

DILUCIDACIÓN DEL MODO DE UNIÓN DE APTÁMEROS ANTI-CHIKUNGUNYA: BIOINFORMÁTICA ESTRUCTURAL E INTELIGENCIA ARTIFICIAL

Francisco M. de los Santos¹, Sara E. Gómez¹, Gonzalo Núñez¹, Belén Pintos¹, Ihara J. Suárez¹, Lizette N. Vela¹ y Pablo D. Dans^{2,3}

¹ Ciclo de Metodología Científica II 2024 - Facultad de Medicina - Universidad de la República, Centro Universitario Litoral Norte (CENUR LN), Salto, Uruguay.

² Grupo de Biofísica Computacional. Departamento de Ciencias Biológicas. CENUR LN. Universidad de la República, Salto, Uruguay.

³ Unidad de Bioinformática. Instituto Pasteur de Montevideo, Uruguay.

ÍNDICE GENERAL

1. INTRODUCCIÓN	1
2. OBJETIVOS	6
2.1. Objetivos generales	6
2.2. Objetivos específicos	6
3. MATERIALES Y MÉTODOS	7
3.1. Búsqueda de patrones en las secuencias de los aptámeros	8
3.2. Predicción de la estructura secundaria (2D)	8
3.3. Predicción de la estructura tridimensional (3D)	9
3.4. Predicción de la dinámica y flexibilidad (4D)	9
3.5. Predicción del modo de unión y proposición de mutantes	10
3.6. Análisis de los datos	11
4. RESULTADOS Y DISCUSIÓN	11
4.1. Patrones de secuencia y motivos secundarios	12
4.2. Estructura tridimensional, dinámica y flexibilidad	14
4.3. Hipótesis del modo de unión y proposición de mutantes	18
5. CONCLUSIONES Y PERSPECTIVAS.....	22
6. REFERENCIAS BIBLIOGRÁFICAS	23
7. ANEXO	25

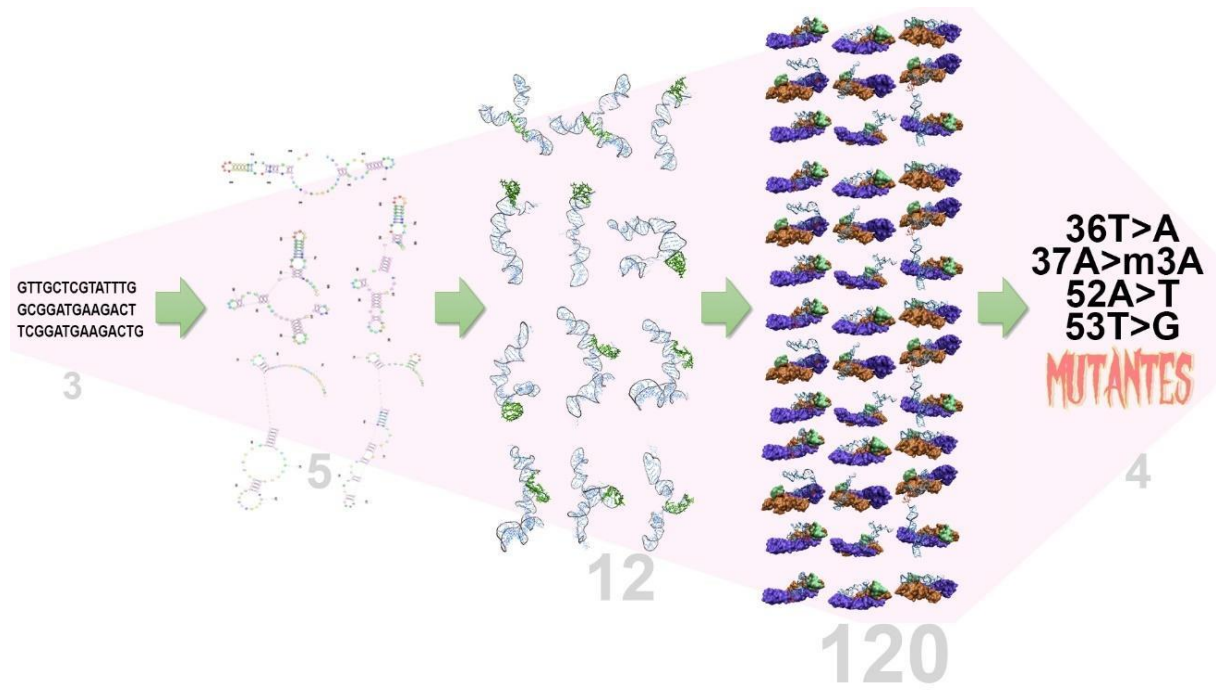
ÍNDICE DE TABLAS Y FIGURAS

Tabla I. Secuencia (1D) de los aptámeros determinados por SELEX.	7
Tabla II. Caracterización de la estructura 2D determinada con RNAstructure en términos de motivos secundarios y su transcripción a la notación vienna	14
Tabla III. Análisis de las proteínas de la cápside del virus en contacto con el aptámero y su orientación relativa	19
Tabla IV. Análisis de las interacciones moleculares entre el aptámero ApVLP11F_1c0p1 y las proteínas de la cápside E1-3.....	21
Tabla V. Análisis de las interacciones moleculares entre el aptámero ApVLP11F_2c0p1 y las proteínas de la cápside E1-3.....	25
Tabla VI. Análisis de las interacciones moleculares entre el aptámero ApVLP4R_2c0p2 y las proteínas de la cápside E1-3.....	26
Tabla VII. Análisis de las interacciones moleculares entre el aptámero ApVLP4R_1c0p4 y las proteínas de la cápside E1-3.....	27
Tabla VIII. Análisis de las interacciones moleculares entre el aptámero ApVLP4F_1c0p10 y las proteínas de la cápside E1-3.....	28
Figura 1. Mapa del Uruguay dónde se muestra un gradiente del índice de potencial transmisión de azul (menor riesgo) a rojo (mayor riesgo).....	1
Figura 2. Esquema de los pasos experimentales involucrados en una selección de aptámeros mediante el método SELEX.	2
Figura 3. Blanco molecular de la cápside de Chikungunya usados en la selección de aptámeros mediante el método SELEX.	3
Figura 4. Antecedentes de resultados experimentales obtenidos por el Hospital Ramón y Cajal sobre los tres aptámeros seleccionados contra la cápside de Chikungunya.....	4
Figura 5. Esquema de las etapas involucradas en la determinación estructural de los aptámeros y la generación de hipótesis de unión.	8
Figura 6. Resultado de la búsqueda de patrones en la secuencia de los tres aptámeros seleccionados usando el algoritmo del programa MEME.	12
Figura 7. Resultado de la selección de confórmeros hecha en base a las predicciones llevadas a cabo con el programa RNAstructure y el análisis detallado de los motivos secundarios.	13
Figura 8. Estructuras generadas con RNAcomposer en base a los motivos secundarios de la Tabla II.....	15
Figura 9. Análisis de las trayectorias de dinámica molecular de los 5 sistemas simulados en las condiciones del SELEX.....	17
Figura 10. Estructura de los 12 sub-estados conformacionales capturados de las dinámicas moleculares.	18
Figura 11. Representación de los 16 complejos seleccionados en los experimentos numéricos de docking molecular.....	20
Figura 12. Resultado de la interacción entre el aptámero ApVLP4R y las proteínas E1-3 obtenida usando la IA Alphafold 3 desarrollada por Google.....	22
Figura 13. Anaglifos de los 16 complejos seleccionados en los experimentos numéricos de docking molecular.....	29

