *et al.*, Structure of the HIV-1 full-length capsid protein in a conformationally trapped unassembled state induced by small-molecule binding. *Journal of molecular biology***406**, 371 (Feb 25, 2011).
- 15. O. Pornillos, B. K. Ganser-Pornillos, M. Yeager, Atomic-level modelling of the HIV capsid. *Nature***469**, 424 (Jan 20, 2011).
- 16. G. Zhao *et al.*, Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature***497**, 643 (May 30, 2013).
- 17. I. J. Byeon *et al.*, Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell***139**, 780 (Nov 13, 2009).
- 18. N. Gillet *et al.*, Mechanisms of leukemogenesis induced by bovine leukemia virus: prospects for novel antiretroviral therapies in human. *Retrovirology***4**, 18 (2007).
- 19. S. M. Rodriguez *et al.*, Preventive and therapeutic strategies for bovine leukemia virus: lessons for HTLV. *Viruses***3**, 1210 (Jul, 2011).
- 20. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta crystallographica*. *Section D, Biological crystallography***69**, 1204 (Jul, 2013).
- 21. G. Cardone, J. G. Purdy, N. Cheng, R. C. Craven, A. C. Steven, Visualization of a missing link in retrovirus capsid assembly. *Nature***457**, 694 (Feb 5, 2009).
- 22. R. L. Kingston *et al.*, Structure and self-association of the Rous sarcoma virus capsid protein. *Structure***8**, 617 (Jun 15, 2000).

- 23. C. Tang, Y. Ndassa, M. F. Summers, Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. *Nature structural biology***9**, 537 (Jul, 2002).
- 24. F. van den Ent, J. Lowe, RF cloning: a restriction-free method for inserting target genes into plasmids. *Journal of biochemical and biophysical methods***67**, 67 (Apr 30, 2006).
- 25. W. Kabsch, Integration, scaling, space-group assignment and post-refinement. *Acta crystallographica*. *Section D, Biological crystallography***66**, 133 (Feb, 2010).
- 26. T. R. Schneider, G. M. Sheldrick, Substructure solution with SHELXD. *Acta crystallographica*. *Section D*, *Biological crystallography***58**, 1772 (Oct, 2002).
- 27. G. Bricogne, C. Vonrhein, C. Flensburg, M. Schiltz, W. Paciorek, Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta crystallographica*. *Section D, Biological crystallography***59**, 2023 (Nov, 2003).
- 28. J. P. Abrahams, A. G. Leslie, Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta crystallographica. Section D, Biological crystallography***52**, 30 (Jan 1, 1996).
- 29. G. Bricogne *et al.* (Global Phasing Ltd. United Kingdom, Cambridge, 2009).
- 30. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta crystallographica*. *Section D, Biological crystallography***66**, 486 (Apr, 2010).
- 31. A. J. McCoy *et al.*, Phaser crystallographic software. *J Appl Crystallogr***40**, 658 (Aug 1, 2007).
- 32. P. V. Afonine *et al.*, Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallographica*. *Section D, Biological crystallography***68**, 352 (Apr, 2012).
- 33. R. Das, D. Baker, Macromolecular modeling with rosetta. *Annual review of biochemistry***77**, 363 (2008).
- 34. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography***66**, 213 (Feb, 2010).
- 35. P. R. Evans, An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta crystallographica. Section D, Biological crystallography***67**, 282 (Apr, 2011).
- 36. J. E. Padilla, T. O. Yeates, A statistic for local intensity differences: robustness to anisotropy and pseudocentering and utility for detecting twinning. *Acta crystallographica. Section D, Biological crystallography***59**, 1124 (Jul, 2003).
- 37. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D*, *Biological crystallography***67**, 235 (Apr, 2011).
- 38. W. Kabsch, A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A***34**, 827 (1978).
- 39. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography***66**, 12 (Jan, 2010).

CAPÍTULO 4

Discusión general y perspectivas

La salida de la partícula retroviral de la célula hospedero es un momento muy particular de la biología de los retrovirus. Una gran cantidad de datos bioquímicos y celulares han cambiado la forma que se describe el papel de la cápside en el ciclo replicativo de los retrovirus. Inicialmente considerada como un "simple" contenedor y protector del genoma viral, rápidamente desechado al infectar una nueva célula, la cápside ha demostrado ser un elemento clave para la infectividad. Cumple un papel esencial (si bien pobremente caracterizado) en controlar los procesos de retro-transcripción y la integración. Ambos procesos están ligados al desensamblado de la cápside, y creciente evidencia indica que ocurre en forma temporal y espacialmente precisa. A su vez, la estructura se mantiene perfectamente estable en el virión maduro en su etapa extracelular. La cápside depende entonces del correcto ensamblado a partir de su único bloque de construcción, la proteína CA, la cual forma una estructura metaestable, capaz de cumplir con su doble función. La CA posee la información para cumplir estos papeles, opuestos e indispensables, economizando en una única proteína la formación de una estructura supramolecular.

Sobre el auto-ensamblado. Durante las pasadas dos décadas aproximadamente, se han investigado intensamente las bases biofísicas y estructurales que gobiernan el auto-ensamblado y arquitectura de la cápside retroviral. La vía de ensamblado no se ha definido experimentalmente, si bien se han logrado avances interesantes mediante dinámica molecular. Los factores fisicoquímicos que controlan el auto-ensamblado se conocen pobremente, habiéndose definido solamente algunos requerimientos principales como ser la elevada fuerza iónica, pero los trabajos se han enfocado casi exclusivamente en HIV-1. Un punto fundamental es que los datos disponibles no permiten proponer un modelo que explique cómo se gatilla el auto-ensamblado en un contexto tipo-fisiológico, ni como se determina la formación de pentámeros además de hexámeros, que permitan formar una cápside cerrada. Esto es particularmente importante en el desarrollo de drogas que interfieren con el ensamblado, ya que el contexto fisicoquímico específico del interior del virión puede determinar condiciones muy distintas de interacción que las evaluadas *in vitro*.

Nuestro resultados mostraron que la CA de BLV ensambla mayoritariamente formando mallas 2D a baja fuerza iónica, a diferencia de la CA de HIV-1 que forma predominantemente tubos, y a elevada fuerza iónica. Un aspecto interesante fue el hecho que el hábito de ensamblado puede modularse mediante iones divalentes como fosfato y sulfato, tanto a baja como alta fuerza iónica aportada por el ión divalente. De hecho, el aumento en fuerza iónica únicamente favorece la velocidad de reacción, sin alterar la

morfología. Nuestra hipótesis es que los efectos del fosfato podrían estar relacionados con un posible papel del anión mimetizando de alguna manera las características del ARN presente en el contexto de la partícula viral. Esto es interesante ya que la CA de BLV tiene la capacidad de ensamblar en condiciones menos extremas: polimeriza óptimamente a la temperatura corporal del bovino y a baja fuerza iónica, si bien se inhibe a fuerza iónica fisiológica. Sin embargo a fuerza iónica en el rango fisiológico pero aportada por dianiones, la proteína polimeriza y además forma partículas que (si bien más pequeñas) son esféricas, indicando una posible inducción de la formación de pentámeros. En este eje de trabajo, nuestro foco estuvo siempre centrado en intentar correlacionar los datos in vitro con el contexto in vivo, ya que éste era un punto controversial y pobremente conocido para retrovirus en general. Un resultado interesante fue demostrar que la proteína no ensambla en absoluto en un tampón CLM (cytosolic like médium) lo cual indicaba que la fuerza iónica monovalente es la principal responsable del arresto. Por tanto las condiciones iónicas del interior de la partícula viral, presumiblemente similares a las del citosol, no son suficientes para gatillar el ensamblado. Esto nos llevó a evaluar el efecto de las dos características salientes del interior del virión: la presencia de ARN y la alta concentración de macromoléculas (crowding macromolecular). Demostramos, hasta donde conocemos por primera en forma sistemática, que en un contexto donde se aporte un solvente CLM y se mimetice la presencia de ARN y crowding molecular similar a las del virión, la proteína no sólo ensambla rápidamente sino que altera su hábito de ensamblado formando exclusivamente partículas esferoidales de tamaño casi idéntico a la cápside auténtica. Este hallazgo es novedoso e importante, ya que establece por primera vez una conexión entre los factores que disparan la reacción in vitro y su pertinencia en el contexto fisiológico en el cual la proteína se encuentra para ensamblar. Imaginamos que una vez libre la alta concentración de CA (≥ 1mM) encuentra un contexto muy favorable para ensamblar, aportado por su propia elevada concentración, la temperatura y pH óptimos del bovino, la presencia de la alta densidad de carga negativa que aporta el ARN viral, y el ambiente de crowding macromolecular aportado por la elevada concentración de macromoléculas presente en la partícula. El efecto del fosfato es interesante en cuanto a que modula el hábito de ensamblado de CA_{BLV} hacia formas esféricas. Al igual que la CA de HIV-1, la CA de BLV presenta un anillo central de Argininas formado por la Arg conservada localizada al inicio de la hélice α 1. Esto sumado al hecho que los dianiones divalente potencial el auto-ensamblado y promueven la formación de exclusivamente partículas esferoidales sugieren que la CA de BLV también podría alternar entre pentámero y hexámero mediante un control electrostático. Esta propiedad de la CA de BLV es compartida con la CA de RSV pero no de HIV-1, indicando que otros factores son necesarios conjuntamente con el control electrostático. En este sentido, el dramático efecto del ARN y el crowding molecular como potenciadores de la polimerización y modificadores del hábito de ensamblado hacia

estructuras tipo cápside permite suponer que las condiciones que la proteína encontrará en el virión una vez libre, le aportarán el ambiente para ensamblar con la rapidez y hacia la arquitectura adecuada.

