



Vigilada Mineducación

TÍTULO

Modelación *in silico* de la región no codificante 5' del genoma del Virus de la Hepatitis C de aislados de pacientes colombianos.

Autor

Santiago Alvarez Cárdenas

Trabajo de Grado

Asesores

María Cristina Navas Navas

Rodrigo Ochoa Deossa

UNIVERSIDAD EAFIT
ESCUELA DE CIENCIAS
BIOLOGÍA
MEDELLÍN
2021

Modelación *in silico* de la región no codificante 5' del genoma del Virus de la Hepatitis C de aislados de pacientes colombianos

Santiago Alvarez-Cardenas^{1,2*}, Rodrigo Ochoa³, Maria-Cristina Navas²

1. Pregrado de Biología, Departamento de Ciencias Biológicas, Universidad EAFIT. Medellín, Colombia.
2. Grupo de Gastrohepatología, Facultad de Medicina, Universidad de Antioquia. Medellín, Colombia.
3. Grupo Tandem Max Planck en Biofísica de las Enfermedades Tropicales, Universidad de Antioquia. Medellín, Colombia.

INFO. ARTÍCULO

Palabras Clave:
Virus de Hepatitis C
IRES
Estructura ARN
Modelación in silico

RESUMEN

El sitio de entrada interno al ribosoma (IRES) del genoma del Virus de la Hepatitis C es una región estructurada ubicada en la región no codificante 5' del genoma. Diversas mutaciones (sustituciones, inserciones) causan anomalías en el plegamiento de la estructura secundaria y, por ende, terciaria del IRES. En este trabajo se discuten las características de las secuencias obtenidas de una cohorte de 7 individuos con infección por VHC, las mutaciones encontradas y las implicaciones de estas en la conformación estructural de los DII, DII del IRES.

1. Introducción

El virus de la hepatitis C (VHC) está clasificado en la Familia Flaviviridae, género Hepacivirus, especie Hepacivirus C (ICTV, 2021; Simmonds et al., 2017). El genoma es un ARN de cadena sencilla, sentido positivo, de aproximadamente 9.6 kb, codifica una poliproteína que se escinde en 10 proteínas, tres estructurales (Core, E2, E1) y siete no estructurales (P7, NS2, NS3, NS4A, NS4B, NS5A y NS5B) (Ashfaq et al., 2011; Li & Lo, 2015; Niepmann & Gerresheim, 2020; Scheel & Rice, 2013).

El genoma presenta en los extremos 5' y 3' regiones no codificantes (NCR, por sus siglas en inglés), que tienen un papel relevante en el reconocimiento del genoma por el ribosoma para la síntesis de la poliproteína (5'NCR) y por la polimerasa viral para la síntesis de nuevas copias del genoma viral (3'NCR) (Ashfaq et al., 2011; Li & Lo, 2015; Preciado et al., 2014). Luego de la adhesión, entrada y descapsidación de la partícula viral, se inicia la traducción del genoma viral gracias a la interacción del 5' NCR con el ribosoma de una manera independiente de la 7 metil-guanosina en el extremo 5' de los ARN mensajeros (CAP, por sus siglas en inglés) (Ashfaq et al., 2011, Université Claude Bernard Lyon 1, 2015).

El 5' NCR presenta una estructura secundaria conocida como sitio de entrada interno al ribosoma (IRES, por sus siglas en inglés) que tiene un papel crítico en la traducción del genoma viral. El IRES consta de cuatro dominios (I-IV) de los cuales el dominio II y III tienen la mayor relevancia en el acoplamiento al ribosoma (Lukavsky, 2009; Lytle et al., 2002; Toroney et al., 2010; van Leeuwen et al., 2004).

El dominio dI, (nt 1-43 en secuencia de referencia subgenotipo 1b), aunque no participa en la interacción con el ribosoma, su relevancia radica en albergar los sitios de unión del micro RNA celular 122 (miR-122) (nt 2-8 y nt 23-28) (Jopling et al., 2008). La unión del miR-122 al genoma viral permite una mayor estabilidad del ARN y mayor eficiencia en la traducción de la poliproteína (Niepmann & Gerresheim, 2020; Ono et al., 2020).

El dominio dII (nt 44-118) (Figura 1) presenta una estructura de hairpin con dos loops y un pseudo-loop y se organiza en un dominio apical (dIIb) y basal (dIIa) que interactúa con la unidad

ribosomal 40S (Beales et al., 2001; Lukavsky, 2009; Lytle et al., 2002; van Leeuwen et al., 2004); adicionalmente, dII es responsable de una reorganización en el ARN ribosomal (Rrna) 18S que facilita la unión del amino ácido Metionina para dar inicio a la traducción de la poliproteína (Malygin et al., 2013).

El dominio dIII (nt 118-306) es el más grande y consiste en varios tallo-bucle ramificados (dIIIabcdef) que interactúan en su mayoría con 40S (dIIIdef) y el factor de iniciación eIF3 (por sus siglas en inglés, Eukaryotic initiation factor 3) (dIIIabc) (Beales et al., 2001; Honda et al., 1996; Lytle et al., 2002).

El dominio dIV (306-360) presenta una estructura en tallo bucle hacia el final de la región no codificante que incluye el codón de inicio en la posición 342 (AUG) estando plegado, que también interactúa con 40S (Lukavsky, 2009; van Leeuwen et al., 2004).

La estructura del IRES ha sido bien caracterizada por métodos experimentales como Resonancia Magnética Nuclear (RMN) (Figura 1) y Cristalografía de Rayos-X, métodos que resultan costosos y logísticamente complejos (Beales et al., 2001). Sin embargo, diversos algoritmos computacionales han sido desarrollados con el fin de predecir la estructura secundaria y terciaria con determinada exactitud de secuencias de ARN.