LISTA DE ABREVIATURAS

2D	Dos dimensiones
3D	Tres dimensiones
4D	Cuatro dimensiones
ABSarbo	Estrategia basada en aptámeros contra arbovirus (<i>Aptamer Based Strategies against ARBOvirus</i>)
ADN	Ácido desoxirribonucleico
ADNasa	Desoxirribonucleasa
A-T	Par de base Adenina-Timina
ARN	Ácido ribonucleico
C-G	Par de base Citosina-Guanina
CHIK-V	Virus de Chikungunya
EULAC	Comunidad Europea-Latinoamérica
IA	Inteligencia Artificial
IRYSIS	Instituto Ramón y Cajal de Investigación Sanitaria
MEME	Herramienta de análisis de motivos en secuencias de ADN/ARN (Multiple Expectation Maximization for Motif Elicitation)
NTD	Enfermedad Tropical Desatendida (<i>Neglected Tropical Disease</i>)
PDB	Banco de Datos de Proteínas (<i>Protein Data Bank</i>)
RMN	Resonancia Magnética Nuclear
RMSD	Desvío cuadrático medio de las posiciones atómicas
SELEX	Evolución Sistemática de Ligandos por Enriquecimiento Exponencial
VLP	Partícula Similar a Virus (<i>Virus Like Particle</i>)
VMD	Visual Molecular Dynamics

RESUMEN GRÁFICO



RESUMEN

El virus de Chikungunya (CHIKV), fue identificado por primera vez en 1952 y, desde entonces, ha causado brotes significativos, incluyendo más de un millón de casos en América Latina y el Caribe en 2014. En Uruguay, los primeros casos autóctonos se reportaron en 2023, con un elevado potencial de transmisión en los departamentos del litoral norte, como Artigas, Salto y Paysandú. Si bien la tasa de mortalidad es baja, el riesgo aumenta en lactantes y personas mayores con comorbilidades. El diagnóstico temprano, durante la fase inicial de la infección cuando la carga viral es mínima, es fundamental para cortar los contagios, pero plantea a su vez un desafío particular para los países con recursos limitados.

Una de las estrategias para la detección del CHIKV es el uso de sondas moleculares, que incluyen anticuerpos monoclonales y aptámeros. Estos son pequeños oligonucleótidos de ADN/ARN de cadena simple seleccionados mediante el método SELEX, capaces de reconocer con alta especificidad y afinidad blancos moleculares. Estas moléculas presentan ventajas sobre los anticuerpos, como menores costos de producción y mayor versatilidad en aplicaciones diagnósticas y terapéuticas. Sus propiedades de unión únicas derivan de su estructura tridimensional y flexibilidad, lo que hace imprescindible su caracterización estructural para entender los mecanismos de unión. En ese marco, este trabajo se centró en la caracterización estructural de tres aptámeros de ADN diseñados para unirse con alta afinidad a las proteínas de la cápside del CHIKV, usando métodos de bioinformática estructural, simulaciones de dinámica molecular e inteligencia artificial. Del estudio de los sub-estados conformacionales visitados por los aptámeros y la generación de decenas de hipótesis de unión, se proponen una serie de mutaciones puntuales que permitirán racionalizar y optimizar el diseño de estas sondas ultra específicas para la detección del virus, con un costo accesible para los países más afectados.

Palabras claves: ADN. Cápside virus. Proteínas E1-3. Predicción estructural. Interacciones moleculares.

ABSTRACT

The Chikungunya virus (CHIKV), transmitted by *Aedes* mosquitoes, was first identified in 1952 and has since caused significant outbreaks, including more than one million cases in Latin America and the Caribbean in 2014. In Uruguay, the first locally transmitted cases were reported in 2023, with high transmission potential identified in northern departments such as Artigas, Salto, and Paysandú. Although the mortality rate is low, the risk increases substantially in infants and older adults with pre-existing health conditions. Early diagnosis during the initial stages of infection—when viral load is minimal—is critical for containing transmission but poses a particular challenge for resource-constrained countries.

One promising strategy for CHIKV detection involves the use of molecular probes, such as monoclonal antibodies and aptamers. Aptamers are small, single-stranded DNA or RNA oligonucleotides selected through the SELEX method (Systematic Evolution of Ligands by Exponential Enrichment). These molecules offer significant advantages over antibodies, including lower production costs and greater versatility in diagnostic and therapeutic applications. Their unique binding properties arise from their three-dimensional structure and inherent flexibility, making structural characterization essential for understanding their binding mechanisms. This study focused on the structural characterization of three DNA aptamers designed to bind with high affinity to CHIKV capsid proteins. Advanced computational methods, including structural bioinformatics, molecular dynamics simulations, and artificial intelligence, were employed to explore the conformational sub-states of these aptamers and generate dozens of binding hypotheses. Based on these analyses, a series of targeted mutations were proposed to refine and optimize the design of these ultra-specific probes for CHIKV detection. The resulting molecular tools aim to provide cost-effective solutions tailored to the needs of the most affected regions.

Keywords: DNA, Viral capsid, E1-3proteins, Structural prediction, Molecular interactions.

1. INTRODUCCIÓN

El Chikungunya es un virus transmitido por el mosquito del género *Aedes*, presente en Uruguay, que pasó de identificarse en Tanzania en 1952 a generar una epidemia con más de 1 millón de casos en Latinoamérica y el Caribe en 2014. En 2023 se reportaron más de 410.000 infectados. La tasa de mortalidad es baja (0,4%), aunque es mayor en niños menores de un año de edad (2,8%) y se incrementa en personas mayores que presentan comorbilidades. Uruguay reportó sus primeros casos a finales de abril de 2023, identificando 51 casos autóctonos y 19 casos importados. Como se muestra en la Figura 1, los valores más altos del índice de potencial transmisión se dan para los departamentos del litoral norte: Artigas, Salto y Paysandú, en ese orden (1).

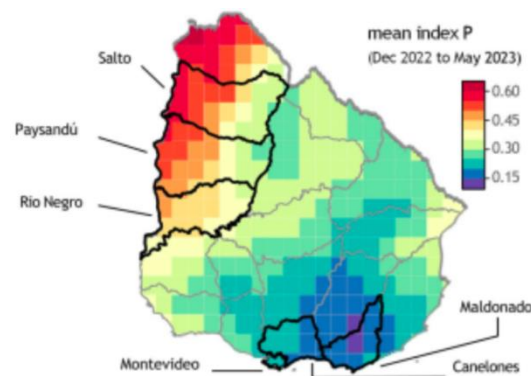


Figura 1. Mapa del Uruguay dónde se muestra un gradiente del índice de potencial transmisión de azul (menor riesgo) a rojo (mayor riesgo). Tomado de referencia (1).

Los síntomas de la enfermedad viral pueden confundirse con otras dolencias y por lo tanto es difícil de diagnosticar sin un análisis específico. Los mismos pueden incluir fiebre mayor a 38° C, dolor articular intenso, náuseas, erupciones rojas en la piel e incluso conjuntivitis. En la primera semana de la enfermedad, el virus se halla en la sangre y si un mosquito pica a la persona con Chikungunya, el mosquito puede infectarse y propagar la enfermedad en la familia y comunidad local. De este modo, durante el reposo que se suele prescribir, el paciente debe permanecer bajo un mosquitero. Esto resalta la importancia de poder diagnosticar al paciente en las primeras horas de la infección, cuando la carga viral en el cuerpo es mínima. Esto requiere del desarrollo de instrumentos de diagnóstico ultra específicos, algo que suele ser muy costoso para los países más afectados que son aquellos del trópico, América Latina y África (2). Es por ello que la enfermedad causada por el

virus de Chikungunya se considera una enfermedad tropical desatendida (NTD, de la sigla en inglés para *Neglected Tropical Disease*) (3).