Sobre la estructura. Por otro lado, la elucidación de la estructura cuaternaria de la cápside retroviral ha constituido uno de los problemas más desafiantes de la retrovirología en las últimas dos décadas, y aún no se ha resuelto completamente. Son varios los motivos que dificultaron/dificultan la elucidación estructural de ensambles de la proteína CA, la cual incluso llegó a ser considerada como no cristalizable. Por otro lado, la tendencia a formar ensambles heterogéneos al polimerizar ha dificultado su elucidación estructural a una resolución suficiente para visualizar los contactos (cadenas laterales) que median la interacción. Durante el desarrollo de esta Tesis, surgieron un conjunto de trabajos realizados por algunos de los grupos líderes en biología estructural de retrovirus que aportaron la primera imagen de la estructura de los capsómeros hexamérico y pentamérico. Estas estructuras permitieron visualizar con detalle atómico las interacciones que mantienen unidos los protómeros, y confirmar las interfaces previamente definidas en estructuras de baja resolución. La comparación de estructuras de baja resolución de arreglos 2D hexaméricos (cristalinas) y de la lattice hexamérica curva (nanotubos) permitieron definir que en ambos casos los hexámeros se unen vía contactos diméricos CTD-CTD. Sin embargo, en la lattice curva, se observó además una interfaz adicional CTD trimérica, la cual se aceptó como exclusiva de la arquitectura curva. Los contactos de esta interfaz se consideraron por tanto necesarios para inducir y/o estabilizar la curvatura de la lattice. Usando esta información y las estructuras de alta resolución de los dominios individuales NTD y CTD (resueltas por cristalografía y/o NMR), recientemente se generó un modelo mediante dinámica molecular (MD) de la cápside completa de HIV-1. Este trabajo es la más grande simulación realizada al momento, y la primera MD de la cápside completa (entre >3 millones de átomos). Este trabajo confirmó la presencia de la interfaz CTD trimérica en la lattice hexagonal curva, y aportó un modelo pseudo-atómico de los contactos inter-capsoméricos⁽¹⁶⁾.

A pesar de estos notables avances, no se ha logrado obtener información estructural con resolución atómica de ensamblados nativos, ya sea de sus bloques de construcción (pentámeros o hexámeros) o de oligómeros de mayor orden. Concretamente, las estructuras cristalinas de los capsómeros provienen de proteínas CA unidas covalentemente mediante puentes de cisteínas a nivel del anillo de NTDs, lo cual tiene el problema de la potencial restricción de la flexibilidad del capsómero. Por otro lado, se mutaron los residuos de CTD que median la dimerización de manera de impedir la polimerización y así generar capsómeros solubles para ensayos de cristalización. Cuando comenzamos los estudios estructurales en este trabajo de Tesis (colaboración con la Unidad de Cristalografía del Institut Pasteur), analizamos la estructura cristalina de la *lattice* 2D hexagonal obtenida usando esta versión mutante de la CA de HIV-1 (PDB 3H4E) y encontramos que las mutaciones en el CTD alteraban profundamente la interacción. Más

específicamente, si bien en el cristal los hexámeros unidos covalentemente forman una capa 2D hexagonal mediante interacciones laterales CTD diméricas, esta interfaz es artefactual, por lo cual los contactos inter-capsómeros no son representativos de la *lattice* auténtica. En principio esto es lógico, ya que los residuos de la hélice α 9 que median la unión CTD-CTD nativa fueron mutados. De hecho, en la estructura la mayor parte de esta hélice (incluyendo los residuos de dimerización Trp184 y Met185) no se ve, indicando que las mutaciones introducidas impiden su plegamiento correcto. Así, de manera fortuita, los CTDs de hexámeros vecinos interaccionan a través de otros contactos, no nativos, vía residuos del extremo de la hélice α 9 que mantuvo su plegamiento normal. Por tanto, los contactos CTD-CTD que median la unión entre hexámeros es artefactual debido a las mutaciones introducidas. En conjunto, si bien se había avanzado mucho en la comprensión estructural de la cápside, aún seguía faltando una estructura que mostrara –a resolución atómica- los detalles de los contactos auténticos, nativos, que mantienen unidos los protómeros formando capsómeros, y los capsómeros formando la lattice madura.

La estructura de la CA de BLV que logramos resolver, tiene varias características de gran relevancia para la comprensión de la cápside retroviral en términos amplios. Por un lado, permite visualizar la estructura del bloque básico de construcción, el capsómero hexamérico, ensamblado a partir de la proteína nativa, no modificada, por lo cual representa fielmente los contactos auténticos. Por otro lado, los hexámeros se mantienen unidos entre sí mediante interacciones laterales vía sus CTDs, reflejando el arreglo hexagonal de la *lattice* madura auténtica, y revelado a alta resolución por primera vez los contactos que median el auto-ensamblado. En conjunto, la estructura mostró que el capsómero de BLV se mantiene unido mediante las interfaces típicas, NTD-NTD, NTD-CTD, y CTD-CTD. Sin embargo, para nuestra sorpresa, encontramos diferencias conceptualmente novedosas e interesantes. Por un lado, el hexámero es asimétrico, con las posiciones de sus protómeros apartándose de la simetría 6 estricta. Por tanto, el eje que pasa a través del anillo hexagonal y que relaciona a cada protómero es de hecho, un eje quasi-6. La asimetría surge del hecho que cada protómero en el hexámero adopta una conformación ligeramente distinta, y tanto el anillo de NTDs como el anillo de CTDs se apartan de la simetría 6 estricta. Este hallazgo contrasta con la visión actual del capsómero retroviral, visión que proviene de las estructuras unidas covalentemente mediante cisteínas. La figura actual del capsómero involucra variaciones en la orientación de los dominios CTDs en relación al anillo central rígido de NTDs. El anillo de NTDs obedece estrictamente la simetría hexagonal, mientras que la movilidad proviene principalmente del cinturón móvil de CTDs; éstos pueden rotar entre sí en la díada CTD-CTD, y respecto al anillo de NTDs mediante el punto de pivote NTD-CTD. Nosotros suponemos que el uso de la CA nativa de BLV precisamente permitió exponer la flexibilidad auténtica del capsómero, al no tener restricción conformacional debida al cross-linking de cisteínas entre protómeros de HIV. Como se ve en la Fig. 12 (a, d), las cisteínas introducidas en el anillo de NTDs de HIV podrían interferir con la libertad

de movimiento de los protómeros, en particular evitando que cada protómero adopte una conformación distinta al "simetrizar" la estructura. Esta posibilidad es, por lo menos, lógica de suponer, ya que los puentes de cisteínas forman un cinturón relativamente rígido localizado justamente en la interfaz entre NTDs. La flexibilidad del capsómero de BLV es funcionalmente interesante, ya que sugiere que el bloque de construcción de la cápside tiene una plasticidad estructural superior a la prevista. El hexámero mantiene la posibilidad de movimientos de tipo cuerpo rígido entre NTDs y CTDs, pero en este caso, ambos anillos son móviles, lo que adiciona un nuevo rango de movimientos. En particular, y favorecido por la extensión de *linker*, la rotación de ambos anillos alrededor del eje de simetría hexagonal permite una deformación importante del capsómero, donde cada anillo puede girar y retorcerse ligeramente. Esto contrasta con la visión de un anillo rígido imperturbable, alrededor del cual el cinturón de CTDs se mueve mediante rotaciones sobre un punto fijo. La "deformabilidad" exhibida por el capsómero de BLV muestra la posibilidad de asociarse a capsómeros vecinos mediante distintas conformaciones, abriendo la puerta a alterar el ángulo y declinaciones relativa a hexámeros adyacentes mediante movimientos más complejos.

Por otro lado, encontramos que los contactos triméricos entre CTDs de hexámeros adyacentes pueden ser variables pero similares, esto es, *quasi*-equivalentes. Lo mismo ocurre con los contactos diméricos CTD, si bien en este caso la variabilidad es mucho menor. Esto plantea un escenario donde los hexámeros se unen a través de dos tipos de contactos a través de sus CTDs: diméricos y triméricos. Los diméricos muestran menos variabilidad y posiblemente esta conservación esté implicada con el mantenimiento de la estructura básica de la *lattice*. Los contactos triméricos son más variables, pudiendo interaccionar mediante interacciones quasi-equivalentes. En el cristal se observan tres tipos de contactos triméricos (3 ejes quasi-3), los cuales cambian el ángulo relativo entre capsómeros vecinos. Este hallazgo es importante ya que demuestran por primera vez a resolución atómica la presencia de contactos *quasi*-equivalentes en la cápside retroviral.