Por lo anterior, el uso de estas herramientas computacionales favorece la exploración de las variaciones y dinámicas estructurales que allí tienen lugar, brindando información de cómo se favorece o perjudica la función del IRES del VHC por la presencia de mutaciones. El objetivo del presente trabajo es modelar la estructura secundaria (2D) y terciaria (3D) de siete secuencias de 5' NCR de aislados de VHC obtenidos de individuos colombianos para predecir el efecto de mutaciones en el plegamiento estructural para cada uno de los dominios del IRES que podrían ser relevantes en el acoplamiento con el ribosoma.

*Autor de correspondencia
e-mail: salvarezc@eafit.edu.co

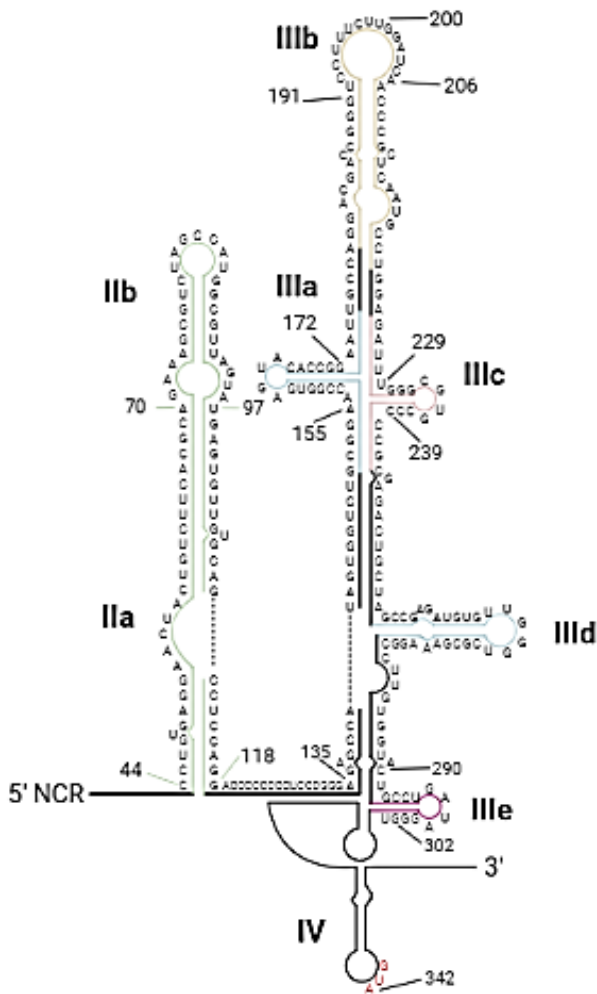


Figura 1. Estructura del IRES. Modelo de la estructura del IRES de VHC de la secuencia referencia D90208, adaptado de (Lukavsky, 2009; Pestova et al., 1998). Creado con BioRender.com

2. Materiales y Métodos

Secuencias y análisis

En el marco del proyecto “Prevalencia y modelación de la infección por el Virus de la Hepatitis C en Colombia” se obtuvieron siete (7) secuencias de VHC correspondientes a 6 del genotipo 1 de VHC, cuatro de subgenotipo 1b, dos del subgenotipo 1a y una muestra de genotipo 4, subgenotipo 4d, fueron incluidas para los análisis (1). Las secuencias corresponden a muestras de suero obtenidas de 4 individuos con historia transfusional antes de 1996, 2 personas que se inyectan y de un paciente con infección crónica y tratamiento previo con IFN.

En el marco del proyecto en mención, previamente se amplificó la región 5'NCR por medio de RT-PCR anidada según lo descrito por Chan et al. La secuencia de los amplicones se determinó mediante secuenciación con el método de Sanger (Macrogen, Korea).

Para la secuencia de referencia, luego de haber revisado las secuencias en publicaciones de Honda, 1996; Lukavsky, 2003; y

los modelos PDB, y tomando como referencia la publicación de van Leeuwen et al., 2004, se realiza un Blast nucleotídico entre las secuencias encontradas (GB: D28833, LADB: FN435993, EU1) encontrando que la secuencia que cumple un consenso y mayor puntaje de Blast para secuencia completa es D90208 (Resultado no mostrado).

Las secuencias fueron alineadas con el algoritmo L-INS-i en MAFFT (MAFFT v.7.453) (Kato & Standley, 2013) y analizadas individualmente para identificar las mutaciones según la secuencia referencia D90208 (VHC genotipo, subgenotipo 1b). Los modelos tridimensionales de 5'NCR obtenidos por Resonancia Magnética Nuclear a partir de un constructo sintético se utilizaron como referencia, disponibles en el PDB (<https://www.rcsb.org>) con los números de acceso: 1P5P (II), 1KH6 (IIIa-c), 1KP7 (IIIb), 1IDV (IIIc), 1F84 (IIId) y 1F85 (IIIe) (Lukavsky, 2009; Lukavsky et al., 2003).

Predicción de la estructura secundaria

Las estructuras secundarias del dominio II y los subdominios IIIa-c, IIIa, IIIb, IIIc, IIId, IIIe de las secuencias 5'NCR fueron calculadas usando UNAFOLD 4.0 (Markham & Zuker, 2008) y RNAFold 2.3 (Lorenz et al., 2011), obteniendo diferentes estructuras con su respectiva energía libre mínima (MFE, en inglés), la entropía posicional de cada nucleótido, el índice de paridad y la estructura secundaria óptima. Estos programas fueron escogidos debido a su reconocimiento en la materia, reportes en la literatura, soporte disponible, la practicidad y la facilidad de uso. Adicionalmente, RNAFold fue escogido como el modelador predilecto debido a que entregaba estructuras más acordes con la reportado en la literatura, cantidad adecuada de resultados y la facilidad de estos para conectarse luego con otros programas. La visualización de la estructura secundaria al igual que la identificación de mutaciones en el plegamiento se realizó con ayuda de la suite interactiva foRNA (Kerpedjiev et al., 2015) en <http://rna.tbi.univie.ac.at/foRNA/>.