La tecnología conocida como sonda molecular es una de las alternativas promisorias para la detección de trazas del virus. La misma suele estar compuesta por un anticuerpo monoclonal o un aptámero que son diseñados para unirse con alta afinidad y especificidad a algún blanco molecular del virus y contienen a su vez un fluoróforo que permite su detección con métodos de imagenología clásicos. Sin embargo, el desarrollo de anticuerpos monoclonales suele ser muy costoso, mientras que el costo de producir aptámeros es significativamente menor, lo que los convierte en una alternativa particularmente interesante para los principales países afectados.

Los aptámeros son pequeños oligonucleótidos simple hebra de ADN o ARN que reconocen blancos moleculares con alta afinidad y especificidad. Son obtenidos al exponer el blanco molecular de interés con una biblioteca de oligonucleótidos de secuencia variable, para posteriormente aislar y enriquecer los más afines en sucesivas rondas. El método se denomina SELEX (de las siglas en inglés, *Systematic Evolution of Ligand by EXponential enrichment*) y existen diferentes variantes según el propósito del aptámero que se desea seleccionar (4). En la Figura 2 se muestra un esquema con los pasos principales de un SELEX.

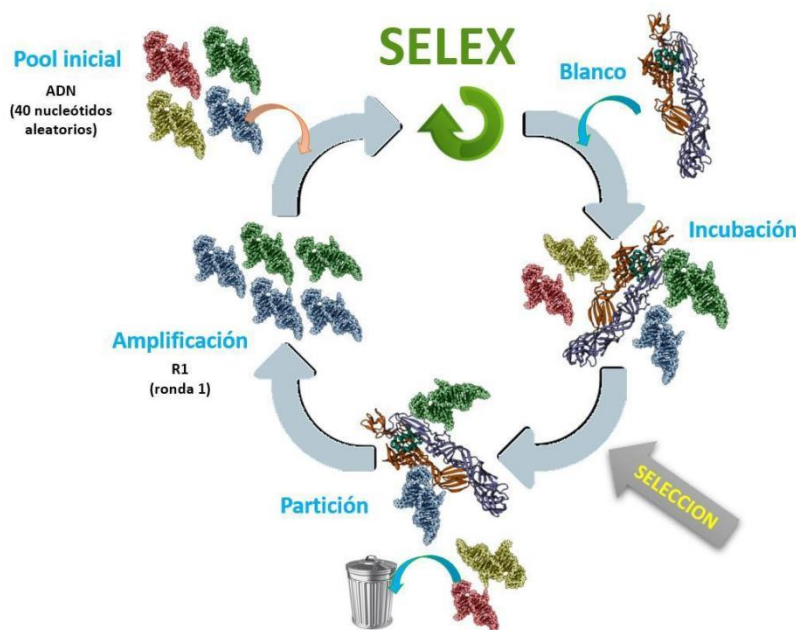


Figura 2. Esquema de los pasos experimentales involucrados en una selección de aptámeros mediante el método SELEX. Generalmente se hacen del orden de 10 rondas de forma de obtener aptámeros que se unen al blanco deseado a concentraciones nanomolares.

Actualmente, los aptámeros son explorados en novedosas y diversas aplicaciones como biofármacos que permiten desde el diagnóstico hasta la terapia. Sus propiedades de unión particulares se deben al plegamiento 3D que los mismos adoptan y a su flexibilidad, la que resulta clave para poder entender cómo es la interacción con el blanco molecular de interés.

En el marco de un proyecto EULAC (Comunidad Europea-Latinoamérica, 2022-2025) liderado por Uruguay (ABSarbo), el Instituto Ramón y Cajal de Investigación Sanitaria (IRYSIS, Madrid, España) seleccionó aptámeros de ADN contra un complejo proteico de la cápside del virus de Chikungunya. El complejo utilizado en los experimentos de SELEX constituye la unidad básica estructural de la cápside, conocido como partícula similar a virus, VLP (de la sigla en inglés para *Virus Like Particle*). Son complejos proteicos que imitan la estructura de un virus, pero no contienen material genético viral. En el caso de Chikungunya, la cápside está formada por un complejo de cuatro proteínas denominadas E1, E2, E3 y Core, que se encuentran resueltas experimentalmente y depositadas en el *Protein Data Bank* (PDB) a diferentes niveles de resolución (ver Figura 3). E1-3 son las proteínas de la cápside expuestas al medio exterior, mientras que la proteína Core se encuentra por debajo de la membrana plasmática, siendo las alfa hélices que unen E1-2 con Core las regiones transmembrana (ver estructura 8FCG en Figura 3).

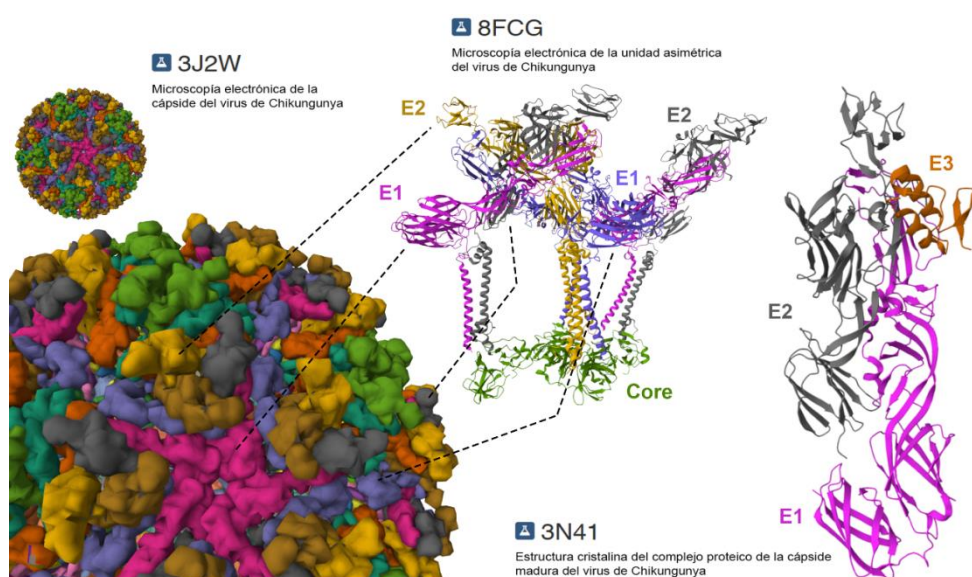


Figura 3. Blanco molecular de la cápside de Chikungunya usados en la selección de aptámeros mediante el método SELEX. Se reportan los códigos de acceso a las estructuras tomadas del *Protein Data Bank*.

La selección llevada a cabo en Madrid, arrojó tres secuencias denominadas ApVLP4F, ApVLP4R y ApVLP11F que mostraron unirse con alta afinidad (concentración nanomolar) y de manera concentración dependiente al blanco (ver Figura 4A). Se corroboró que los aptámeros hallados se unieran específicamente a las proteínas Core, E1 y E2 (ver Figura 4B), y se ensayó su estabilidad exponiéndolos a la DNAasa I (ver Figura 4C). Sin embargo, estos estudios no proporcionan ningún tipo de información sobre dónde y cómo se unen los aptámeros al complejo proteico, algo que requeriría de la caracterización estructural a nivel molecular.