La cápside retroviral es de hecho una estructura supramolecular asimétrica, cuya curvatura varía continuamente entre regiones casi planas y otras de gran curvatura, con capsómeros que requieren establecer contactos diferentes con cada capsómero adyacente, en ambientes locales variables y distribuidos no homogéneamente. Nosotros nos imaginamos que la plasticidad del hexámero de BLV explica este pleiotropismo del capsómero, inferido pero no formalmente demostrado. El mantenimiento de la lattice tiene lugar mediante los contactos más conservados CTD diméricos, si bien es posible que se adapten ligeramente, presumiblemente en sitios donde las distancias inter-capsómero cambien un poco (por ejemplo, regiones planas de hexámeros versus hexámeros unidos a un mismo pentámero). La posibilidad de rotación alrededor del eje *quasi*-2 que relaciona ambos CTD (ya visto en HIV-1) provee movilidad adicional. Las interfaces triméricas por su parte aparecen como la fuente de alteración del

ángulo entre hexámeros advacentes. La posibilidad de formar contactos quasi-equivalente que alteran la declinación entre capsómeros vecinos provee un mecanismo potencial para la generación de curvatura. En el cristal, las distintas interacciones mantenidas vía los ejes quasi-3 permiten distintas declinaciones entre hexámeros adyacentes que permiten lograr el empaquetamiento. En analogía con la situación del cristal, es lógico suponer la posibilidad de transmitir a largo rango diferentes grados de declinación entre hexámeros vecinos, que finalmente permitan inducir curvatura. La combinación de un hexámero plástico, capaz de deformar su simetría estricta y "retorcerse" sobre sí mismo puede alterar el ángulo entre vecinos en mayor medida que un capsómero rígido. En combinación con contactos diméricos capaces de rotar respecto al anillo de NTDs, pero manteniendo el contacto dimérico relativamente invariable, el capsómero proporciona una combinatoria de interacciones. La estructura en el cristal aporta una imagen, un "snapshot" del rango de conformaciones que posiblemente pueda adoptar el capsómero. Imaginamos que el hexámero puede oscilar entre diferentes conformaciones isoenergéticas (energéticamente equivalentes, y por tanto equi-probables), las cuales son seleccionadas según el microambiente local de interacción y curvatura. Consideramos que los ejes quasi-3 son una fuente mayor de ruptura local de la simetría hexagonal de la *lattice*, cuyos efectos se expresan a largo rango en la asimetría global de la cápside.

Los contactos que median la interacción entre protómeros son relativamente escasos y débiles, lo cual concuerda con el hecho que la estructura es metaestable. Una gran cantidad de contactos débiles mantiene la lattice cooperativamente, hasta que el contexto adecuado favorezca su desensamblado. La estructura de BLV y de otras CA retrovirales sugieren que los contactos que mantiene unidos los protómeros son relativamente laxos, y en su mayoría mediados por puentes de moléculas de agua. Esto tiene implicancias en la transición hexámero-pentámero si bien la información disponible no permite claramente predecir qué condiciones favorecerán la predominancia de uno u otro.

Reflexión final y perspectivas. Globalmente, los resultados de esta Tesis permitieron establecer que en condiciones tipo-fisiológicas, la inducción del auto-ensamblado y la adquisición de la morfología supramolecular típica de la cápside madura de *Deltaretrovirus* dependen de las condiciones fisicoquímicas específicas de la partícula viral. Globalmente, estos hallazgos representan un avance muy importante en la comprensión de las bases biofísicas y estructurales del proceso de auto-ensamblado de la proteína de cápside retroviral. Por otro lado, la nueva información sobre la plasticidad y tipo de contactos entre capsómeros, sugiere un mecanismo por el cual puede inducirse y mantener declinaciones distintas entre sí, lo cual es lógico suponer que puede usarse para generar la curvatura de la *lattice*. Este resultado provino no sólo gracias a varios años de esfuerzo sino también al hecho que las propiedades de ensamblado de la CA del BLV, diferenciales respecto a otras como planteamos en nuestra hipótesis de

trabajo, son favorables para lograr su cristalización. Específicamente, la tendencia predominante a formar mallas 2D no sólo fue determinante para la cristalización de esta proteína flexible y "no cristalizable" sino que además permitió elucidar a estructura atómica del hexámero en un ensamble biológicamente relevante, mimetizando la *lattice* madura.

Destacamos también que la producción de CABLV en gran escala también la usamos para el desarrollo de un kit formato ELISA para el diagnóstico de Leucosis Bovina en nuestro país. Así, logramos articular los productos de la ciencia básica con su aplicación biotecnológica.

Un desafío a futuro será correlacionar la información biofísica y estructural de manera de generar un modelo que permita predecir, a partir de una proteína CA determinada, el tipo de morfología supramolecular que adoptará al auto-ensamblar. Esta información será ampliamente valiosa para el desarrollo de drogas anti-retrovirales que interfieran con el ensamblado. La CA madura, así como formando parte de Gag, son blancos potencialmente importantes para bloquear la replicación viral. Asimismo, el conocimiento generado también aportará en el diseño racional de nanoestructuras, un campo de intenso desarrollo, basado en el uso de bloques de construcción capaces de auto-ensamblar.

Nuestra perspectiva es doble. Por un lado, usar la información sobre auto-ensamblado y estructural para evaluar compuestos que puedan interferir con la formación de la cápside. Por otro, pensamos abordar la elucidación de la estructura de los ensambles tubulares de CA_{BLV}, de manera de determinar cómo cambian los contactos en la transición plano-2D de la *lattice* hexamérica.

ANEXO 1.



Subfamilia Orthoretrovirinae				
Genus	Features	Examples		
1. Alpharetrovirus	Simple, Onco	Avian leucosis virus, RSV		
2. Betaretrovirus	Simple, Onco	Mouse Mammary Tumor Virus		
3. Gammaretrovirus	Simple, Onco	Murine leukemia virus (Moloney, Harvey)		
4. Deltaretrovirus	Complex, Onco	Bovine Leukemia (BLV), Human T Cell Leukemia (HTLV)		
5. Epsilonretrovirus	Complex, Onco	Walleye Dermal Sarcoma		
6. Lentivirus	Complex	HIV, Visna, EIAV		
Subfamila Spumaretrovirinae				
Genus	Features	Examples		
1. Spumavirus	Complex	Simian Foamy virus		

Taxonomía de los retrovirus. La clasificación está basada en el último reporte (2013) del Comité Internacional sobre Taxonomía de Virus (ICTV-www.ictvonline.org). (**Arriba**) Filogenia de los Retrovirus. Los géneros que incluyen genomas endógenos se indican con asterisco rojo (Tomado de Weiss et al Retrovirology 2006, 3:67). (**Abajo**) La Familia Retroviridae. Se muestran los distintos géneros de las dos subfamilias de retrovirus. Retrovirus Simples: aquellos que además de los genes que codifican para gag, pro, pol, y env, pueden poseer una región codificante adicional, pero que no dan lugar a proteínas esenciales para la síntesis o procesamiento del ARN; el procesamiento alternativo del ARN genómico puede dar lugar a solamente un ARN mensajero. Retrovirus complejos: Pueden tener múltiples y complejos patrones de procesamiento alternativo, resultando en una amplia variedad de productos génicos, uno de los cuales codifica para la proteína tax, un activador transcripcional. (Tomado de Coffin, J.M. Retroviruses. 1997. Cold Spring Harbor.

ANEXO 2





El icosaedro es una estructura isométrica, con 12 vértices pentaméricos y 20 caras triangulares. Esto determina la presencia de un conjunto de tres ejes de simetría: 6 ejes 5-fold^a que pasan a través de los 12 vértices del icosaedro (esto es, los vértices quedan enfrentados entre sí), 10 ejes 3-fold que pasan a través de las 20 caras triangulares, y 15 ejes 2-fold que conectan las caras vecinas.