Modelación de la estructura terciaria

Los modelos obtenidos en notación punto-corchete por RNAFold fueron usados para modelar la estructura terciaria haciendo uso del servidor RNAComposer (<http://rnacomposer.cs.put.poznan.pl>) (Antczak et al., 2016; Popenda et al., 2012) obteniendo las estructuras en formato PDB para luego ser evaluados en similitud con las referencias respectivas usando el método de RMSD en RNAssess (<http://rnassess.cs.put.poznan.pl>) (Lukasiak et al., 2013, 2015) y un análisis visual comparativo con las estructuras esperadas. Además, las estructuras fueron revisadas para identificar el efecto de las mutaciones en la conformación tridimensional de la estructura haciendo uso de Chimera 1.15rc (Pettersen et al., 2004).

3. Resultados

De las 7 secuencias de VHC obtenidas, S1-S4 corresponden a individuos con historial transfusional antes de 1996, S5 a un paciente con infección crónica por VHC y tratamiento previo con IFN y S6-S7 a personas que se inyectan.

ID	Factor de riesgo de la infección por VHC	Genotipo /Subgenotipo VHC*	Longitud de la Secuencia	Cobertura (secuencia referencia D90208) nt-nt	Mutaciones 5'NCR de VHC
S1	Historial transfusional antes de 1996	1b	280 nt	1-284	C184U, U204C, C205U
S2	Historial transfusional antes de 1996	1b	220 nt	63-282	A73
S3	Historial transfusional antes de 1996	4d	211 nt	92-302	C104U, G107A, U176C, U180C, C184U, C205A, U221C, G225A
S4	Historial transfusional antes de 1996	1b	222 nt	63-285	-
S5	Infección crónica por VHC y tratamiento previo con IFN	1b	221 nt	63-283	A73
S6	PID	1a	243 nt	63-306	-
S7	PID	1a	230 nt	73-302	C205A

Tabla 1. Características de las secuencias de VHC. Se detallan los factores de riesgo asociados a cada caso y las características de la secuencia correspondiente, la cobertura respecto a la referencia D90208 (GenBank) y las mutaciones encontradas. Únicamente la secuencia S1 posee cobertura completa del DII. * El genotipo y subgenotipo fue caracterizado previamente por López MC, 2021.

Si bien la estrategia de amplificación, identificación y secuenciación fue la misma para todas las muestras, la longitud de la secuencia no es idéntica entre las muestras (cebador HCV-211 nt 266 – 292 y cebador HCV-940 nt 41 – 60). En consecuencia, solo la secuencia S1 presenta una cobertura completa del DII (nt 44-118 en D90208), mientras que las secuencias S3, S6 y S7 son las únicas con cobertura completa del subdominio IIIe (nt 291-302) (Tabla 1).

El alineamiento de las 7 secuencias de 5'NCR de VHC correspondientes a los dominios II (nt 44-118), IIIa-c (nt 150-242), IIIa (nt 150-174), IIIb (179-220), IIIc (nt 224-242), IIId (nt 253-279) y IIIe (290-302) mostró un total de 10 sustituciones en tres secuencias, S1, S3, S7, y 1 inserción (A73) en las secuencias S2 y S5 (Tabla 1, Figura 2).

De las 10 sustituciones, 9 corresponden a transiciones (C104U, G107A, U176C, U180C, C184U, U204C, C205U, U221C, G225A) y 1 transversión (C205A). La mayoría de las transiciones (78%) y la transversión se identificaron en la secuencia S3 del genotipo 4 de VHC, subgenotipo 4d (Tabla 1). La secuencia S1 (subgenotipo 1b) presenta 3 mutaciones (C184U, U204C, C205U), mientras que la secuencia 7 (subgenotipo 1a) presenta la transversión (C205A) (Tabla 1). Las secuencias S4 y S6 no presentaron ningún tipo de mutación según el análisis de la secuencia disponible.

La estructura secundaria de cada uno de los dominios y subdominios fue modelada con RNAFold 2.3 y UNAFold 4.0 usando los algoritmos descritos previamente (Mathews et al., 2007; Zuker & Stiegler, 1981) con el fin de comparar los modelos de secuencias de VHC del estudio comparados con los predichos para la secuencia referencia y reportados en la literatura,

observando así los efectos de las mutaciones en el plegamiento y organización estructural. La estructura secundaria se visualiza con la probabilidad de unión de cada nucleótido como se ilustra en la Figura 3. Las mutaciones se encontraron principalmente en el Dominio II (30%) y Subdominio IIIb (60%) (Figura 1, 4).

La secuencia S3, al ser portadora de un alto número de mutaciones en comparación con las otras secuencias, propende variaciones estructurales altamente notorias y relevantes a nivel bidimensional (Figura 4) y tridimensional (Figura 5).

La predicción bidimensional muestra un apareamiento incorrecto en los subdominios IIIa y IIIb de S3 debido a las transiciones U176C, A180G, C184U, modificando la estructura de plegamiento según lo observado con la secuencia de referencia, debido a la formación de enlaces G-C, más estables que G-U y que en consecuencia le confieren el plegamiento modificado y una mayor estabilidad de la estructura comparada con la referencia (Figuras 4 y 5), evidenciada en un aumento de 2,7 kcal/mol en la MFE. Los resultados de los análisis de RMSD, MFE se muestran en la Tabla 2.

Adicionalmente, se realizó una comparación de los modelos reportados en el PDB con respecto a los modelos obtenidos por RNAComposer, usando como base el análisis del modelo 1P5P (DII) y su secuencia. Siguiendo la misma estrategia, se obtuvo la estructura secundaria usando RNAFold y llevando la estructura punto-corchete a RNAComposer. Como se observa en la Figura 6, las estructuras divergen fuertemente hacia los nt U64 y U103, obteniendo en consecuencia un RMSD 26.193 entre la estructura PDB y el modelo de RNAComposer. Luego, el RMSD de la secuencia S1 era de 0,743 con respecto a la secuencia referencia.