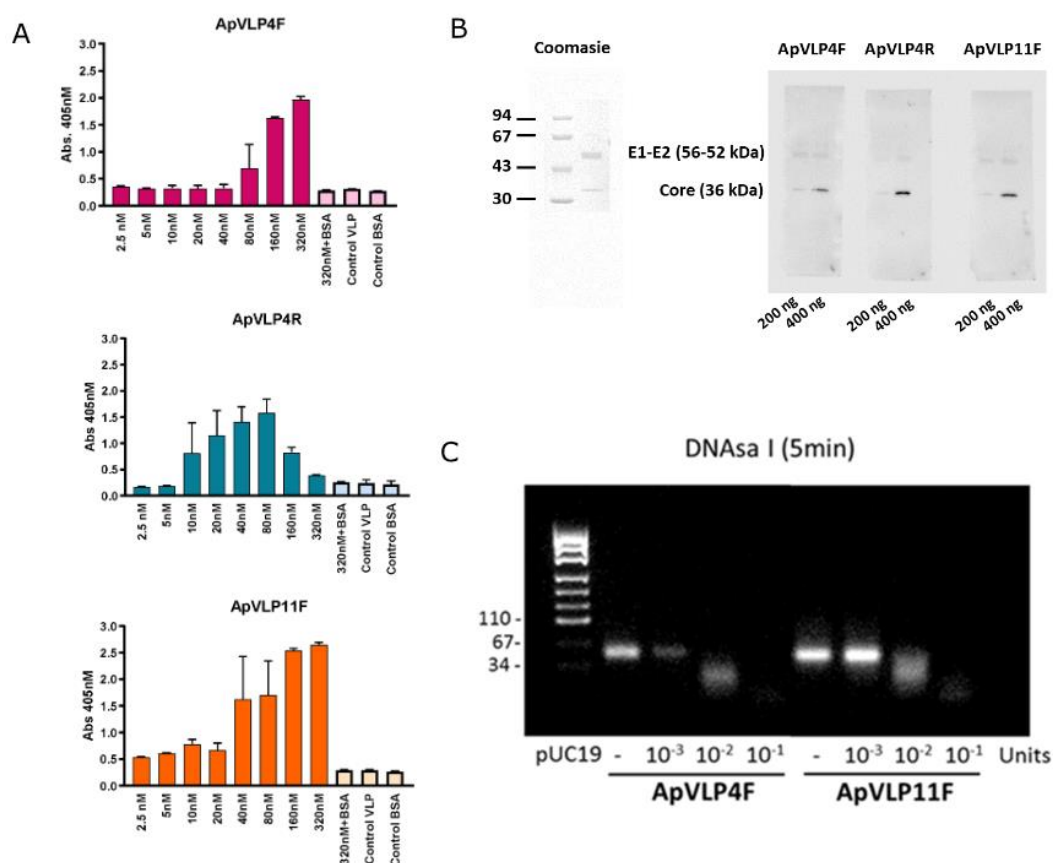


Figura 4. Antecedentes de resultados experimentales obtenidos por el Hospital Ramón y Cajal sobre los tres aptámeros seleccionados contra la cápside de Chikungunya. A) Estudio de afinidad concentración dependiente. B) Ensayos de especificidad de unión utilizando la técnica ELONA. C) Análisis de estabilidad usando ADNasa I.

Poder generar hipótesis sobre el modo de unión aptámero-blanco molecular es fundamental para poder hacer racionalmente un ajuste estructural fino del aptámero en base al detalle de las interacciones moleculares. Este conocimiento sobre la estructura molecular de los aptámeros y las interacciones que forma con el blanco puede obtenerse a través de métodos experimentales como la difracción de rayos-X usando cristales de los complejos o mediante estudios de Resonancia

Magnética Nuclear (RMN). Ambos enfoques requieren de equipos muy costosos y técnicos especializados, algo de difícil acceso para los países afectados por el virus de Chikungunya y que en todos los casos permiten determinar unos pocos complejos (lejos de las decenas de hipótesis de unión que deben investigarse por cada aptámero seleccionado). Una alternativa teórica y por lo tanto mucho más accesible tanto desde el punto de vista técnica como de costos lo representan los métodos de la bioinformática estructural, química computacional e inteligencia artificial (5, 6). Sin embargo, a la fecha, aún no es sencillo predecir con certeza la conformación 3D de motivos inusuales de ADN a partir de la secuencia. Los métodos más desarrollados y usados permiten llegar a la predicción de la estructura secundaria, lo que denominamos predicción 2D (7). La inmensa mayoría de los algoritmos para la predicción 2D fueron desarrollados para ARN y existen muy pocas herramientas para ADN. Un algoritmo recursivo de plegamiento 2D, que hace uso de datos termodinámicos generados para ADN, se ha implementado en un programa llamado RNAstructure (7). La gran limitación de estos métodos es que solo tienen en cuenta para su predicción los segmentos que logran aparearse de forma canónica a través de enlaces de hidrógeno de Watson-Crick, A-T y C-G para ADN. De este modo, ninguno de estos métodos 2D logra predecir con precisión los motivos secundarios no canónicos (horquillas, bucles, uniones, protuberancias, etc.) que son fundamentales a la hora de explicar la estructura 3D que pueden adoptar los aptámeros y que son cruciales para entender su modo de unión al blanco molecular.

Si bien la predicción de la estructura secundaria resulta útil para tener una aproximación de orden cero a la estabilidad de cierta secuencia de ADN y para buscar patrones entre la secuencia y los motivos secundarios, algunos autores han trabajado en métodos que permiten generar una predicción 3D inicial (8–10). Tal es el caso de RNAcomposer (8, 9), unas de las pocas alternativas para ADN, que se basa en fragmentos pequeños generados a partir de las estructuras determinadas experimentalmente y depositadas en el PDB (11). Las estructuras resultantes requieren siempre de ajustes que pueden obtenerse haciendo optimizaciones de geometría o simulaciones de dinámica molecular. Por lo tanto, una solución a la predicción de la estructura 3D de motivos inusuales de ADN y su flexibilidad (4D) a partir de la secuencia primaria solo puede encontrarse adaptando e integrando varios métodos de la bioinformática estructural, química

computacional y técnicas de simulación. La inteligencia artificial a través de los métodos de aprendizaje profundo está jugando un papel importante en esta área (5), dónde también cumple un rol los programas de *Docking* molecular (6).

2. OBJETIVOS

2.1. Objetivos generales

Dilucidar la estructura tridimensional, flexibilidad y el modo de unión de secuencias de aptámeros seleccionados mediante SELEX contra la cápside del virus de Chikungunya utilizando métodos de la bioinformática estructural, de la biofísica/química computacional e inteligencia artificial con la finalidad de generar hipótesis sobre el modo de unión aptámero-blanco molecular.

Generar pautas (mutantes), asistidas por las técnicas *in silico*, para el ajuste estructural fino de los aptámeros seleccionados con el propósito de usarlos como sondas moleculares ultra específicas, eficientes, y de bajo costo para diagnosticar la presencia viral.

2.2. Objetivos específicos

- i. Indagar sobre qué son los aptámeros, el método SELEX y las características generales del virus Chikungunya.
- ii. Conocer cuáles son las técnicas de visualización y predicción de la estructura secundaria y tridimensional específicas para modelar secuencias de ADN.
- iii. Incorporar las técnicas de dinámica molecular y el estudio de la flexibilidad en condiciones de laboratorio.
- iv. Llevar a cabo experimentos numéricos de *docking* molecular y técnica de inteligencia artificial que permitan caracterizar el modo de unión aptámero-proteínas de la cápside E1-3.
- v. Proponer mutaciones para aumentar o disrumpir la unión aptámero-Proteínas E1-3, que serán probadas en el laboratorio del Hospital Ramón y Cajal de Madrid, que puedan confirmar o rechazar las hipótesis de unión generadas en este trabajo con las técnicas computacionales.

