

## DESCUBRIENDO AL VIH: EL GENOMA

Hoy en día, y al contrario de lo que ocurre con otras enfermedades graves causadas por virus, es raro encontrar una persona que no sepa medianamente lo que es el VIH o la enfermedad que ocasiona (SIDA); quizás sea porque nos afecta de lleno en nuestro bonito primer mundo. Pero también, desgraciadamente, una parte de estas personas sienten un tremendo rechazo hacia los afectados por este virus, aún cuando hay suficiente información accesible sobre el mismo que desmiente muchos de los tópicos que aún persisten en la población.

Desde hace tiempo investigadores de todo el mundo están realizando un esfuerzo inmenso para tratar de comprender los secretos moleculares que guarda, y pese a que se ha avanzado mucho en el control de la infección, aún no se ha conseguido dar con la clave para acabar con esta epidemia de distribución mundial.

Un aspecto importante del VIH es su genoma, en el cual nos vamos a detener en los siguientes párrafos tratando de explicar las distintas partes que lo componen, así como las proteínas a las que da lugar, las cuales permiten que finalmente nuevas partículas víricas salgan de la célula hospedadora; sin embargo antes de meternos de lleno en este tema vamos a hacer una pequeña introducción para situar al lector.

El VIH o virus de la inmunodeficiencia humana es un virus perteneciente a la familia *Retroviridae*, subfamilia *Orthoretrovirinae* y al género *Lentivirus*. Como todos sabréis, es el causante del SIDA o síndrome de inmunodeficiencia adquirida.

Se conocen dos tipos de VIH genética y antigénicamente diferentes, denominados VIH-1 y VIH-2. El VIH-1 es el agente responsable de la epidemia mundial de SIDA, mientras que el VIH-2 es un virus endémico del África Oriental y es raro encontrarlo fuera de esta región. Los genomas de estos dos tipos de virus tienen una similitud de sólo el 40-50%; sin embargo ocasionan una enfermedad clínicamente indistinguible, aunque el VIH-2 presenta una patogenicidad más baja y una progresión más lenta a SIDA, por lo que la enfermedad es menos agresiva que la producida por el VIH-1. En esta ocasión, debido a su mayor importancia y para no complicar más las cosas, sólo vamos a hacer referencia al genoma y a la estructura del VIH-1.

Cada virión presenta, a grandes rasgos, una estructura esférica cuya parte externa está rodeada por una envuelta lipídica procedente de la membrana plasmática de la célula hospedadora. En el interior de esta membrana se encuentra una matriz que separa la envuelta de la cápside, la cual es de tipo cilíndrico-cónico y contiene en su interior el genoma vírico (ARN monocatenario diploide de polaridad positiva) junto con varias enzimas, entre las que se encuentra la transcriptasa inversa. Vamos a ver una imagen estructural del VIH de forma que, más adelante, cuando se expliquen los distintos productos génicos poder tener una mejor visión de la implicación de cada uno en la estructura final del virus:

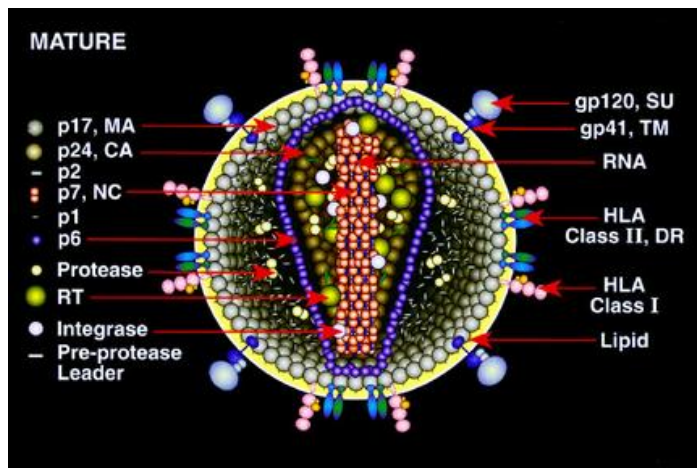


Imagen 1. Estructura del VIH. Vía: [wikimedia](https://www.wikimedia.org/)