^(a) eje de simetría n-fold: esto es, si miramos la cápside a través de cada eje 2-fold, 3-fold, 5-fold, ésta retornará a la misma posición inicial luego de rotarla 180°, 120°, y 72° respectivamente

REFERENCIAS

- 1. B. K. Ganser-Pornillos, M. Yeager, O. Pornillos, *Adv Exp Med Biol* 726, 441.
- 2. M. Yeager, J Mol Biol 410, 534 (Jul 22).
- 3. N. Arhel, *Retrovirology* **7**, 96.
- 4. N. Gillet *et al.*, *Retrovirology* **4**, 18 (2007).
- 5. B. K. Ganser-Pornillos, M. Yeager, W. I. Sundquist, *Curr Opin Struct Biol* **18**, 203 (Apr, 2008).
- 6. R. Swanstrom, J. W. Willis, *Retroviruses*. J. M. Coffin, S. M. Hughes, H. E. Varmus, Eds., (Cold Spring Harbor Laboratory Press, 1997).
- 7. W. I. Sundquist, H. G. Krausslich, *Cold Spring Harb Perspect Med* **2**, a006924 (Jul).
- 8. J. R. Lingappa, J. C. Reed, M. Tanaka, K. Chutiraka, B. A. Robinson, Virus Res, (Jul 24).
- 9. S. Du et al., J Mol Biol 406, 371 (Feb 25).
- 10. B. K. Ganser-Pornillos, A. Cheng, M. Yeager, *Cell* **131**, 70 (Oct 5, 2007).
- 11. L. Balvay, M. Lopez Lastra, B. Sargueil, J. L. Darlix, T. Ohlmann, Nat Rev Microbiol 5, 128 (Feb, 2007).
- 12. V. D'Souza, M. F. Summers, *Nat Rev Microbiol* **3**, 643 (Aug, 2005).
- 13. A. Rein, S. A. Datta, C. P. Jones, K. Musier-Forsyth, Trends Biochem Sci 36, 373 (Jul).
- 14. A. G. Cashikar *et al.*, *Elife*, e02184 (May 30).
- 15. I. J. Byeon *et al.*, *Cell* **139**, 780 (Nov 13, 2009).
- 16. G. Zhao *et al.*, *Nature* **497**, 643 (May 30).
- 17. O. Pornillos, B. K. Ganser-Pornillos, M. Yeager, *Nature* **469**, 424 (Jan 20).
- 18. B. N. Fields, P. M. Howley, Fields Virology. R. Press, Ed., (New York, ed. 3rd, 1996).
- 19. M. Brahic, J. F. Bureau, Ann Neurol 42, 984 (Dec, 1997).
- 20. R. F. DiGiacomo, S. G. Hopkins, Vet Clin North Am Food Anim Pract 13, 177 (Mar, 1997).
- 21. N. Rosenberg, P. Jolicoeur, (1997).
- 22. R. C. Tustin, J. S. Afr. Vet. Med. Assoc. 40, 3 (1969).
- 23. R. Johnson, J. B. Kaneene, Vet Bull 62, 287 (1992).
- 24. J Exp Med 201, 320 (Feb 7, 2005).
- 25. L. Gross, Proc Soc Exp Biol Med **76**, 27 (1951).
- 26. L. Gross, Proc Soc Exp Biol Med 94, 767 (1957).
- 27. B. J. Poiesz et al., Proc Natl Acad Sci U S A 77, 7415 (Dec, 1980).
- 28. K. Barmak, E. Harhaj, C. Grant, T. Alefantis, B. Wigdahl, Virology 308, 1 (Mar 30, 2003).
- 29. F. Barre-Sinoussi et al., Science 220, 868 (May 20, 1983).
- 30. M. S. Gottlieb *et al.*, *N Engl J Med* **305**, 1425 (Dec 10, 1981).
- 31. G. Maartens, C. Celum, S. R. Lewin, *Lancet* **384**, 258 (Jul 19).
- 32. E. S. Lander *et al., Nature* **409**, 860 (Feb 15, 2001).
- 33. E. B. Chuong, M. A. Rumi, M. J. Soares, J. C. Baker, Nat Genet 45, 325 (Mar).
- 34. D. Haig, *Bioessays* **35**, 845 (Oct).
- 35. V. M. Vogt, (1997).
- 36. A. B. Rabson, *Retroviruses*. J. M. Coffin, S. M. Hughes, H. E. Varmus, Eds., (Cold Spring Harbor Laboratory Press, 1997).
- 37. Y. Aida, H. Murakami, M. Takahashi, S. N. Takeshima, *Front Microbiol* **4**, 328.
- 38. G. Gutierrez et al., Viruses 6, 2416 (Jun).
- 39. Y. Somura, E. Sugiyama, H. Fujikawa, K. Murakami, Arch Virol, (Jun 12).
- 40. G. C. Buehring et al., Emerg Infect Dis 20, 772 (May).
- 41. G. Gutierrez, I. Alvarez, R. Merlini, F. Rondelli, K. Trono, *BMC Vet Res* **10**, 82.
- 42. G. Gutierrez et al., BMC Vet Res 8, 187.
- 43. S. M. Rodriguez *et al.*, *Viruses* **3**, 1210 (Jul).
- 44. R. Zaffaroni et al., V Jornadas Técnicas Veterinarias, 150 (2007).
- 45. R. Swanstrom, J. W. Wills, (1997).
- 46. A. Rein, Arch Virol Suppl **9**, 513 (1994).
- 47. R. Berkowitz, J. Fisher, S. P. Goff, *Curr Top Microbiol Immunol* **214**, 177 (1996).

- 48. N. A. Jewell, L. M. Mansky, J Gen Virol 81, 1889 (Aug, 2000).
- 49. T. Jacks *et al.*, *Nature* **331**, 280 (Jan 21, 1988).
- 50. L. Llames *et al.*, *Virus Res* **79**, 47 (Nov 5, 2001).
- 51. Y. Yoshinaka, I. Katoh, T. D. Copeland, G. W. Smythers, S. Oroszlan, J Virol 57, 826 (Mar, 1986).
- 52. G. Franchini, R. Fukumoto, J. R. Fullen, Int J Hematol 78, 280 (Nov, 2003).
- 53. N. Sagata *et al., EMBO J* **3**, 3231 (Dec 20, 1984).
- 54. B. K. Felber, D. Derse, A. Athanassopoulos, M. Campbell, G. N. Pavlakis, *New Biol* 1, 318 (Dec, 1989).
- 55. A. Florins *et al.*, *J Virol* **81**, 10195 (Sep, 2007).
- 56. L. Willems *et al.*, *Proc Natl Acad Sci U S A* **91**, 11532 (Nov 22, 1994).
- 57. V. M. Vogt, M. N. Simon, *J Virol* **73**, 7050 (Aug, 1999).
- 58. J. A. Briggs et al., Nat Struct Mol Biol **11**, 672 (Jul, 2004).
- 59. J. Ghysdael, R. Kettmann, A. Burny, *J Virol* **29**, 1087 (Mar, 1979).
- 60. J. M. Miller, L. D. Miller, C. Olson, K. G. Gillette, J Natl Cancer Inst 43, 1297 (Dec, 1969).
- 61. I. Katoh, T. Yasunaga, Y. Yoshinaka, J Virol 67, 1830 (Apr, 1993).
- 62. X. Gao, R. X. Shen, W. H. Xiang, J. H. Zhou, *Bing Du Xue Bao* 24, 487 (Nov, 2008).
- 63. J. C. Paillart, M. Shehu-Xhilaga, R. Marquet, J. Mak, Nat Rev Microbiol 2, 461 (Jun, 2004).
- 64. C. Giroud, N. Chazal, L. Briant, *Retrovirology* **8**, 71.
- 65. D. E. Ott, *Rev Med Virol* **18**, 159 (May-Jun, 2008).
- 66. T. Jacob, C. Van den Broeke, H. W. Favoreel, *J Virol* **85**, 1158 (Feb).
- 67. D. E. Ott, *Rev Med Virol* **7**, 167 (Sep, 1997).
- 68. S. Cen *et al.*, *J Virol* **75**, 5043 (Jun, 2001).
- 69. E. K. Franke, H. E. Yuan, J. Luban, *Nature* **372**, 359 (Nov 24, 1994).
- 70. J. Gabor, S. Cen, H. Javanbakht, M. Niu, L. Kleiman, J Virol 76, 9096 (Sep, 2002).
- 71. D. E. Ott et al., AIDS Res Hum Retroviruses 11, 1003 (Sep, 1995).
- 72. M. Thali *et al., Nature* **372**, 363 (Nov 24, 1994).
- 73. J. Lanman *et al.*, *Nat Struct Mol Biol* **11**, 676 (Jul, 2004).
- 74. E. Morita *et al.*, *Cell Host Microbe* **2**, 41 (Jul 12, 2007).
- 75. E. Morita *et al., EMBO J* **26**, 4215 (Oct 3, 2007).
- 76. A. Patnaik, V. Chau, J. W. Wills, *Proc Natl Acad Sci U S A* **97**, 13069 (Nov 21, 2000).
- 77. A. Patnaik, J. W. Wills, *J Virol* **76**, 2789 (Mar, 2002).
- 78. A. Scott *et al.*, *Proc Natl Acad Sci U S A* **102**, 13813 (Sep 27, 2005).
- 79. W. I. Sundquist *et al., Mol Cell* **13**, 783 (Mar 26, 2004).
- 80. U. K. von Schwedler *et al.*, *Cell* **114**, 701 (Sep 19, 2003).
- 81. D. M. Ward et al., J Biol Chem 280, 10548 (Mar 18, 2005).
- 82. Q. Zhai *et al., Nat Struct Mol Biol* **15**, 43 (Jan, 2008).
- 83. S. Andreadis *et al.*, *J Virol* **74**, 1258 (Feb, 2000).
- 84. J. Kimpton, M. Emerman, J Virol 66, 2232 (Apr, 1992).
- 85. M. Piatak, Jr. et al., Science **259**, 1749 (Mar 19, 1993).
- 86. S. Zarkik *et al.*, *FEBS Lett* **406**, 205 (Apr 7, 1997).
- 87. Y. Kubo, H. Hayashi, T. Matsuyama, H. Sato, N. Yamamoto, *Adv Virol* **2012**, 640894.
- 88. G. B. Mortuza *et al.*, *Nature* **431**, 481 (Sep 23, 2004).
- 89. B. M. Forshey, U. von Schwedler, W. I. Sundquist, C. Aiken, J Virol 76, 5667 (Jun, 2002).
- 90. D. McDonald *et al.*, *J Cell Biol* **159**, 441 (Nov 11, 2002).
- 91. D. Warrilow *et al., J Virol* **82**, 1425 (Feb, 2008).
- 92. Y. Suzuki, R. Craigie, Nat Rev Microbiol 5, 187 (Mar, 2007).
- 93. M. Bukrinsky, *Mol Med* **10**, 1 (Jan-Jun, 2004).
- 94. J. D. Dvorin, M. H. Malim, *Curr Top Microbiol Immunol* **281**, 179 (2003).
- 95. J. Lehmann-Che, A. Saib, *AIDS Rev* 6, 199 (Oct-Dec, 2004).
- 96. D. Warrilow, G. Tachedjian, D. Harrich, *Rev Med Virol* **19**, 324 (Nov, 2009).
- 97. G. J. Klarmann, C. A. Schauber, B. D. Preston, *J Biol Chem* **268**, 9793 (May 5, 1993).
- 98. N. J. Arhel *et al., EMBO J* **26**, 3025 (Jun 20, 2007).
- 99. D. J. Dismuke, C. Aiken, *J Virol* **80**, 3712 (Apr, 2006).
- 100. K. Lee *et al., Cell Host Microbe* **7**, 221 (Mar 18).