3. MATERIALES Y MÉTODOS

Para este proyecto de investigación original se partió de las secuencias de nucleótidos de tres aptámeros seleccionados mediante SELEX y cedidos al grupo de Biofísica Computacional por el Hospital Ramón y Cajal de Madrid (ambos partes del consorcio ABSabro). Los mismos fueron etiquetados como ApVLP4R, ApVLP4F y ApVLP11F (ver Tabla I). Por otro lado, se contaba con la estructura 3D de las proteínas que conforman la cápside del virus de Chikungunya, a varios niveles de resolución, reportadas experimentalmente (rayos-X y microscopía electrónica de baja temperatura) y disponibles en el PDB bajo los códigos: 3J2W, 8FCG y 3N41 (ver Figura 3).

Tabla I. Secuencia (1D) de los aptámeros determinados por SELEX.

Etiqueta	Secuencia
ApVLP4R	GTTGCTCGTATTTAGGGCACAAAGGTAACGCTCACCAAAGCAACCGTTTCA TGCGACCGACACCAGTCTTCATCCGC
ApVLP4F	GCGGATGAAGACTGGTGTGCGGTGCGCATGAAACGGTTGCTTTTGGTGAGCGT TACCTTTGTGCCCTAAATACGAGCAAC
ApVLP11F	TCGGATGAAGACTGGTGTGGCGGTGAGTGGTCATGTAAGCATTCTGATTGT GGTTGGGGATGTGCCCTAAATACGAGCAAC

Para hacerlos experimentos de unión entre los aptámeros y el complejo proteico de la cápside del virus, fue necesario caracterizar la estructura 3D de los aptámeros a través de una serie de experimentos numéricos que se llevaron a cabo en cuatro grandes etapas que se esquematizan en la Figura 5. A partir de las secuencias de los tres aptámeros obtenidos por secuenciación, se buscó patrones en las mismas y se llevó a cabo la predicción de la estructura secundaria (Etapa 1), la predicción de la estructura 3D (Etapa 2), el estudio de la flexibilidad y los sub-estados conformacionales (Etapa 3) y finalmente se pasó a predecir la unión aptámero-complejo E1-3 (Etapa 4), de modo de poder generar hipótesis sobre el modo de unión.

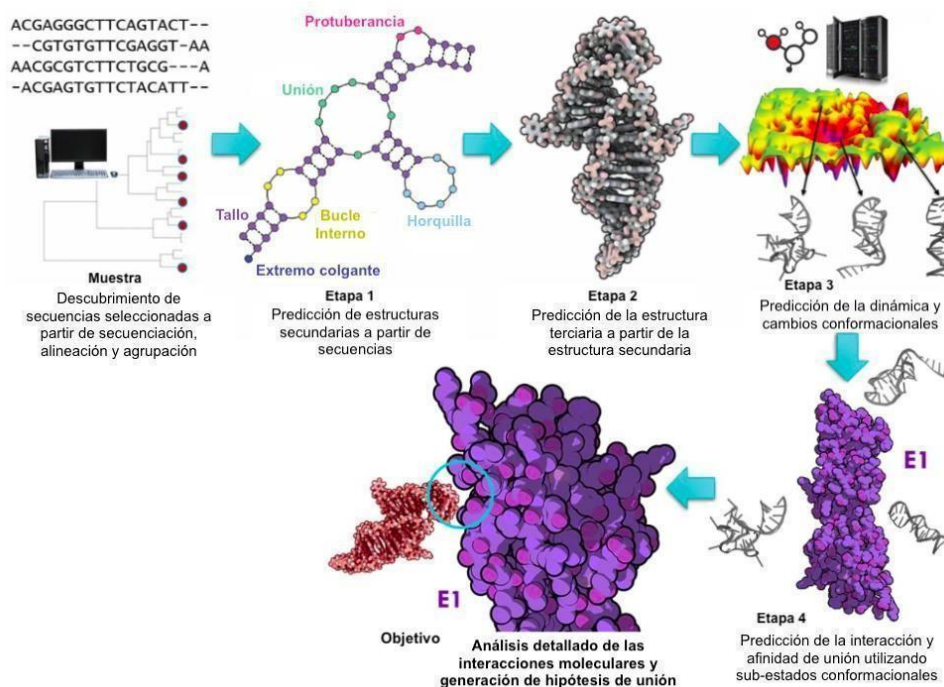


Figura 5. Esquema de las etapas involucradas en la determinación estructural de los aptámeros y la generación de hipótesis de unión.

3.1. Búsqueda de patrones en las secuencias de los aptámeros

Se llevó a cabo una búsqueda de patrones entre las secuencias de los tres aptámeros seleccionados utilizando el algoritmo proporcionado por el programa MEME (12). El mismo se encuentra optimizado para secuencias cortas, sin cortes (*gaps*) como ocurre con los aptámeros. Se usó el programa en el modo clásico, indicando que se trataba de secuencias de ADN y se le pidió al algoritmo que encontrara un máximo de 3 patrones comunes. Como los aptámeros comparten los primeros y últimos 18 nucleótidos que se usan como cebador (*primer* en inglés) para la etapa de amplificación durante el SELEX, eso deja la posibilidad de encontrar un solo patrón de secuencia común entre los tres aptámeros analizados en los 40 nucleótidos centrales.

3.2. Predicción de la estructura secundaria (2D)

Como las secuencias de los aptámeros seleccionados (Tabla I) no mostraron regiones con varias guaninas consecutivas, descartando la presencia de una estructura llamada G-cuádruplex, se utilizó el algoritmo de predicción de la estructura secundaria llamado RNAstructure (13). El mismo se basa en datos termodinámicos experimentales determinados sobre secuencias de ADN doble hebra canónicas, la aproximación de siguientes vecinos (en inglés *nearest-neighbors*) (14) y un algoritmo recursivo que busca todos los posibles modos de

apareamiento de Watson-Crick. La determinación se hizo de dos formas: i) Fijando el algoritmo como si las secuencias fueran de ARN a 333 K, según se ha recomendado en una revisión reciente (7); y ii) Considerando que se trataba de ADN a 310 K. Los demás parámetros se fijaron con los valores por defecto que proporciona el servidor (1). Al buscar las regiones de los aptámeros que pueden aparearse según Watson-Crick (tallos), los nucleótidos que no se organizan en tallos forman las estructuras secundarias llamadas horquillas, uniones, bucles internos, protuberancias y extremos colgantes (ver Etapa 1 en Figura 5).

3.3. Predicción de la estructura tridimensional (3D)

Para obtener una estructura inicial tridimensional de los aptámeros seleccionados, se usó la predicción 2D generada con el programa RNAstructure. Para eso, se transcribieron los motivos secundarios predichos a la notación vienna de puntos y paréntesis para bases desapareadas y apareadas respectivamente (15). La secuencia y la notación vienna son las únicas entradas que necesita el programa RNAcomposer para arrojar una estructura tridimensional inicial (2). Su concepto se basa en un sistema de traducción automática similar al empleado en el campo de la lingüística computacional. El método aprovecha la base de datos de RNA conocida como FRABASE, adaptada al diccionario que relaciona la estructura secundaria del RNA con los elementos de la estructura terciaria 3D (ver Etapa 2 en Figura 5).

El algoritmo del programa RNAcomposer sólo sirve para estructuras de ARN y por lo tanto las estructuras 3D predichas contienen uracilo en lugar de timina y los azúcares son ribosas en lugar de desoxirribosas presentes en el ADN. De este modo, los archivos en formato PDB que devuelve el servidor deben editarse para convertir las estructuras de nuevo a ADN. Los archivos PDB se editan con un editor de texto y se borra el hidrógeno unido al N3 de los uracilos, y los átomos O2' y HO2' de todas las ribosas. El grupo metilo unido al N3 de las timinas y el átomo de hidrógeno H2' de las desoxirribosas se agrega automáticamente usando el programa tleap del paquete AMBER 24 (3).