Como acabamos de ver, el virión del VIH contiene en su interior dos copias de ARNm de polaridad positiva. ¿Esto qué quiere decir? La polaridad positiva del ARN vírico refleja que ésta es igual a la del ARN mensajero, es decir, aquel ARN que se obtiene al transcribir el ADN y a partir del cual se producen las proteínas por traducción. En el caso contrario se hablaría de polaridad negativa. De esta forma, se puede deducir que un virus con ARNm de polaridad positiva podrá usar ésta cadena como un ARNm desde donde se producirán las proteínas víricas directamente. ¿Ocurre esto en el VIH? Pues no, los retrovirus (entre los que se incluye el VIH), a pesar de tener este tipo de ARN no pueden utilizarlo directamente como ARNm, sino que necesitan un paso intermedio en el cual, a partir del ARN, se obtiene una molécula de ADN bicatenario (ADNbc), denominada provirus, que se integra en el genoma de la célula hospedadora y a partir de la cual se producen las proteínas víricas previo paso por ARNm. Como podemos ver este proceso rompe con el dogma central de la biología molecular en el sentido de que permite la producción de ADN a partir de ARN, hecho por lo cual se denomina transcripción inversa o retrotranscripción y está a cargo de una enzima llamada transcriptasa inversa o retrotranscriptasa; pero todo este fascinante proceso se explicará de forma detallada en una futura entrada.

En ésta ocasión vamos a describir las distintas partes en las que se divide el ARN vírico, así como el papel que tiene cada una de estas en la estructura final del virión. Cabe decir que aunque normalmente se nos presenta esta molécula como una estructura lineal, lo cierto es que la verdadera estructura secundaria dista mucho de ser así, sino que más bien está formada por una serie de bucles y horquillas donde se encuentran las diferentes regiones. Pero aún falta bastante por saber y existe una exhaustiva investigación en este campo, por lo que intentaré exponer aquí de forma entendible la organización del genoma vírico sin entrar demasiado en la arquitectura molecular del mismo.

Si observamos los extremos en la imagen siguiente, cada molécula de ARN vírico presenta un **Cap** en 5' y una cola de **poli-A** en el extremo 3'.

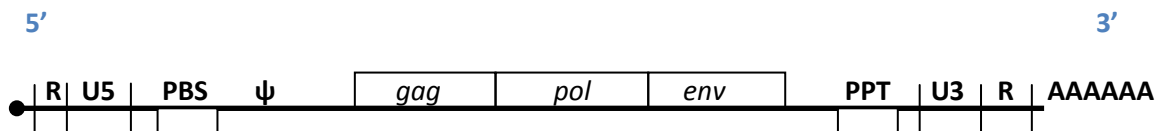


Imagen 2. Organización general del genoma vírico. Se representan algunas de las secuencias más importantes, así como el cap 5' (●) y la cola de poli-A en 3' (AAA...)

En el extremo 5', justo tras el Cap, nos encontramos la región 5'-UTR cuya estructura secundaria consiste en una serie de horquillas conectadas entre sí como son:

- **Región de transactivación (TAR):** juega un papel esencial en la activación transcripcional mediada por la proteína Tat.
- **Señal de poliadenilación (polyA):** como su propio nombre indica, esta región contiene la señal de poliadenilación AAUAAA, la cual también se encuentra cerca del extremo 3'. Aunque se sabe que la señal de poliadenilación localizada en 3' funciona en la maduración del mRNA, no se sabe mucho acerca de la función de la misma señal localizada en 5'.
- **U5:** contiene una secuencia ATT necesaria para la integración del genoma vírico en el ADN de la célula hospedadora y para el empaquetamiento del genoma del virus.
- **Sitio PBS:** es crítico para la replicación ya que sirve como sitio de unión para un tRNA<sup>Lys</sup> celular, el cebador para la transcripción inversa, y además sirve como elemento regulador que afecta a la iniciación de la misma.
- **Sitio de inicio de la dimerización (DIS):** actúa en la dimerización de las dos moléculas de RNA que finalmente formarán parte del virión.
- **SD (splice donor):** parece que juega un rol importante en el splicing de los mRNA víricos.
- **ψ:** se piensa que esta secuencia es la que actúa como señal para el empaquetamiento del genoma del virus, aunque parece que no es la única responsable de este proceso.

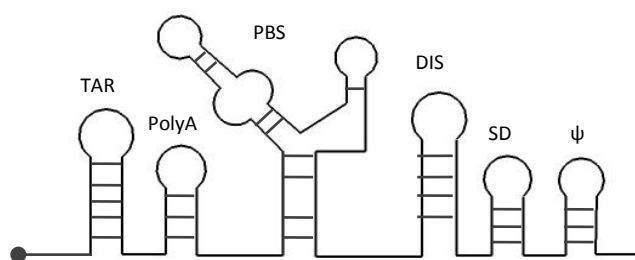


Imagen 3. Detalle la región 5'-UTR del ARN vírico. Se muestra la estructura secundaria orientativa de las distintas regiones que lo componen.