- 101. M. Yamashita, M. Emerman, J Virol **78**, 5670 (Jun, 2004).
- 102. M. Yamashita, O. Perez, T. J. Hope, M. Emerman, *PLoS Pathog* **3**, 1502 (Oct 26, 2007).
- 103. L. Karageorgos, P. Li, C. Burrell, AIDS Res Hum Retroviruses 9, 817 (Sep, 1993).
- 104. Z. Ambrose, C. Aiken, *Virology* **454-455**, 371 (Apr).
- 105. K. Warren, D. Warrilow, L. Meredith, D. Harrich, Viruses 1, 873 (Dec, 2009).
- 106. D. Warrilow, K. Warren, D. Harrich, *PLoS One* **5**, e13229.
- 107. S. P. Goff, *J Gene Med* **3**, 517 (Nov-Dec, 2001).
- 108. S. Li, C. P. Hill, W. I. Sundquist, J. T. Finch, *Nature* 407, 409 (Sep 21, 2000).
- 109. J. Balzarini *et al.*, *Virology* **192**, 246 (Jan, 1993).
- 110. H. Zhang et al., AIDS Res Hum Retroviruses 9, 1287 (Dec, 1993).
- 111. A. E. Hulme, O. Perez, T. J. Hope, Proc Natl Acad Sci U S A 108, 9975 (Jun 14).
- 112. Y. Yang, T. Fricke, F. Diaz-Griffero, J Virol 87, 683 (Jan).
- 113. H. Zhang, G. Dornadula, J. Orenstein, R. J. Pomerantz, J Hum Virol 3, 165 (May-Jun, 2000).
- 114. V. Arfi *et al., J Virol* **83**, 7524 (Aug, 2009).
- 115. D. Warrilow, D. Stenzel, D. Harrich, *Retrovirology* 4, 77 (2007).
- 116. A. Roa et al., J Virol 86, 1717 (Feb).
- 117. C. Tipper, J. Sodroski, J Virol 87, 3628 (Apr).
- 118. C. Aberham, S. Weber, W. Phares, J Virol 70, 3536 (Jun, 1996).
- 119. D. Braaten *et al., J Virol* **70**, 5170 (Aug, 1996).
- 120. C. Cartier *et al., J Virol* **71**, 4832 (Jun, 1997).
- 121. P. Gupta *et al., J Mol Biol* **410**, 681 (Jul 22).
- 122. S. Misumi et al., J Biol Chem 285, 25185 (Aug 13).
- 123. Y. Li, A. K. Kar, J. Sodroski, J Virol 83, 10951 (Nov, 2009).
- 124. T. Hatziioannou, D. Perez-Caballero, S. Cowan, P. D. Bieniasz, J Virol 79, 176 (Jan, 2005).
- 125. M. Qi, R. Yang, C. Aiken, *J Virol* **82**, 12001 (Dec, 2008).
- 126. T. Schaller *et al.*, *PLoS Pathog* **7**, e1002439 (Dec).
- 127. L. M. Ylinen *et al.*, *J Virol* **83**, 2044 (Feb, 2009).
- 128. N. Arhel, F. Kirchhoff, *Biochim Biophys Acta* 1802, 313 (Mar).
- 129. V. P. Basu et al., Virus Res 134, 19 (Jun, 2008).
- 130. A. Telesnitsky, S. P. Goff, (1997).
- 131. M. I. Bukrinsky et al., Proc Natl Acad Sci U S A 90, 6125 (Jul 1, 1993).
- 132. A. Fassati, S. P. Goff, *J Virol* **75**, 3626 (Apr, 2001).
- 133. M. D. Miller, C. M. Farnet, F. D. Bushman, J Virol **71**, 5382 (Jul, 1997).
- 134. N. Arhel et al., Nat Methods **3**, 817 (Oct, 2006).
- 135. F. Diaz-Griffero *et al., Virology* **369**, 400 (Dec 20, 2007).
- 136. S. Nisole, J. P. Stoye, A. Saib, *Nat Rev Microbiol* **3**, 799 (Oct, 2005).
- 137. C. Besnier, Y. Takeuchi, G. Towers, Proc Natl Acad Sci U S A 99, 11920 (Sep 3, 2002).
- 138. S. Cowan et al., Proc Natl Acad Sci U S A 99, 11914 (Sep 3, 2002).
- 139. G. Towers et al., Proc Natl Acad Sci U S A 97, 12295 (Oct 24, 2000).
- 140. M. G. Grutter, J. Luban, *Curr Opin Virol* **2**, 142 (Apr).
- 141. S. Neil, P. Bieniasz, J Interferon Cytokine Res 29, 569 (Sep, 2009).
- 142. M. Stremlau et al., Nature 427, 848 (Feb 26, 2004).
- 143. L. Krishnan *et al., J Virol* **84**, 397 (Jan).
- 144. K. A. Matreyek, S. S. Yucel, X. Li, A. Engelman, *PLoS Pathog* 9, e1003693.
- 145. A. Engelman, G. Englund, J. M. Orenstein, M. A. Martin, R. Craigie, J Virol 69, 2729 (May, 1995).
- 146. G. Englund, T. S. Theodore, E. O. Freed, A. Engelman, M. A. Martin, J Virol 69, 3216 (May, 1995).
- 147. R. L. LaFemina et al., J Virol 66, 7414 (Dec, 1992).
- 148. H. Sakai *et al., J Virol* **67**, 1169 (Mar, 1993).
- 149. M. Stevenson, T. L. Stanwick, M. P. Dempsey, C. A. Lamonica, EMBO J 9, 1551 (May, 1990).
- 150. H. Murakami *et al., Virus Res* **156**, 107 (Mar).
- 151. M. P. Sherman, W. C. Greene, *Microbes Infect* 4, 67 (Jan, 2002).
- 152. C. Bolinger, K. Boris-Lawrie, *Retrovirology* 6, 8 (2009).
- 153. N. Jouvenet, P. D. Bieniasz, S. M. Simon, *Nature* **454**, 236 (Jul 10, 2008).