3.4. Predicción de la dinámica y flexibilidad (4D)

Las estructuras 3D iniciales determinadas con RNAcomposer se usaron como punto de partida para llevar a cabo simulaciones de dinámica molecular de modo

de relajar las interacciones moleculares teniendo en cuenta las condiciones de temperatura (25°C), presión (1 atm) y *buffer* (0.15 M de K⁺Cl⁻) usadas en el SELEX. Además de optimizar las interacciones intramoleculares del aptámero e intermoleculares con el solvente, estas simulaciones permitieron muestrear el espacio conformacional arrojando los sub-estados estructurales que los aptámeros pueden adoptar a lo largo del tiempo (ver Etapa 3 en la Figura 5). Se hicieron simulaciones de producción atomísticas en la escala del microsegundo (1 a 2 μ s según el aptámero) usando el campo de fuerza PARMBSC1 para representar al ADN (18), el modelo OPC para el agua (19) y los parámetros de Li y Merz para los iones K⁺ y Cl⁻ (20). Se utilizaron la suite de programas de dinámica molecular AMBER 24 (3) y los servidores de cálculo del grupo de Biofísica Computacional.

Las trayectorias de dinámica molecular obtenidas fueron procesadas usando el programa ccpptraj del paquete AMBER 24. El mismo permitió calcular el desvío cuadrático medio de la posición de todos los átomos a lo largo del tiempo de simulación (RMSD), los factores de temperatura de cada nucleótido que dan idea de la flexibilidad local y determinar los sub-estados conformacionales visitados durante la simulación haciendo un *k-cluster* y fijando *k*=5 (esto le pide al algoritmo que genere 5 agrupaciones en base al RMSD). Los centroides de dichas agrupaciones fueron tomados como representativos de 5 sub-estados conformacionales que adoptaron los aptámeros durante la simulación.

3.5. Predicción del modo de unión y proposición de mutantes

Del análisis de la dinámica y flexibilidad de los aptámeros obtenidos en la etapa anterior, se eligieron 12 sub-estados conformacionales para enfrentar al blanco molecular (código PDB 3N41, ver Figura 3) en experimentos de *docking* (ver Etapa 4 en la Figura 5). Como la estructura tomada del PDB proviene de un experimento de difracción de rayos-X, tuvimos que agregar los átomos de hidrógeno que no están determinados con esa técnica. Lo cual se realizó con el programa PDB2PQR (21), fijando el pH en 7 y el formato de salida en AMBER. Los ensayos de unión se llevaron a cabo usando el programa HDock con las opciones por defecto, pero sesgando los residuos activos en los aptámeros al patrón de secuencia común predicho por MEME. Se calcularon 10 poses por cada sub-estado conformacional generando 120 hipótesis de unión diferentes.

Para predecir la interacción entre nuestros aptámeros y las proteínas E1-3, también se utilizó el servidor de la inteligencia artificial (IA) llamada Alphafold3 (5) desarrollada por Google (22). Esta IA solo requiere las secuencias de los aptámeros y de las proteínas E1, E2 y E3 que fueron bajadas en formato fasta de la estructura 3N41 desde el PDB.

Del análisis de los 120 complejos aptámero-E1-3, se seleccionó 50 estructuras correspondientes a 10 poses de los 5 sub-estados mayoritarios (etiquetados como ApVLP4F_1c0, ApVLP4R_1c0, ApVLP4R_2c0, ApVLP11F_1c0 y ApVLP11F_2c0). De esas 50 estructuras, se eligió aquellas donde la interacción del aptámero E1-3 se diera por el lado de la superficie expuesta al exterior de la cápside, haciendo una inspección visual de cada complejo. Esto dejó 16 estructuras finales que fueron analizadas a nivel atómico, contacto por contacto, de modo de clasificar las interacciones moleculares en: i) Específicas, aquellas que se dan a una distancia menor a 3.5 Å entre un átomo de la base nitrogenada y un residuo de proteína; ii) Inespecíficas, aquellas que se dan a una distancia menor a 3.5 Å entre los átomos que conforman el *backbone* del ADN (grupo fosfato y desoxirribosa) y un residuo de proteína; y iii) Artefactos, cuando en una posible interacción está involucrando al menos un átomo de Carbono -que no puede formar enlace de hidrógeno- a una distancia menor a 3.5 Å, sea este átomo del aptámero o de la proteína.

3.6. Análisis de los datos

Para llevar a cabo los experimentos numéricos, en todas las etapas, se utilizaron estructuras, programas y algoritmos, todos de acceso libre. Para el procesamiento de los datos generados se operó sobre una estación de trabajo con Linux (Ubuntu 22.04), con el editor de texto atom (23), el programa de edición de imágenes GIMP 2 (24) y el paquete LibreOffice 24 (25). Los datos se graficaron usando grace (26) y gnuplot (27) y las representaciones moleculares se llevaron a cabo con Mol* (28) y VMD 1.9.7 (29).

4. RESULTADOS Y DISCUSIÓN

Al tratarse este de un trabajo de investigación longitudinal, el mismo se organizó en 4 etapas principales que permitieron ir de la secuencia de los aptámeros a generar varias hipótesis de unión con las proteínas de la cápside del virus de

Chikungunya. En el marco de la estrategia utilizada, las etapas debieron llevarse de forma sucesiva, siendo los resultados de una etapa las entradas necesarias para la siguiente (ver Figura 5). De este modo, resulta más natural y claro presentar los resultados junto a su discusión de forma de ir entendiendo las diferentes decisiones y caminos que seguimos.

4.1. Patrones de secuencia y motivos secundarios

La búsqueda de patrones en la secuencia permite hacer una primera suposición sobre el modo de unión que resulta crucial para el desarrollo del resto del trabajo. Si se encuentran motivos de secuencia común entre los aptámeros seleccionados que provienen del mismo experimento SELEX, se podría estar frente a una competencia entre las tres secuencias más representadas en el experimento. Es decir, que los aptámeros ApVLP4F, ApVLP4R y ApVLP11F se estén uniendo al complejo de proteínas de la cápside del virus E1-3 y Core a través del motivo de secuencia común. Para detectar la presencia de motivos se utilizó el algoritmo del programa MEME, pidiéndole que busque un máximo de 3 motivos (ver Materiales y métodos). Como los aptámeros comparten los primeros y últimos 18 nucleótidos, esto deja la detección de un motivo más en la región de 40 nucleótidos centrales.

Como se puede ver en la Figura 6, se encontró un motivo común entre los tres aptámeros con la secuencia CATGWAA (en el caso de los aptámeros *forward* "F", ApVLP4F y ApVLP11F) o TTWCATG (para el aptámero *reverse* "R", ApVLP4R), donde W representa una Adenina o Timina. *Forward* significa que la secuencia de ADN se lee de 5' a 3' y *reverse* de 3' a 5'. Si bien el motivo encontrado en la región central (ver rectángulo verde en la Figura 6) es detectado, no tiene la significancia estadística que tienen los cebadores en la confianza. En base a este resultado, se analizó con particular interés la región 25-31 en ApVLP4F, 48-54 en ApVLP4R y 32-38 en ApVLP11F. A partir de aquí, se asumió que los tres aptámeros se unieron a las proteínas de las cápsides a través de un motivo de secuencia común.