Si nos vamos al otro extremo del ARN vírico, nos encontramos la región 3'-UTR la cual está caracterizada por presentar una cola de poliadenilación (poli-A), una región U3 y una secuencia denominada ppt. La región U3 al igual que la U5 también contiene un sitio ATT, mientras que el sitio ppt es una región de polipurinas compuesta por 9 nucleótidos que actúan como cebadores para la síntesis de la cadena positiva de ADN en la transcripción inversa.

Tanto el 5'-UTR como el 3'-UTR son regiones no codificantes del genoma vírico. ¿Esto qué quiere decir? Pues que estas secuencias no son “leídas” por los ribosomas y por tanto no se traducen a ningún tipo de proteína. ¿Eso implica que no sean importantes? Al contrario, estas regiones contienen motivos regulatorios que son cruciales para que prácticamente todo el ciclo replicativo del virus se lleve a cabo de forma correcta. Entre las distintas funciones que llevan a cabo están: activar la transcripción, iniciar la transcripción inversa, facilitar la dimerización del genoma, empaquetamiento del virus, interactuar con proteínas virales y del hospedador, etc.

Una vez descritos los dos extremos del ARN vírico, y antes de pasar a las secuencias codificantes del genoma, creo que es oportuno hablar aquí de la transcripción inversa en el sentido de los cambios que se producen en los extremos de la molécula de ADN generada por este proceso.

La nueva molécula de ADN obtenida por acción de la transcriptasa inversa, además de ser de doble cadena, contiene en sus extremos lo que se denominan LTR o repeticiones largas terminales. Estas secuencias se producen en el proceso de retrotranscripción y son claves, primero para la integración del ADN vírico en el genoma de la célula hospedadora, y luego para la transcripción correcta del mismo. Cada LTR está dividida en 3 zonas (U3, R y U5) tal como se puede ver en la imagen siguiente:

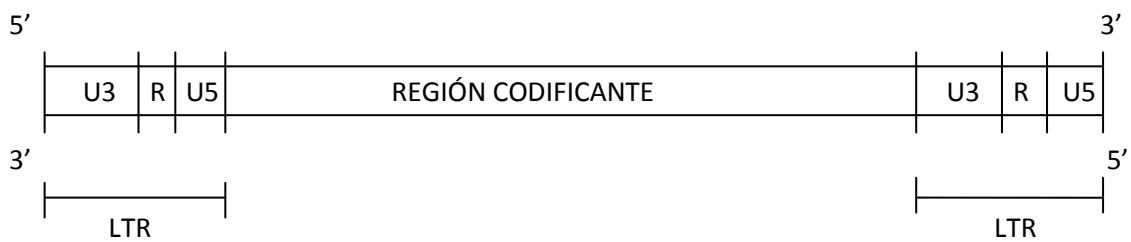


Figura 4. Imagen de la molécula de DNA del VIH donde se muestran las secuencias LTR en ambos extremos.

La generación de estos LTR tiene lugar en el proceso de transcripción inversa como ya hemos dicho anteriormente y es uno de los motivos de por qué el DNA es algo más grande que el RNA original. La importancia de estas regiones así como la forma en la que se generan se explicará detalladamente en un próximo capítulo sobre este apasionante tema.

Ahora que ya hemos pasado por los extremos del genoma nos vamos a ir a la región central, donde se encuentran las secuencias codificantes y en las que podemos encontrar tres genes principales, comunes a todos los retrovirus:

- **gag:** codifica una poliproteína Gag (o antígeno específico de grupo) de 55 kDa (p55) que se escinde por la proteasa viral para formar las siguientes proteínas estructurales
  - o La proteína de matriz **p17 (MA)** que se encuentra anclada al interior de la membrana gracias a una miristilación previa en su extremo amino-terminal.
  - o La proteína de la cápside **p24 (CA)** que, como su propio nombre indica, es la que, mediante polimerización, da lugar a la cápside cilíndrica-cónica del virus.