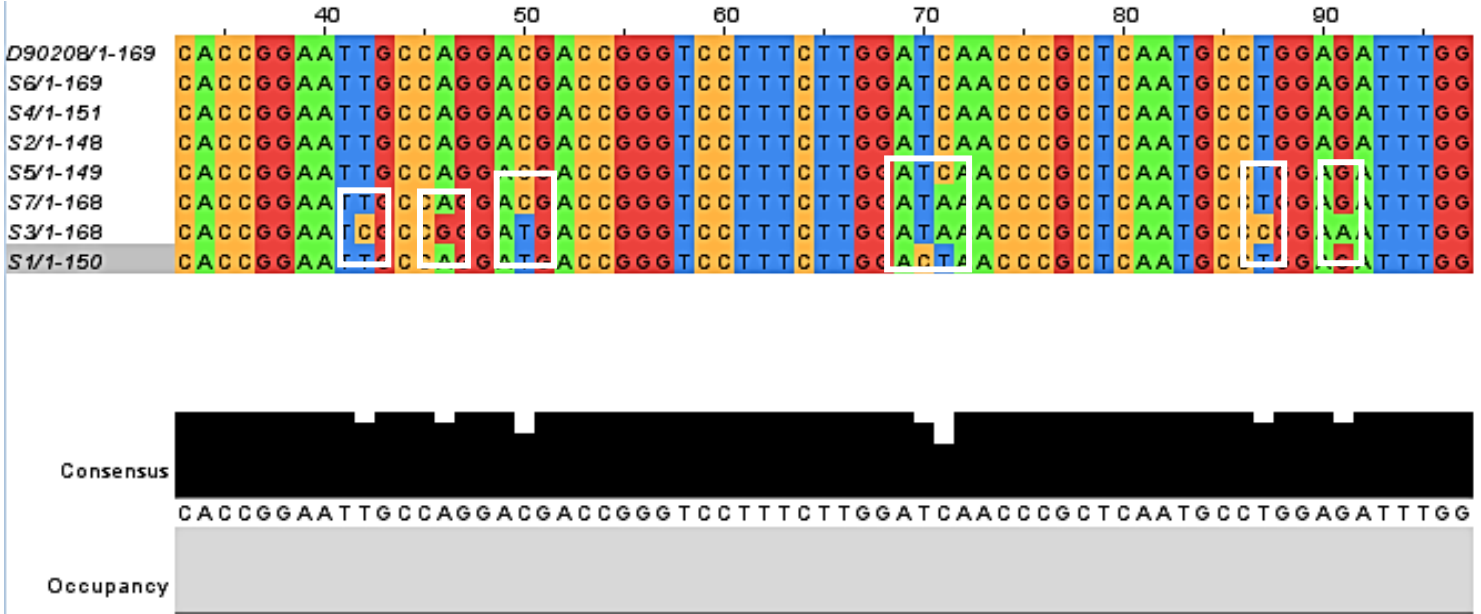


Figura 2. Alineamiento múltiple de la secuencia referencia D90208 de VHC y las secuencias S1-S7 usando MAFFT, visualizado en Jalview. Se resaltan en recuadros blancos las diferentes mutaciones encontradas en la región alineada del dIIIabc del IRES de VHC y se muestran las gráficas de consenso y ocupación de la región.

Para los demás modelos, 1KH6 (IIIa-c), 1KP7 (IIIb), 1IDV (IIIc) la correspondencia no era similar debido a las diferencias en tamaño y organización de las secuencias (RMSD > 30, resultados no mostrados), aunque en los segmentos alineados la conservación era del 100%. Por su parte, 1F84 (IIId) y 1F85 (IIIe) obtuvieron una correspondencia del 100% y valores de RMSD < 0,7 (resultados no mostrados).

Khan et al., 2020; Shi & Lai, 2006), probablemente se deba a la necesidad de conservar la estructura del IRES allí presente para el reconocimiento e interacción con proteínas celulares que participan en la traducción de la poliproteína (Moretti et al., 2010). La mayoría de las mutaciones se encontraron en la secuencia S3, como era de esperarse debido a ser un genotipo totalmente diferente al resto (Genotipo 4, subgenotipo 4d) y a la secuencia referencia (subgenotipo 1b). Estas mutaciones causaron un cambio conformacional mayor en modelo tridimensional debido a la mayor afinidad por formar enlaces G-C que, adicionalmente, le confieren una estabilidad estructural mayor al disminuir cada vez más sus valores de energía mínima libre (Tabla 2). Dicho fenómeno también se observó en cada sustitución que implicó el cambio de una Adenina o Uracilo por Citosina o Guanina (S1, en IIIb) (Tabla 2). No fue posible evaluar adecuadamente el efecto de la inserción A73 en las secuencias S2 y S5 debido a la imposibilidad de contar con una cobertura completa del DII para estas secuencias. Las sustituciones como C205A, que cambian nucleótidos de triple enlace (C/G) por nucleótidos de doble (U/A) no parecen tener un mayor efecto en la conformación estructural de las moléculas, como se evidencia para la S7 en IIIb (Tabla 2). Lo anterior podría ser debido a que favorecen precisamente la formación del *hairpin* en el subdominio IIIb en vez de favorecer la unión cruzada de los nucleótidos (Figura 4).

Figura 3. Ejemplo de estructura 2D del dII del IRES de VHC de la secuencia S1 obtenida por RNAFold. La escala de colores corresponde a la probabilidad de apareamiento entre las bases en cuestión (0-100%).

4. Discusión

La 5'NCR es una de las regiones más conservadas del genoma del VHC entre los diferentes genotipos (> 90% de homología) (S.