- 154. S. Ivanchenko *et al., PLoS Pathog* 5, e1000652 (Nov, 2009).
- 155. K. Fahr, H. C. Oppermann, E. Willich, *Rofo* **126**, 571 (Jun, 1977).
- 156. M. J. Hayman, Virology 85, 241 (Mar, 1978).
- 157. H. C. Oppermann, L. Wille, U. Bleyl, M. Obladen, Pediatr Radiol 5, 137 (Mar 17, 1977).
- 158. H. G. Gottlinger, *AIDS* **15 Suppl 5**, S13 (2001).
- 159. S. B. Kutluay, P. D. Bieniasz, *PLoS Pathog* 6, e1001200.
- 160. M. A. Checkley, B. G. Luttge, E. O. Freed, J Mol Biol 410, 582 (Jul 22).
- 161. M. C. Johnson, AIDS Res Hum Retroviruses 27, 239 (Mar).
- 162. P. R. Tedbury, E. O. Freed, *Trends Microbiol* 22, 372 (Jul).
- 163. T. Murakami, *Mol Biol Int* **2012**, 682850.
- 164. N. Sharova, C. Swingler, M. Sharkey, M. Stevenson, EMBO J 24, 2481 (Jul 6, 2005).
- 165. K. Simons, D. Toomre, Nat Rev Mol Cell Biol 1, 31 (Oct, 2000).
- 166. P. D. Bieniasz, *Cell Host Microbe* **5**, 550 (Jun 18, 2009).
- 167. J. G. Carlton, J. Martin-Serrano, Science **316**, 1908 (Jun 29, 2007).
- 168. E. Morita, W. I. Sundquist, Annu Rev Cell Dev Biol 20, 395 (2004).
- 169. Y. Usami et al., Biochem Soc Trans 37, 181 (Feb, 2009).
- 170. J. H. Hurley, S. D. Emr, Annu Rev Biophys Biomol Struct 35, 277 (2006).
- 171. R. L. Williams, S. Urbe, *Nat Rev Mol Cell Biol* **8**, 355 (May, 2007).
- 172. E. O. Freed, J Virol 76, 4679 (May, 2002).
- 173. H. G. Gottlinger, T. Dorfman, J. G. Sodroski, W. A. Haseltine, *Proc Natl Acad Sci U S A* 88, 3195 (Apr 15, 1991).
- 174. M. Huang, J. M. Orenstein, M. A. Martin, E. O. Freed, J Virol 69, 6810 (Nov, 1995).
- 175. B. A. Puffer, L. J. Parent, J. W. Wills, R. C. Montelaro, *J Virol* **71**, 6541 (Sep, 1997).
- 176. B. Yuan, X. Li, S. P. Goff, *EMBO J* 18, 4700 (Sep 1, 1999).
- 177. O. Pornillos, S. L. Alam, D. R. Davis, W. I. Sundquist, *Nat Struct Biol* 9, 812 (Nov, 2002).
- 178. O. Pornillos *et al.*, *EMBO J* **21**, 2397 (May 15, 2002).
- 179. O. Pornillos, J. E. Garrus, W. I. Sundquist, *Trends Cell Biol* 12, 569 (Dec, 2002).
- 180. I. Le Blanc, M. C. Prevost, M. C. Dokhelar, A. R. Rosenberg, J Virol 76, 10024 (Oct, 2002).
- 181. H. Wang, N. J. Machesky, L. M. Mansky, J Virol 78, 1503 (Feb, 2004).
- 182. H. Wang, K. M. Norris, L. M. Mansky, *J Virol* **76**, 8485 (Aug, 2002).
- 183. C. S. Adamson, E. O. Freed, Adv Pharmacol 55, 347 (2007).
- 184. E. O. Freed, *Virology* **251**, 1 (Nov 10, 1998).
- 185. L. Garnier, L. J. Parent, B. Rovinski, S. X. Cao, J. W. Wills, J Virol 73, 2309 (Mar, 1999).
- 186. C. A. Carter, L. S. Ehrlich, Annu Rev Microbiol 62, 425 (2008).
- 187. A. Dordor, E. Poudevigne, H. Gottlinger, W. Weissenhorn, Future Microbiol 6, 1159 (Oct).
- 188. J. Martin-Serrano, S. J. Neil, Nat Rev Microbiol 9, 519 (Jul).
- 189. S. Santos, Y. Obukhov, S. Nekhai, M. Bukrinsky, S. Iordanskiy, Retrovirology 9, 65.
- 190. N. M. Bell, A. M. Lever, Trends Microbiol 21, 136 (Mar).
- 191. M. A. Massiah et al., J Mol Biol 244, 198 (Nov 25, 1994).
- 192. S. Matthews et al., Nature **370**, 666 (Aug 25, 1994).
- 193. C. Tang et al., Proc Natl Acad Sci U S A 101, 517 (Jan 13, 2004).
- 194. C. P. Hill, D. Worthylake, D. P. Bancroft, A. M. Christensen, W. I. Sundquist, *Proc Natl Acad Sci U S A* **93**, 3099 (Apr 2, 1996).
- 195. E. Hamard-Peron, D. Muriaux, *Retrovirology* **8**, 15.
- 196. V. Chukkapalli, A. Ono, J Mol Biol **410**, 512 (Jul 22).
- 197. I. Katoh, H. Kyushiki, Y. Sakamoto, Y. Ikawa, Y. Yoshinaka, J Virol 65, 6845 (Dec, 1991).
- 198. M. D. Resh, *Biochim Biophys Acta* 1451, 1 (Aug 12, 1999).
- 199. W. Zhou, M. D. Resh, *J Virol* **70**, 8540 (Dec, 1996).
- 200. L. Hermida-Matsumoto, M. D. Resh, J Virol 73, 1902 (Mar, 1999).
- 201. F. Li, C. Wild, Curr Opin Investig Drugs 6, 148 (Feb, 2005).
- 202. C. R. Erdie, J. W. Wills, J Virol 64, 5204 (Oct, 1990).
- 203. A. K. Dalton, P. S. Murray, D. Murray, V. M. Vogt, J Virol 79, 6227 (May, 2005).
- 204. J. S. Saad *et al., J Mol Biol* **382**, 434 (Oct 3, 2008).

- 205. J. Inlora, V. Chukkapalli, D. Derse, A. Ono, *J Virol* **85**, 3802 (Apr).
- 206. A. Ono, S. D. Ablan, S. J. Lockett, K. Nagashima, E. O. Freed, *Proc Natl Acad Sci U S A* **101**, 14889 (Oct 12, 2004).
- 207. J. S. Saad et al., Proc Natl Acad Sci U S A 103, 11364 (Jul 25, 2006).
- 208. J. Vlach, J. S. Saad, Proc Natl Acad Sci U S A 110, 3525 (Feb 26).
- 209. R. A. Dick, S. L. Goh, G. W. Feigenson, V. M. Vogt, Proc Natl Acad Sci U S A 109, 18761 (Nov 13).
- 210. R. C. Aloia, H. Tian, F. C. Jensen, Proc Natl Acad Sci U S A 90, 5181 (Jun 1, 1993).
- 211. R. C. Aloia, F. C. Jensen, C. C. Curtain, P. W. Mobley, L. M. Gordon, *Proc Natl Acad Sci U S A* **85**, 900 (Feb, 1988).
- 212. G. van Meer, D. R. Voelker, G. W. Feigenson, Nat Rev Mol Cell Biol 9, 112 (Feb, 2008).
- 213. A. G. Stephen *et al.*, *J Biomol Tech* **18**, 259 (Sep, 2007).
- 214. J. A. Thomas, R. J. Gorelick, Virus Res 134, 39 (Jun, 2008).
- 215. R. D. Berkowitz, A. Ohagen, S. Hoglund, S. P. Goff, J Virol 69, 6445 (Oct, 1995).
- 216. M. Cruceanu, R. J. Gorelick, K. Musier-Forsyth, I. Rouzina, M. C. Williams, *J Mol Biol* **363**, 867 (Nov 10, 2006).
- 217. J. G. Levin, J. Guo, I. Rouzina, K. Musier-Forsyth, Prog Nucleic Acid Res Mol Biol 80, 217 (2005).
- 218. L. J. Parent, N. Gudleski, J Mol Biol 410, 553 (Jul 22).
- 219. S. A. Datta et al., J Mol Biol 365, 812 (Jan 19, 2007).
- 220. S. A. Datta et al., J Mol Biol **365**, 799 (Jan 19, 2007).
- 221. S. A. Datta et al., J Mol Biol 406, 205 (Feb 18).
- 222. S. D. Fuller, T. Wilk, B. E. Gowen, H. G. Krausslich, V. M. Vogt, *Curr Biol* **7**, 729 (Oct 1, 1997).
- 223. A. Alfadhli, R. L. Barklis, E. Barklis, *Virology* **387**, 466 (May 10, 2009).
- 224. M. A. Accola, B. Strack, H. G. Gottlinger, *J Virol* **74**, 5395 (Jun, 2000).
- 225. Y. Zhang, H. Qian, Z. Love, E. Barklis, *J Virol* **72**, 1782 (Mar, 1998).
- 226. S. Campbell, A. Rein, J Virol **73**, 2270 (Mar, 1999).
- 227. A. Cimarelli, S. Sandin, S. Hoglund, J. Luban, J Virol 74, 3046 (Apr, 2000).
- 228. D. Muriaux, J. Mirro, K. Nagashima, D. Harvin, A. Rein, J Virol 76, 11405 (Nov, 2002).
- 229. J. L. Neira, FEBS J 276, 6110 (Nov, 2009).
- 230. J. A. Briggs et al., Proc Natl Acad Sci U S A 106, 11090 (Jul 7, 2009).
- 231. T. A. Bharat et al., Proc Natl Acad Sci U S A 111, 8233 (Jun 3).
- 232. T. A. Bharat *et al., Nature* **487**, 385 (Jul 19).
- 233. L. A. Carlson et al., Cell Host Microbe 4, 592 (Dec 11, 2008).
- 234. E. R. Wright et al., EMBO J 26, 2218 (Apr 18, 2007).
- 235. J. A. Briggs, H. G. Krausslich, J Mol Biol 410, 491 (Jul 22).
- 236. T. Wilk et al., J Virol **75**, 759 (Jan, 2001).
- 237. S. Campbell *et al.*, *Proc Natl Acad Sci U S A* **98**, 10875 (Sep 11, 2001).
- 238. I. Gross et al., EMBO J 19, 103 (Jan 4, 2000).
- 239. F. H. Crick, J. D. Watson, *Nature* **177**, 473 (Mar 10, 1956).
- 240. A. Klug, D. L. Caspar, *Adv Virus Res* **7**, 225 (1960).
- 241. H. P. Arnold, U. Ziese, W. Zillig, Virology 272, 409 (Jul 5, 2000).
- 242. A. Bize et al., Res Microbiol **159**, 358 (Jun, 2008).
- 243. D. Prangishvili, P. Forterre, R. A. Garrett, Nat Rev Microbiol 4, 837 (Nov, 2006).
- 244. D. Prangishvili, R. A. Garrett, Biochem Soc Trans 32, 204 (Apr, 2004).
- 245. D. Prangishvili, R. A. Garrett, *Trends Microbiol* **13**, 535 (Nov, 2005).
- 246. D. Prangishvili, K. Stedman, W. Zillig, *Trends Microbiol* 9, 39 (Jan, 2001).
- 247. K. A. Dryden *et al.*, *J Cell Biol* **122**, 1023 (Sep, 1993).
- 248. S. Hafenstein et al., Proc Natl Acad Sci U S A 104, 6585 (Apr 17, 2007).
- 249. O. T. Fackler, B. M. Peterlin, *Curr Biol* **10**, 1005 (Aug 24, 2000).
- 250. T. Forest, S. Barnard, J. D. Baines, *Nat Cell Biol* **7**, 429 (Apr, 2005).
- 251. V. L. Giranda *et al.*, *Proc Natl Acad Sci U S A* **89**, 10213 (Nov 1, 1992).
- 252. U. F. Greber, M. Way, *Cell* **124**, 741 (Feb 24, 2006).
- 253. U. F. Greber, M. Willetts, P. Webster, A. Helenius, *Cell* **75**, 477 (Nov 5, 1993).
- 254. D. Marchant, S. J. Neil, K. Aubin, C. Schmitz, A. McKnight, *J Virol* **79**, 9410 (Aug, 2005).