- Péptido espaciador 1, **p2 (SP1)** compuesto sólo por 14 aminoácidos, participa en el ensamblaje de CA para formar la cápside.
  - La proteína de la nucleocápside **p7 (NC)** cuya función principal (o más conocida) es reconocer el ARN del virus y formar la nucleocápside; sin embargo también se une de modo específico a la señal de empaquetamiento  $\psi$  y realiza una serie de funciones distintas sumamente importantes.
  - Péptido espaciador 2, **p1 (SP2)** de 16 aminoácidos, el cual parece que podría ser importante para la incorporación de Gag y Pol en el virión.
  - La proteína **p6** que junto con la proteína p7 (NC) forman la nucleocápside, y además, por otro lado, se une a una proteína accesoria del virus (Vpr) conduciendo la incorporación de la misma al interior de los viriones.
- **pol:** las proteínas que da lugar este gen se producen a partir de una proteína precursora llamada Gag-Pol (p160), la cual se genera por un cambio de lectura ribosomal, provocado por un motivo específico del ARN en la región distal del gen *gag*. A partir de este precursor se obtienen:
- **Proteasa (p10):** es una proteína dimérica que actúa escindiendo los precursores poliproteicos Gag y Gag-Pol durante la maduración del virus.
  - **Transcriptasa inversa:** es la polimerasa vírica, cuya forma funcional predominante es un heterodímero de **p65** y **p50**. Actúa generando una molécula de ADN bicatenaria a partir del ARN monocatenario original del virus.
  - **Integrasa (p31):** esta proteína es la que actúa en la inserción del ADN vírico en el ADN de la célula hospedadora infectada.
- **env:** da lugar a una proteína precursora de 160 KDa (**gp160**) la cual es sintetizada en el retículo endoplasmático para posteriormente ser transportada al aparato de Golgi, donde será glicosilada antes de sufrir una escisión por una proteasa celular, para dar lugar a dos glicoproteínas que se encuentran unidas de forma no covalente formando las espículas, que se proyectan en la superficie externa del virus y que son esenciales para el reconocimiento y la infección de la célula hospedadora:
- **gp120 (SU):** forma un heterotrímero que se aloja en la zona más externa de la membrana
  - **gp41 (TM):** forma un tronco transmembrana de 3 moléculas de esta proteína al que se unen de forma no covalente las tres moléculas de gp120 para formar una espícula de la envuelta vírica.

Además de estos tres genes principales, común a todos los retrovirus, el VIH presenta una serie de genes adicionales que codifican proteínas reguladoras como son:

- **Tat:** este gen codifica la proteína Tat, un activador de la transcripción. La proteína Tat se une al ARN vírico (a diferencia de los factores de transcripción convencionales) en una estructura en horquilla en la región 5' del genoma conocida como TAR y que hemos visto anteriormente. La unión de esta proteína activa la transcripción de los genes del VIH unas 1000 veces más.

- **Rev:** codifica una proteína denominada Rev, que se une de forma específica al ARN vírico y favorece la transición de la expresión génica del VIH de la fase temprana a la fase tardía. Rev se une a una estructura secundaria del ARN denominada RRE (elemento de respuesta a Rev) que se encuentra dentro de la región codificante del gen env.

Aparte de estas dos proteínas reguladoras tenemos una serie de genes que codifican proteínas accesorias, las cuales no son imprescindibles para la replicación vírica *in vitro*, pero representan factores críticos para la virulencia *in vivo*. La mayoría de estas proteínas tienen múltiples funciones:

- **Nef:** codifica una proteína temprana miristilada del virus, cuyo único exón está localizado cerca del extremo 3'. Entre sus funciones destacan:
  - o La disminución de la tasa de expresión del receptor CD4 de los linfocitos T.
  - o Interferencia en la activación de linfocitos T, de forma que parece que puede ejercer distintos efectos sobre la activación de estas células según el contexto de su expresión.
  - o Aumento de la infectividad de los viriones.
  - o Favorece la progresión hacia SIDA.
- **Vif:** codifica un polipéptido de 23 KDa con las siguientes funciones:
  - o Altera la actividad de la proteína antivírica humana APOBEC3G, una citidina desaminasa con capacidad de provocar mutaciones en el genoma vírico.
  - o Aumenta la infectividad del VIH entre 100 y 1000 veces
- **Vpu:** produce una fosfoproteína integral de membrana cuyo gen solapa en su extremo 3' con el gen *env*. De esta forma Vpu y Env se expresan a partir del mismo ARNm. Sin embargo la proteína Vpu se produce en concentraciones mucho más bajas que la proteína Env. Se piensa que esto es debido a un mecanismo conocido como "leaky scanning" donde se traducen más proteínas en función de que el contexto alrededor del codón de iniciación de la proteína sea más o menos favorable para la lectura a cargo de los ribosomas. De esta forma, el codón de iniciación de Env es más favorable que el de Vpu, lo que explica la diferencia en la concentración de ambas proteínas. Las dos principales funciones de Vpu son:
  - o Regulación negativa del receptor CD4, de forma que reduce la cantidad de las moléculas de este receptor que se expresan en la superficie celular.
  - o Facilita el ensamblaje y favorece la liberación de los viriones de la superficie celular.
- **Vpr:** codifica una proteína que se incorpora al interior de las partículas víricas mediante interacciones específicas con el precursor Gag (p55) y con una gran participación de la proteína p6. Entre sus funciones destacan dos:

- Parece acelerar el proceso de replicación vírica, actuando como un factor de transporte del ADN vírico al núcleo, facilitando así la localización del complejo de preintegración.
- Produce un bloqueo de la división celular.

En la siguiente imagen podéis ver la disposición de los genes que acabamos de ver así como las distintas proteínas a las que finalmente dan lugar:

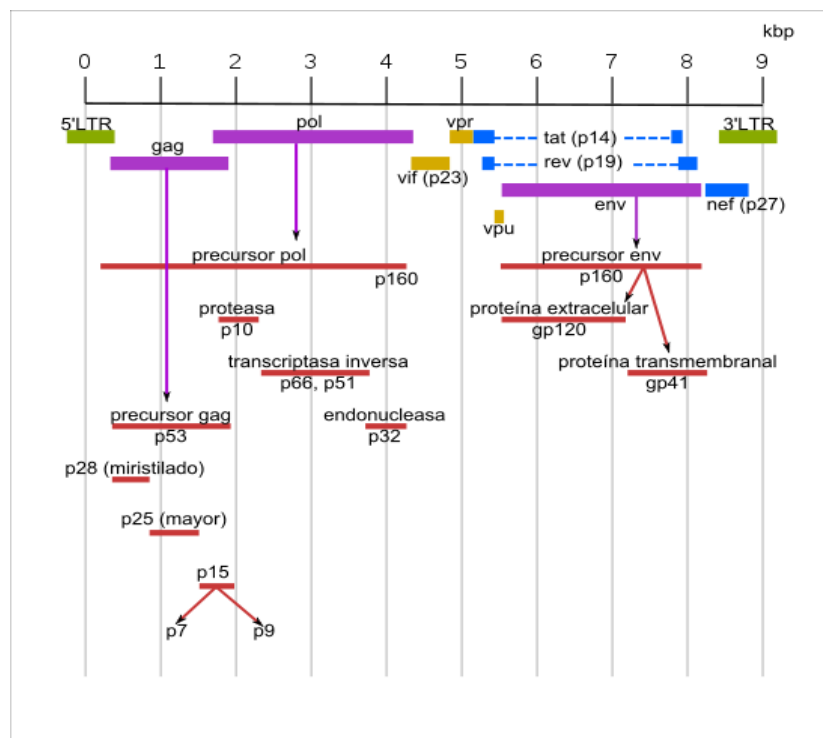


Imagen 5. Se muestra la localización en el genoma de los distintos genes y secuencias que lo componen así como las proteínas a las que dan lugar finalmente. NOTA: algunas proteínas no coinciden con las descritas en el texto, aunque se pueden intuir y/o identificar con el nombre. Vía: [wikimedia](https://commons.wikimedia.org/wiki/File:HIV1_Genome_Map.png). Licencia: [CC](https://creativecommons.org/licenses/by/4.0/)

Para finalizar me gustaría decir que, como suele pasar en estos casos, aún no se conoce en su totalidad la arquitectura genómica del VIH, así como tampoco algunas de las funciones que llevan a cabo las proteínas aquí descritas; incluso a veces hay contradicciones entre los propios investigadores, lo que hace aún más difícil la tarea de intentar explicar este tema de una forma clara. Sin embargo, y a pesar de todo esto, también hay que recalcar que no son pocos los avances en este campo, ni lo son tampoco los esfuerzos realizados en la lucha contra este temido, pero apasionante virus.

## REFERENCIAS

- Bell NM, Lever AML (2013). HIV Gag polyprotein: processing and early viral particle assembly. *Trends in Microbiology*. 21 (3): 136-142
- Delgado R (2011). Características virológicas del VIH. *Enf Infecc Microbio Clin*. 29(1):58-65
- Lu K, Heng X, Summers MF (2011). Structural Determinants and Mechanism of HIV-1 Genome Packaging. *J. Mol. Biol* 410: 609-633
- Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Swanstrom R, Burch CL, Weeks KM (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460: 711-716.