A propósito de las diferencias encontradas entre los modelos obtenidos con RNAComposer y los obtenidos del PDB, es de anotar que una multiplicidad de variables ambientales, tales como la concentración de iones a nivel celular, la misma estrategia usada para realizar la visualización con RMN y, por otro lado, el mismo algoritmo de predicción estructural pueden dar cuenta de las divergencias encontradas. Futuros análisis más extensos y complejos, en lo posible con un soporte experimental físico, serían necesarios para entender la dinámica que subyace a las variaciones estructurales evidenciadas.

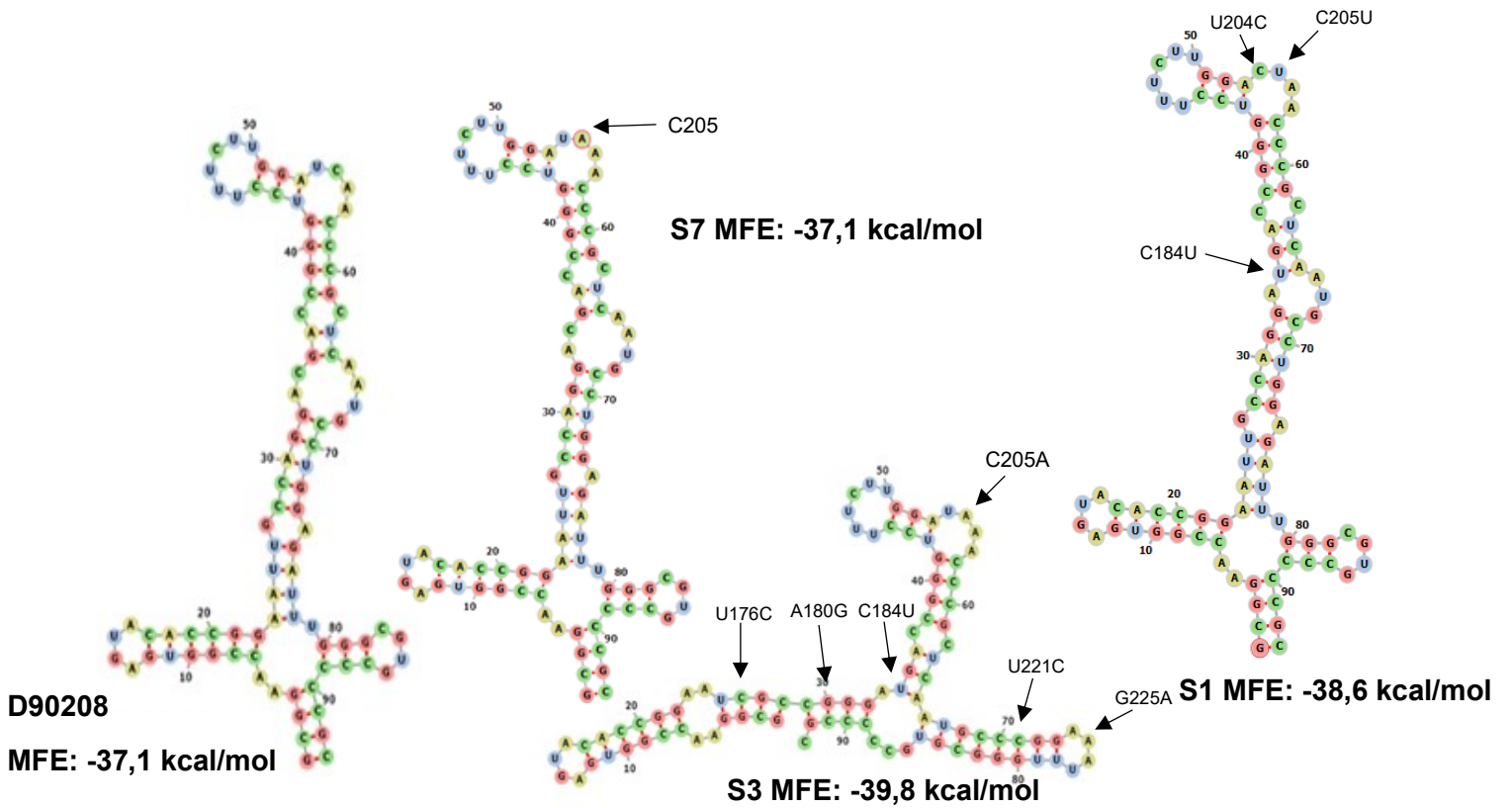
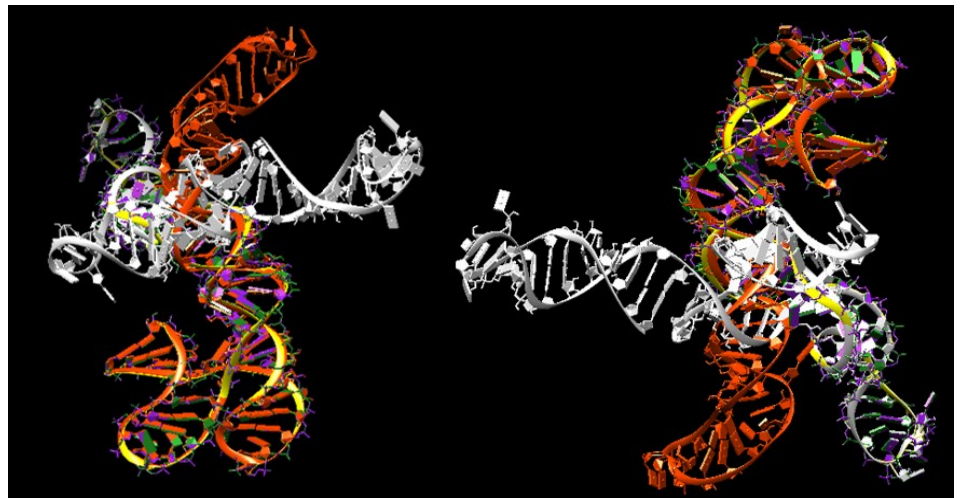


Figura 4. Comparación de la estructura 2D del dIIIabc del IRES de VHC de la secuencia referencia D90208 y las secuencias S7, S3, S1, visualizadas en foRNA. Las sustituciones se señalan con una flecha y se muestra la mínima energía libre de cada estructura.