- 255. T. Maurin, D. Fenard, G. Lambeau, A. Doglio, *J Mol Biol* **367**, 702 (Mar 30, 2007).
- 256. E. J. Platt, M. M. Gomes, D. Kabat, J Virol 88, 4304 (Apr).
- 257. N. L. Goicochea *et al., J Mol Biol* **410**, 667 (Jul 22).
- 258. B. K. Ganser, S. Li, V. Y. Klishko, J. T. Finch, W. I. Sundquist, Science 283, 80 (Jan 1, 1999).
- 259. E. Barklis et al., Journal of Molecular Biology 387, 375 (2009).
- 260. I. M. Jones, Y. Morikawa, *Reviews in medical virology* **8**, 87 (1998).
- 261. H. G. Krausslich, A. M. Traenckner, F. Rippmann, Adv Exp Med Biol **306**, 417 (1991).
- 262. C. Tang, J. M. Louis, A. Aniana, J. Y. Suh, G. M. Clore, *Nature* **455**, 693 (Oct 2, 2008).
- 263. Y. X. Feng et al., Proc Natl Acad Sci U S A 93, 7577 (Jul 23, 1996).
- 264. W. Fu, R. J. Gorelick, A. Rein, *J Virol* **68**, 5013 (Aug, 1994).
- 265. W. Fu et al., J Virol 80, 1242 (Feb, 2006).
- 266. W. Fu, A. Rein, *J Virol* **67**, 5443 (Sep, 1993).
- 267. M. D. Moore *et al., Virology* **379**, 152 (Sep 15, 2008).
- 268. D. Muriaux, J. Mirro, D. Harvin, A. Rein, *Proc Natl Acad Sci U S A* **98**, 5246 (Apr 24, 2001).
- 269. B. N. Kelly et al., Biochemistry 45, 11257 (Sep 26, 2006).
- 270. C. Tang, Y. Ndassa, M. F. Summers, Nature structural biology 9, 537 (Jul, 2002).
- 271. D. Ivanov et al., Proc Natl Acad Sci U S A 104, 4353 (Mar 13, 2007).
- 272. D. K. Worthylake, H. Wang, S. Yoo, W. I. Sundquist, C. P. Hill, *Acta Crystallogr D Biol Crystallogr* **55**, 85 (Jan, 1999).
- 273. J. L. Neira, FEBS J 276, 6097 (Nov, 2009).
- 274. L. Deshmukh et al., J Am Chem Soc 135, 16133 (Oct 30).
- 275. R. K. Gitti *et al., Science* **273**, 231 (Jul 12, 1996).
- 276. T. R. Gamble *et al., Science* **279**, 849 (1997).
- 277. G. Cardone, J. G. Purdy, N. Cheng, R. C. Craven, A. C. Steven, *Nature* **457**, 694 (2009).
- 278. S. Khorasanizadeh, R. Campos-Olivas, M. F. Summers, J Mol Biol 291, 491 (Aug 13, 1999).
- 279. U. K. von Schwedler *et al., EMBO J* **17**, 1555 (Mar 16, 1998).
- 280. E. B. Monroe, S. Kang, S. K. Kyere, R. Li, P. E. Prevelige, Jr., Structure 18, 1483 (Nov 10).
- 281. D. Muriaux, J. L. Darlix, A. Cimarelli, *Curr Pharm Des* **10**, 3725 (2004).
- 282. X. Yu et al., Biomacromolecules **10**, 390 (Feb 9, 2009).
- 283. J. McDermott, L. Farrell, R. Ross, E. Barklis,

Journal of Virology **70**, Structural analysis of human immunodeficiency virus type 1 Gag protein interactions (1996).

- 284. L. S. Ehrlich, T. Liu, S. Scarlata, B. Chu, A. Carter, Journal of Virology 66, 4874 (1992).
- 285. I. Gross, H. Hohenberg, H. G. Krausslich, *Eur J Biochem* **249**, 592 (Oct 15, 1997).
- 286. R. L. Kingston *et al., Structure* **8**, 617 (2000).
- 287. J. G. Purdy, J. M. Flanagan, I. J. Ropson, R. C. Craven, Journal of Molecular Biology 386, 300 (2009).
- 288. J. G. Purdy, J. M. Flanagan, I. J. Ropson, K. E. Rennoll-Bankert, R. C. Craven, J Virol 82, 5951 (Jun, 2008).
- 289. J. Lanman et al., J Mol Biol **325**, 759 (Jan 24, 2003).
- 290. C. Butan, D. C. Winkler, J. B. Heymann, R. C. Craven, A. C. Steven, J Mol Biol 376, 1168 (Feb 29, 2008).
- 291. J. B. Heymann, C. Butan, D. C. Winkler, R. C. Craven, A. C. Steven, *Comput Math Methods Med* 9, 197 (2008).
- 292. Benjamin J., B. K. Ganser-Pornillos, W. F. Tivol, W. I. Sundquist, G. J. Jense, *Journal of Molecular Biology* **346**, 577 (2005).
- 293. B. K. Ganser-Pornillos, U. K. von Schwedler, K. M. Stray, C. Aiken, W. I. Sundquist, *Journal of virology* **78**, 2545 (Mar, 2004).
- 294. P. W. Fowler, D. E. Manolopoulas, *An Atlas of Fulleres*. M. Dover Publications, N.Y., Ed., (1995), pp. 1-392.
- 295. J. Lanman, J. Sexton, M. Sakalian, P. E. Prevelige, Jr., J Virol **76**, 6900 (Jul, 2002).
- 296. M. del Alamo, M. G. Mateu, J Mol Biol 345, 893 (Jan 28, 2005).
- 297. M. del Alamo, G. Rivas, M. G. Mateu, J Virol 79, 14271 (Nov, 2005).
- 298. C. C. Douglas, D. Thomas, J. Lanman, P. E. Prevelige, Jr., Biochemistry 43, 10435 (Aug 17, 2004).
- 299. J. A. Briggs et al., Structure 14, 15 (Jan, 2006).
- 300. J. A. Briggs, T. Wilk, R. Welker, H. G. Krausslich, S. D. Fuller, *EMBO J* 22, 1707 (Apr 1, 2003).