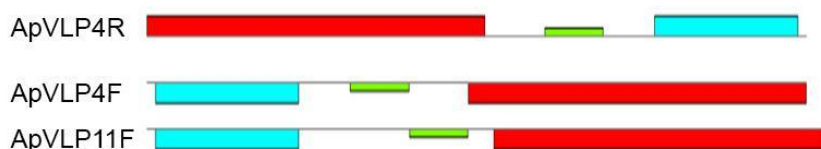


Figura 6. Resultado de la búsqueda de patrones en la secuencia de los tres aptámeros seleccionados usando el algoritmo del programa MEME. Se encontraron 3 motivos esquematizados con rectángulos color rojo, verde y cian.

A nivel molecular, se puede asumir que un mismo motivo de secuencia adopte una estructura tridimensional también común bajo el paradigma de que la secuencia determina la estructura. Un primer paso en la dilucidación de la conformación que pueda adoptar una molécula simple-hebra de ADN como los aptámeros, es intentar predecir qué segmentos son plausibles de unirse a través de enlaces de Watson-Crick en segmentos doble-hebra canónicos. La estabilidad de los aptámeros tanto térmica como frente a DNAasas dependerá directamente de la cantidad de estos segmentos canónicos y de lo rico que son en pares G-C (que son más estables que el par A-T). En ese entendido, se utilizó el algoritmo recursivo implementado en el programa RNAstructure para predecir los segmentos canónicos presentes en nuestros tres aptámeros. La determinación de los segmentos doble-hebra canónicos (también llamados tallos) deja a los demás nucleótidos formando los motivos de estructura secundaria denominados bucles internos, protuberancias, horquillas, uniones y extremos colgantes (ver Etapa 1 en Figura 5).

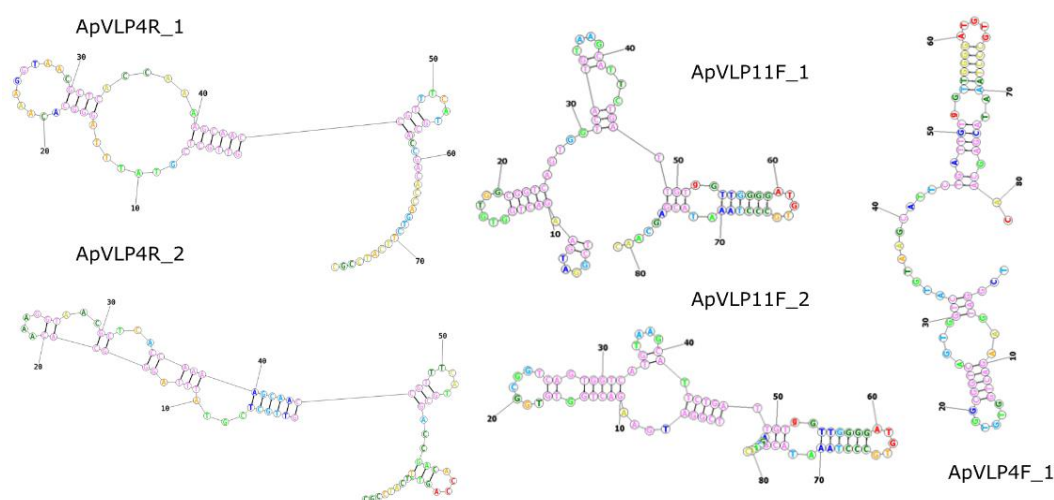


Figura 7. Resultado de la selección de confórmeros hecha en base a las predicciones llevadas a cabo con el programa RNAstructure y el análisis detallado de los motivos secundarios.

La representación gráfica de los confórmeros que se seleccionaron (cinco) se muestran en la Figura 7. Se realizaron varias predicciones cambiando las opciones del algoritmo (ver Materiales y métodos) y se eligió 5 soluciones: dos para el aptámero 4R, dos para el 11F y una para el 4F, etiquetados como ApVLP4R_1, ApVLP4R_2, ApVLP11F_1, ApVLP11F_2 y ApVLP4F_1 respectivamente. La selección de estos 5 confórmeros se realizó en base a un análisis detallado de los motivos secundarios obtenidos, favoreciendo las estructuras que mostraron la mayor cantidad de tallos. Ese análisis se muestra en la Tabla II, junto al código vienna que se determinó para cada caso.

Tabla II. Caracterización de la estructura 2D determinada con RNAstructure en términos de motivos secundarios y su transcripción a la notación vienna.

Nombre	Opciones	Análisis de motivos	Código vienna
ApVLP4R_1	ARN 333	3 tallos con 13 bases apareadas 2 horquillas 1 bucle interno 1 extremo colgante	(((((.....((((.....)))).....)))) (((.....))).....
ApVLP4R_2	ADN 310	7 tallos con 21 bases apareadas 3 horquillas 4 protuberancias 1 extremo colgante	(((((.....((((.....))....))....))))))(((.....)))...(((.....))).....
ApVLP11F_1	ARN 333	6 tallos con 21 bases apareadas 1 unión 4 horquillas 1 protuberancia 1 bucle interno 1 extremo colgante	...(((.....((((.....)))).....)).....((..(((.....((((.....)))).....)))..))..
ApVLP11F_2	ARN 333	6 tallos con 23 bases apareadas 3 horquillas 3 bucles internos 1 extremo colgante	(((((.....((((.....)))).....))....)))))..(((.....((((.....)))).....)).....
ApVLP4F_1	ADN 310	5 tallos con 20 bases apareadas 3 horquillas 1 bucle interno 1 unión 1 extremo colgante	(((.....((((.....)))).....)).....(((.....((((.....)))...((.....))).....))

4.2. Estructura tridimensional, dinámica y flexibilidad

La secuencia de los aptámeros y el código vienna de los motivos secundarios predichos en la etapa anterior fueron usados para generar una primera estructura tridimensional de los 5 confórmeros. No existen utilidades que permitan generar una predicción fiable para moléculas de ADN, así que se usó una herramienta desarrollada para ARN llamada RNAcomposer. Las estructuras generadas en formato PDB con esta herramienta, contienen uracilos en lugar de timinas y ribosas en la conformación C3'-endo (canónica del A-ARN) en lugar de desoxirribosas en la conformación C2'-endo (canónica del B-ADN). Las soluciones obtenidas requieren entonces una edición (ver Materiales y métodos) y sólo pueden ser consideradas como una predicción primaria y aproximada de la estructura 3D de los aptámeros de ADN (Figura 8).

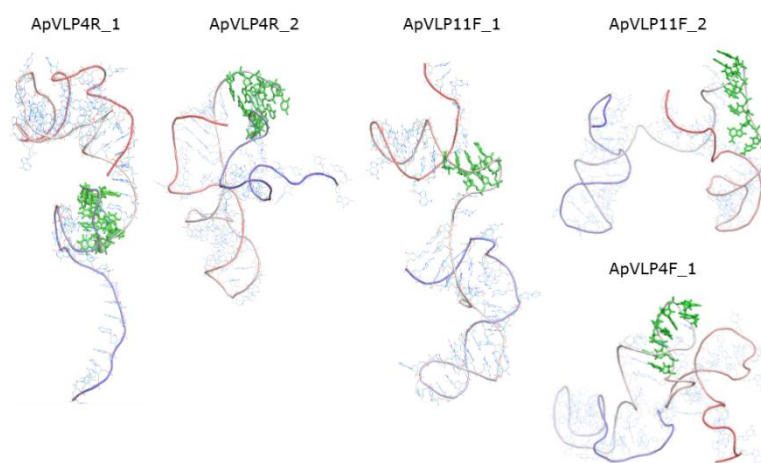


Figura 8. Estructuras generadas con RNAcomposer en base a los motivos secundarios de la Tabla II. Los residuos que se corresponden con el motivo común encontrado por el algoritmo MEME se muestran en verde. El *backbone* del ADN se representa como un tubo con un gradiente de rojo (extremo 5') a azul (extremo 3') pasando por blanco.