Domino II			Domino IIIa-c			Domino IIIa			Domino IIIb			Domino IIIc			Domino IIId			Domino IIIe		
Secuencia	RMSD	MFE	Secuencia	RMSD	MFE	Secuencia	RMSD	MFE	Secuencia	RMSD	MFE	Secuencia	RMSD	MFE	Secuencia	RMSD	MFE	Secuencia	RMSD	MFE
S1 (1b)	0,743	-23,7	S1 (1b)	9,54	-38,6	S1 (1b)	0	-9,9	S1 (1b)	7,31	-12,5	S1 (1b)	0	-3,4	S1 (1b)	0	-5,8	S3 (4d)*	4,27	-0,9
S2 (1b)*	12,44	-13,3	S2 (1b)	0	-37,1	S2 (1b)	0	-9,9	S2 (1b)	0	-11	S2 (1b)	0	-3,4	S2 (1b)	0	-5,8	S6 (1a)	0	-2,7
S3 (4d)*	14,79	-2,6	S3 (4d)	21,03	-39,8	S3 (4d)	0	-9,9	S3 (4d)	6,43	-14,2	S3 (4d)	0,49	-3,4	S3 (4d)	0	-5,8	S7 (1a)*	4,27	-0,9
S4 (1b)*	13,32	-13	S4 (1b)	0	-37,1	S4 (1b)	0	-9,9	S4 (1b)	0	-11	S4 (1b)	0	-3,4	S4 (1b)	0	-5,8			
S5 (1b)*	12,44	-13,3	S5 (1b)	0	-37,1	S5 (1b)	0	-9,9	S5 (1b)	0	-11	S5 (1b)	0	-3,4	S5 (1b)	0	-5,8			
S6 (1a)*	13,32	-13	S6 (1a)	0	-37,1	S6 (1a)	0	-9,9	S6 (1a)	0	-11	S6 (1a)	0	-3,4	S6 (1a)	0	-5,8			
S7 (1a)*	18,89	-10,4	S7 (1a)	0,11	-37,1	S7 (1a)	0	-9,9	S7 (1a)	0,12	-11	S7 (1a)	0	-3,4	S7 (1a)	0	-5,8			

Tabla 2. Resultados en evaluación RMSD para cada dominio y subdominio del IRES de las secuencias S1-S7 comparado con D90208, usando RNAAssess. Se muestra la mínima energía libre (MFE) de cada estructura. *Secuencia con cobertura incompleta del dominio/subdominio.

Figura 5. Comparativa de la estructura terciaria del subdominio IIIabc del IRES de VHC de la secuencia referencia D90208 y las secuencias S1-S7. Se observa como la mayoría de las secuencias de 5'NCR de VHC (5/7) se alinean perfectamente, relativamente cercanas a la secuencia referencia (amarillo). Secuencia S1 (rojo) presenta ligera desviación, principalmente por la sustitución C184U. La secuencia S3 (blanco) es altamente divergente.



El dominio II mostró ser bastante conservado (>98%) entre las secuencias (o los fragmentos de éstas) evaluadas, sin embargo, es necesario un análisis más completo de las secuencias en vista de no contar con el dominio completo para 6 de las 7 secuencias. Cabe resaltar además que la secuenciación Sanger, mediante la cual se obtuvieron estas secuencias, muy probablemente entrega la cuasiespecie más representativa o dominante de las múltiples posibles presentes en la muestra que puede verse favorecida por las dinámicas biológicas que subyacen en cada paciente. Un estudio con técnicas de secuenciación de próxima generación (NGS, por sus siglas en inglés) como la SMRT de PacBio (Yamashita et al., 2020) sería una excelente aproximación para conocer las dinámicas de las cuasiespecies en muestras obtenidas en diferentes tiempos de la infección en un individuo y su relación directa con los cambios conformacionales en el IRES, otorgando más pistas sobre la dinámica estructural a nivel intracelular.

Finalmente, la NGS favorecería también la secuenciación completa de la 5' NCR, evitando el paso de la amplificación por PCR que además de tener una determinada tasa de error y eficiencia, restringe la posibilidad de explorar la diversidad genética de la región. De esta manera, se permitiría el estudio de la completitud de los dominios, así como la modelación completa del IRES y el estudio de todas las dinámicas conformacionales. Dichas ventajas podrán ser aplicadas en estudios posteriores permitiendo estudiar luego su acople (docking) con la subunidad 40S del ribosoma humano y los cambios conformacionales que allí tienen lugar y que, terminan por definir la importancia del docking con la replicación viral y en el fitness, abriendo la posibilidad de conocer más y descubrir nuevas características de la infección por VHC en humanos.

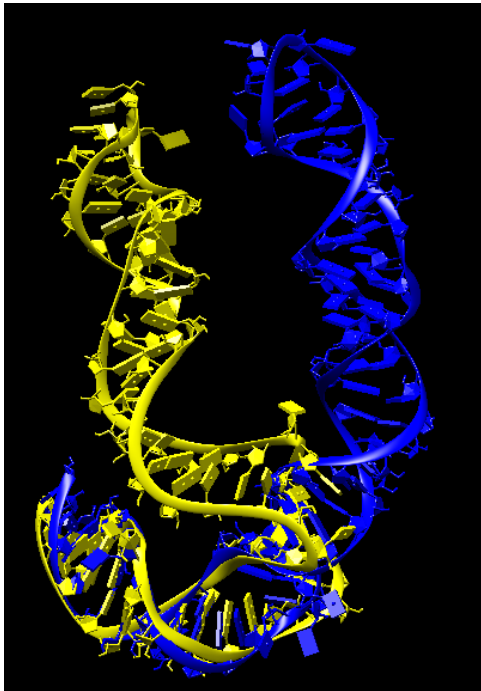


Figura 6. Estructura terciaria del dominio II del IRES de VHC. Secuencia reportada en modelo PDB (1P5P, azul) (Lukavsky, 2003) vs. Secuencia D90208 modelada en RNAComposer (amarillo). RMSD=26.193.