- 301. J. Benjamin, B. K. Ganser-Pornillos, W. F. Tivol, W. I. Sundquist, G. J. Jensen, *Journal of molecular biology* **346**, 577 (Feb 18, 2005).
- 302. O. Pornillos *et al.*, *Cell* **137**, 1282 (Jun 26, 2009).
- 303. U. K. von Schwedler, K. M. Stray, J. E. Garrus, W. I. Sundquist, J Virol 77, 5439 (May, 2003).
- 304. K. Chen, G. Piszczek, C. Carter, N. Tjandra, J Biol Chem 288, 1511 (Jan 18).
- 305. J. B. Bowzard, J. W. Wills, R. C. Craven, J Virol **75**, 6850 (Aug, 2001).
- 306. L. A. Alcaraz, M. del Alamo, F. N. Barrera, M. G. Mateu, J. L. Neira, Biophys J 93, 1264 (Aug 15, 2007).
- 307. V. Bartonova et al., J Biol Chem 283, 32024 (Nov 14, 2008).
- 308. H. C. Wong, R. Shin, N. R. Krishna, *Biochemistry* **47**, 2289 (Feb 26, 2008).
- 309. R. Campos-Olivas, J. L. Newman, M. F. Summers, J Mol Biol 296, 633 (Feb 18, 2000).
- 310. B. Meng, A. M. Lever, *Retrovirology* **10**, 5.
- 311. N. Kaushik-Basu, A. Basu, D. Harris, *BioDrugs* **22**, 161 (2008).
- 312. S. M. Rodriguez *et al.*, *Viruses* **3**, 1210 (Jul, 2011).
- 313. V. M. Vogt, in *Retroviruses, J. M. Coffin, Hughes, H. E., Varmus, H. E., Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1997), pp. 27-70.*
- 314. Z. Jin, L. Jin, D. L. Peterson, C. L. Lawson, J Mol Biol 286, 83 (Feb 12, 1999).
- 315. A. J. Birkett, B. Yelamos, I. Rodriguez-Crespo, F. Gavilanes, D. L. Peterson, *Biochim Biophys Acta* **1339**, 62 (Apr 25, 1997).
- 316. G. B. Mortuza et al., J Mol Biol **386**, 1179 (Mar 6, 2009).
- 317. J. K. Hyun, M. Radjainia, R. L. Kingston, A. K. Mitra, J Biol Chem 285, 15056 (May 14).
- 318. M. Yeager, J Mol Biol 410, 534 (Jul 22, 2011).
- 319. F. Weiland, S. Ueberschar, Arch Virol 52, 187 (1976).
- 320. J. M. Andreu, S. N. Timasheff, *Methods Enzymol* 130, 47 (1986).
- 321. F. N. Barrera, M. del Alamo, M. G. Mateu, J. L. Neira, *Biophys J* 94, L8 (Jan 15, 2008).
- 322. S. Leikin, D. C. Rau, P. V.A, Nat Struct Biol 2, 205 (Mar, 1995).
- 323. A. Alfadhli, T. C. Dhenub, A. Still, E. Barklis, J Virol 79, 14498 (Dec, 2005).
- 324. D. Huseby, R. L. Barklis, A. Alfadhli, E. Barklis, J Biol Chem 280, 17664 (May 6, 2005).
- 325. Y. Morikawa, T. Goto, F. Momose, *Journal of Biological Chemistry* **279**, 31964 (2009).
- 326. J. E. Swatton, C. W. Taylor, *J Biol Chem* **277**, 17571 (May 17, 2002).
- 327. A. M. Rossi, C. W. Taylor, *Nat Protoc* **6**, 365 (Mar).
- 328. J. A. Briggs, H. G. Krausslich, Journal of molecular biology **410**, 491 (Jul 22, 2011).
- 329. B. K. Ganser-Pornillos, M. Yeager, O. Pornillos, *Advances in experimental medicine and biology* **726**, 441 (2012).
- 330. T. R. Gamble *et al., Science* **278**, 849 (Oct 31, 1997).
- 331. C. C. Cornilescu, F. Bouamr, X. Yao, C. Carter, N. Tjandra, *Journal of molecular biology* **306**, 783 (Mar 2, 2001).
- 332. S. Du et al., Journal of molecular biology 406, 371 (Feb 25, 2011).
- 333. O. Pornillos, B. K. Ganser-Pornillos, M. Yeager, *Nature* **469**, 424 (Jan 20, 2011).
- 334. G. Zhao *et al., Nature* **497**, 643 (May 30, 2013).
- 335. P. R. Evans, G. N. Murshudov, *Acta crystallographica. Section D, Biological crystallography* **69**, 1204 (Jul, 2013).
- 336. G. Cardone, J. G. Purdy, N. Cheng, R. C. Craven, A. C. Steven, *Nature* **457**, 694 (Feb 5, 2009).
- 337. R. L. Kingston *et al.*, *Structure* **8**, 617 (Jun 15, 2000).
- 338. F. van den Ent, J. Lowe, *Journal of biochemical and biophysical methods* **67**, 67 (Apr 30, 2006).
- 339. W. Kabsch, Acta crystallographica. Section D, Biological crystallography 66, 133 (Feb, 2010).
- 340. T. R. Schneider, G. M. Sheldrick, *Acta crystallographica. Section D, Biological crystallography* **58**, 1772 (Oct, 2002).
- 341. G. Bricogne, C. Vonrhein, C. Flensburg, M. Schiltz, W. Paciorek, *Acta crystallographica. Section D, Biological crystallography* **59**, 2023 (Nov, 2003).
- 342. J. P. Abrahams, A. G. Leslie, *Acta crystallographica. Section D, Biological crystallography* **52**, 30 (Jan 1, 1996).
- 343. G. Bricogne *et al.* (Global Phasing Ltd. United Kingdom, Cambridge, 2009).

- 344. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta crystallographica. Section D, Biological crystallography* **66**, 486 (Apr, 2010).
- 345. A. J. McCoy *et al.*, *J Appl Crystallogr* **40**, 658 (Aug 1, 2007).
- 346. P. V. Afonine *et al., Acta crystallographica. Section D, Biological crystallography* **68**, 352 (Apr, 2012).
- 347. R. Das, D. Baker, Annual review of biochemistry 77, 363 (2008).
- 348. P. D. Adams *et al.*, *Acta crystallographica*. *Section D, Biological crystallography* **66**, 213 (Feb, 2010).
- 349. P. R. Evans, Acta crystallographica. Section D, Biological crystallography 67, 282 (Apr, 2011).
- 350. J. E. Padilla, T. O. Yeates, Acta crystallographica. Section D, Biological crystallography **59**, 1124 (Jul, 2003).
- 351. M. D. Winn et al., Acta crystallographica. Section D, Biological crystallography 67, 235 (Apr, 2011).
- 352. W. Kabsch, Acta Crystallographica Section A **34**, 827 (1978).
- 353. V. B. Chen et al., Acta crystallographica. Section D, Biological crystallography 66, 12 (Jan, 2010).

Tabla de contenido

RESUMEN	4	
CAPÍTULO 1 Biología Celular y Estructural del Ensamblado de los Retrovirus	8	
1.1 LOS RETROVIRUS	8	
Deltaretrovirus y el Virus de la Leucemia Bovina	8	
Organización Genómica de los Retrovirus	9	
Estructura y Composición de la Partícula Retroviral	11	
Ciclo Replicativo de los Retrovirus	14	
La fase temprana: desde el ingreso a la célula a la integración del genoma viral		
La fase tardía: desde la producción y ensamlado de los componentes virales, a la salida del virión		
1.2 ENSAMBLADO Y ARQUITECTURA DE LA PARTÍCULA RETROVIRAL INMADURA	21	
Un paso previo a la producción de una partícula infecciosa	21	
Gag, una maquinaria molecular de ensamblado de retrovirus		
Unión de Gag a la membrana plasmática: El dominio MA		
Selección y empaquetamiento del genoma viral: el papel de NC y su vínculo con MA	25	
Ensamblado y organización de la lattice inmadura: el papel del dominio CA		
1.3 ENSAMBLADO Y ARQUITECTURA DE LA PARTÍCULA RETROVIRAL INMADURA	21	
La cápside viral: solución elegante y eficaz con recursos limitados		
Maduración de la <i>lattice</i> inmadura	32	
La proteína de cápside	33	
¿Cómo se ensambla la cápside retroviral?		
Arquitectura de la cápside madura: múltiples enfoques para unaestructura desafiante		
La interfaz NTD- NTD		
La interfaz NTD- CTD	40	
La interfaz CTD-		
CTD	41	

El elusivo mecanismo de generación de curvatura en la cápside	41
Sobre esta Tesis: ánimo y objetivos	45
Objetivo General y Objetivos Específicos	46
CAPÍTULO 2 Auto-ensamblado de CA _{BLV}	47
Previa al capítulo: "Self-assembly of Bovine Leukemia Virus capsid protein: implications for del assembly"	taretroviral 47
Índice de Figuras Capítulo 2	51
Resultados y Discusión	52
CAPÍTULO 3 Estructura de CA _{BLV}	87
Previa al capítulo: "Crystal Structure of Native Hexameric Lattice Reveals High Conformational Plasticity"	
Índice de Figuras Capítulo 3	90
Resultados y Discusión	