5. Conclusiones

Los estudios en modelación bi- y tridimensional de ARN y su dinámica estructural permiten la visualización de nuevas perspectivas sobre el papel que estas estructuras llevan a cabo en los sistemas biológicos y que resultan crípticas en muchos casos. Específicamente, el papel que desempeña el plegamiento de la región 5'-NCR en el VHC es fundamental pues, ante un incorrecto apareamiento, no se podría iniciar la traducción de las proteínas virales evitando el desarrollo de la enfermedad.

Las mutaciones C184U, C205A, C205U en loops, según la modelación realizada del dIIIabc del IRES del VHC no tienen efecto en la conformación estructural según se visualizaron en foRNA. En contraposición las mutaciones U176C, A180G en el dIIIabc favorecen la formación de enlaces entre nucleótidos complementarios ausentes en la secuencia de referencia, modificando la conformación estructural del IRES.

El análisis de la secuencia correspondiente al genotipo 4 de VHC, subgenotipo 4d, mostró un dIIIabc sustancialmente diferente. La mayoría de las mutaciones tuvieron lugar en dIIIabc, mientras que dIIId y dIIIE mostraron ser conservados para las 7 secuencias evaluadas. Un análisis que incluya un mayor número de secuencias de VHC, de diferentes genotipos, aportaría un análisis más robusto de este fenómeno.

Las herramientas de modelación computacional utilizadas poseen limitaciones en los modelos otorgados, por lo que suponen la necesidad de la refinación de los resultados obtenidos en aras de acercarse a la realidad biológica modelada. Por su parte, el mejoramiento y la modernización de las técnicas moleculares aplicadas otorgará las herramientas necesarias para desarrollar el ambiente de trabajo bioinformático ideal al brindar nueva y completa información, ayudando en la modelación *in silico*. Estudios próximos que conlleven análisis más profundos, completos y exigentes (en recursos y dedicación) nos darán pistas sobre los alcances de la modelación computacional en la representación biológica de los fenómenos. Entre estos, la secuenciación por NGS de las cuasiespecies presentes en una muestra nos permitiría conocer la dinámica de las mutaciones en cada hospedero en un momento dado y cómo estas pueden influir en la conformación del IRES.

Agradecimientos

Este trabajo se realizó en el marco del proyecto “Prevalencia y Modelación de la infección por VHC en Colombia”, financiado por MinCiencias código 1115-777-57692.

Bibliografía

Antczak, M., Popena, M., Zok, T., Sarzynska, J., Ratajczak, T., Tomczyk, K., Adamiak, R. W., & Szachniuk, M. (2016). New functionality of RNAComposer: An application to shape the axis of miR160 precursor structure. *Acta Biochimica Polonica*, 63(4), 737–744. https://doi.org/10.18388/abp.2016_1329

Ashfaq, U. A., Javed, T., Rehman, S., Nawaz, Z., & Riazuddin, S. (2011). An overview of HCV molecular biology, replication and

- immune responses. *Virology Journal*, 8, 1–10. <https://doi.org/10.1186/1743-422X-8-161>
- Beales, L. P., Rowlands, D. J., & Holzenburg, A. (2001). The internal ribosome entry site (IRES) of hepatitis C virus visualized by electron microscopy. *Rna*, 7(5), 661–670. <https://doi.org/10.1017/S1355838201001406>
- Chan, S. W., McOmish, F., Holmes, E. C., Dow, B., Peutherer, J. F., Follett, E., Yap, P. L., & Simmonds, P. (1992). Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants. *Journal of General Virology*, 73(5), 1131–1141. <https://doi.org/10.1099/0022-1317-73-5-1131>
- Honda, M., Brown, E. A., & Lemon, S. M. (1996). Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *Rna*, 2, 955–968.
- ICTV. (2021). Genus: Hepacivirus. ICTV Online Report. https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/flaviviridae/362/genus-hepacivirus
- Jopling, C. L., Schütz, S., & Sarnow, P. (2008). Position-Dependent Function for a Tandem MicroRNA miR-122-Binding Site Located in the Hepatitis C Virus RNA Genome. *Cell Host and Microbe*, 4(1), 77–85. <https://doi.org/10.1016/j.chom.2008.05.013>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kerpedjiev, P., Hammer, S., & Hofacker, I. L. (2015). Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics*, 31(20), 3377–3379. <https://doi.org/10.1093/bioinformatics/btv372>
- Li, H. C., & Lo, S. Y. (2015). Hepatitis C virus: Virology, diagnosis and treatment. *World Journal of Hepatology*, 7(10), 1377–1389. <https://doi.org/10.4254/wjh.v7.i10.1377>
- López Osorio, María Camila. Infección por el Virus de la Hepatitis C en pacientes con historia transfusional antes de 1996 en tres ciudades de Colombia. Trabajo de investigación de MSc, CCB, Universidad de Antioquia, 2021.
- Lorenz, R., Bernhart, S. H., Höner zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6(1). <https://doi.org/10.1186/1748-7188-6-26>
- Lukasiak, P., Antczak, M., Ratajczak, T., Bujnicki, J. M., Szachniuk, M., Adamiak, R. W., Popena, M., & Blazewicz, J. (2013). RNAnalyzer - Novel approach for quality analysis of RNA structural models. *Nucleic Acids Research*, 41(12), 5978–5990. <https://doi.org/10.1093/nar/gkt318>
- Lukasiak, P., Antczak, M., Ratajczak, T., Szachniuk, M., Popena, M., Adamiak, R. W., & Blazewicz, J. (2015). RNAssess - A Web server for quality assessment of RNA 3D structures. *Nucleic Acids Research*, 43(W1), W502–W506. <https://doi.org/10.1093/nar/gkv557>
- Lukavsky, P. J. (2009). Structure and function of HCV IRES domains. *Virus Research*, 139(2), 166–171. <https://doi.org/10.1016/j.virusres.2008.06.004>
- Lukavsky, P. J., Kim, I., Otto, G. A., & Puglisi, J. D. (2003). Structure of HCV IRES domain II determined by NMR. *Nature Structural Biology*, 10(12), 1033–1038. <https://doi.org/10.1038/nsb1004>
- Lytle, J. R., Wu, L., & Robertson, H. D. (2002). Domains on the hepatitis C virus internal ribosome entry site for 40s subunit binding. *Rna*, 8(8), 1045–1055. <https://doi.org/10.1017/S1355838202029965>
- Malygin, A. A., Kossinova, O. A., Shatsky, I. N., & Karpova, G. G. (2013). HCV IRES interacts with the 18S rRNA to activate the 40S ribosome for subsequent steps of translation initiation. *Nucleic Acids Research*, 41(18), 8706–8714. <https://doi.org/10.1093/nar/gkt632>
- Markham, N. R., & Zuker, M. (2008). UNAFOLD Software for Nucleic Acid Folding and Hybridization. In *Methods in molecular biology* (Vol. 453, Issue 1, pp. 1–31). <https://doi.org/10.1007/978-1-60327-429-6>
- Mathews, D. H., Turner, D. H., & Watson, R. M. (2007). RNA Secondary Structure Prediction. In *Curr Protoc Nucleic Acid Chem* (Vol. 1, Issue 1, pp. 1–29). <https://doi.org/10.1002/0471142700.nc1102s28.RNA>
- Niepmann, M., & Gerresheim, G. K. (2020). Hepatitis C virus translation regulation. *International Journal of Molecular Sciences*, 21(7). <https://doi.org/10.3390/ijms21072328>
- Ono, C., Fukuhara, T., Li, S., Wang, J., Sato, A., Izumi, T., Fauzyah, Y., Yamamoto, T., Morioka, Y., Dokholyan, N. V., Standley, D. M., & Matsuura, Y. (2020). Various miRNAs compensate the role of miR-122 on HCV replication. *PLoS Pathogens*, 16(6 June), 1–22. <https://doi.org/10.1371/journal.ppat.1008308>
- Pestova, T. V., Shatsky, I. N., Fletcher, S. P., Jackson, R. J., & Hellen, C. U. T. (1998). A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes and Development*, 12(1), 67–83. <https://doi.org/10.1101/gad.12.1.67>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Popena, M., Szachniuk, M., Antczak, M., Purzycka, K. J., Lukasiak, P., Bartol, N., Blazewicz, J., & Adamiak, R. W. (2012). Automated 3D structure composition for large RNAs. *Nucleic Acids Research*, 40(14), 1–12. <https://doi.org/10.1093/nar/gks339>
- Preciado, M. V., Valva, P., Escobar-Gutierrez, A., Rahal, P., Ruiz-Tovar, K., Yamasaki, L., Vazquez-Chacon, C., Martinez-Guarneros,

A., Carpio-Pedroza, J. C., Fonseca-Coronado, S., & Cruz-Rivera, M. (2014). Hepatitis C virus molecular evolution: Transmission, disease progression and antiviral therapy. *World Journal of Gastroenterology*, 20(43), 15992–16013. <https://doi.org/10.3748/wjg.v20.i43.15992>

Scheel, T. K., & Rice, C. M. (2013). Understanding the hepatitis C virus life cycle paves the way for highly effective therapies. *Nature Medicine*, 19(7), 837–849. <https://doi.org/10.1038/nm.3248>.Understanding

Simmonds, P., Becher, B., Bukh, J., Gould, E. A., Meyers, G., Monath, T., Muerhoff, S., Pletnev, A., Rico-Hesse, R., Smith, D. B., Stapleton, J. T., & ICTV Report Consortium. (2017). ICTV virus taxonomy profile: Flaviviridae. *Journal of General Virology*, 98, 2–3. <https://doi.org/10.1099/jgv.0.001143>

Toroney, R., Nallagatla, S. R., Boyer, J. A., Cameron, C. E., & Bevilacqua, P. C. (2010). Regulation of PKR by HCV IRES RNA: Importance of Domain II and NS5A. *Journal of Molecular Biology*, 400(3), 393–412. <https://doi.org/10.1016/j.jmb.2010.04.059>

Université Claude Bernard Lyon 1 [Université Claude Bernard Lyon 1]. (2015, 10, 15). Hepatitis C Virus Life Cycle [Archivo de video]. Recuperado de https://www.youtube.com/watch?v=fV-jhNQs_WE

van Leeuwen, H. C., Reusken, C. B. E. M., Roeten, M., Dalebout, T. J., Riezu-Boj, J. I., Ruiz, J., & Spaan, W. J. M. (2004). Evolution of naturally occurring 5' non-translated region variants of hepatitis C virus genotype 1b in selectable replicons. *Journal of General Virology*, 85(7), 1859–1866. <https://doi.org/10.1099/vir.0.79924-0>

Yamashita, T., Takeda, H., Takai, A., Arasawa, S., Nakamura, F., Mashimo, Y., Hozan, M., Ohtsuru, S., Seno, H., Ueda, Y., & Sekine, A. (2020). Single-molecular real-time deep sequencing reveals the dynamics of multi-drug resistant haplotypes and structural variations in the hepatitis C virus genome. *Scientific Reports*, 10(1), 1–12. <https://doi.org/10.1038/s41598-020-59397-2>

Zuker, M., & Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9(1), 133–148. <https://doi.org/10.1093/nar/9.1.133>