Es necesario solvatar, termalizar (llevar a temperatura del experimento) y equilibrar las estructuras usando como punto de partida los PDB obtenidos con el RNAcomposer. Esto se logra con técnicas de dinámica molecular. De este modo, las 5 estructuras de la Figura 8 fueron solvatadas reproduciendo las condiciones usadas experimentalmente en el SELEX (ver Materiales y métodos) y simuladas en la escala de los microsegundos (μ s), que son tiempos estado del arte en el área (30).

Usando el conjunto de herramientas de análisis que forman parte de la suite de programas bajo el nombre AMBER 24 (17) se analizaron las trayectorias de dinámica molecular para calcular el desvío cuadrático medio de todos los átomos a lo largo del tiempo (RMSD), de forma de comprobar la convergencia de las simulaciones. Se corrió por defecto un mínimo 1 μ s para cada estructura, pero cuando el RMSD aún fluctuaba mucho al final del primer microsegundo las simulaciones se extendieron hasta alcanzar 2 μ s. Es necesario tener en cuenta que este es el paso más costoso computacionalmente, que requirió de varias semanas de cálculos continuos en los servidores del grupo de Biofísica Computacional. Se analizó la flexibilidad local de los aptámeros calculando los factores de temperatura por nucleótido y se utilizó un método de *clustering* (ver Materiales y métodos) de las trayectorias de dinámica molecular para dilucidar los sub-estados conformacionales “navegados” por los 5 sistemas sobre las respectivas superficies de energía libre (ver Figura 9).

Como se desprende de la Figura 9, todos los sistemas se apartan en promedio de 8 a 15 Å de la estructura propuesta por el RNAcomposer en términos del desvío cuadrático medio de todos los átomos a lo largo del tiempo de simulación (RMSD). Se trata de desvíos importantes que resaltan la necesidad de equilibrar las estructuras 3D generadas. Una parte importante de ese desvío proviene de la transición que hacen las desoxirribosas de C3'-endo (forma canónica del ARN impuesta por el RNAcomposer) a C2'-endo (configuración observada en ADN). La estabilización del RMSD es una medida de orden cero de la convergencia en la dinámica molecular. Cuando se estabiliza alrededor de un valor más o menos constante durante varias decenas de nanosegundos, la simulación puede considerarse convergida. Cada sistema fue simulado por un mínimo de 1 μ s y se extendieron aquellas trayectorias que aún mostraban fluctuaciones importantes en el RMSD al cumplir ese tiempo de simulación. De esta forma, se extendieron los sistemas etiquetados como ApVLP4F_1, ApVLP4R_1 y ApVLP11F_1 hasta alcanzar estabilizar el RMSD. Sobre las trayectorias convergidas, se calculó los factores de temperatura por nucleótido que dan idea de la flexibilidad local de los aptámeros y se usó una técnica de *clustering* para extraer los sub-estados conformacionales visitados durante la simulación (ver Materiales y métodos). Se puede observar en la Figura 9 que todos los sistemas muestran una flexibilidad local heterogénea, dónde la región de secuencia común (áreas sombreadas) no muestra un comportamiento consistente. En algunos casos dichos nucleótidos se encuentran en picos de flexibilidad (ApVLP4R_1, ApVLP4R_2 y ApVLP11F_1), en otros casos en regiones más rígidas (ApVLP4F_1 y ApVLP11F_2). En términos absolutos, la secuencia ApVLP11F_2 fue la más flexible de todas, lo que podría explicar su mayor resistencia frente a la digestión con DNAasa I (ver Figura 4C). Finalmente, se presentan los sub-estados visitados por los 5 sistemas, dónde se evaluó el tiempo total en el que se observó dicho estado y el momento de la trayectoria que el mismo surge. De este modo, se conservó aquellos sub-estados que estaban presentes más de 15% de tiempo priorizando los que surgieron a partir de 500 ns (Figura 9).

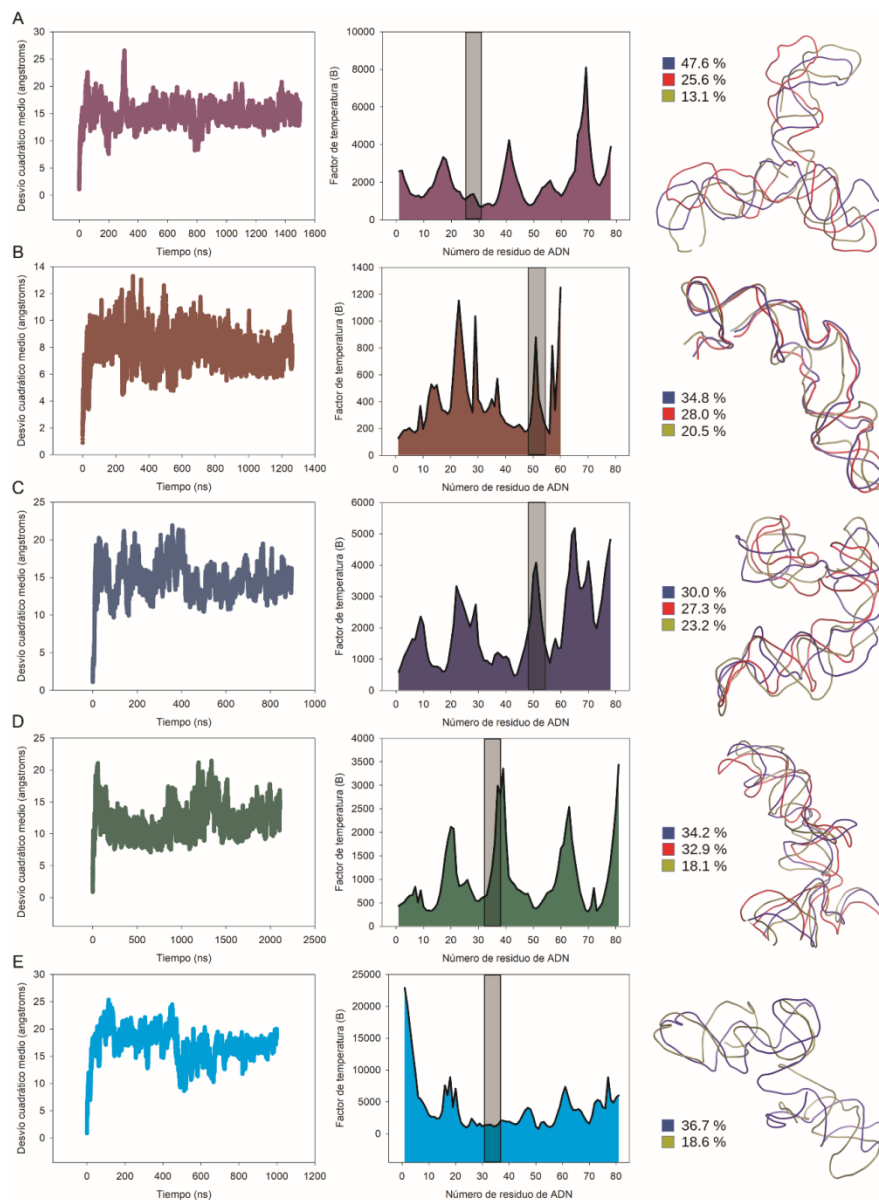


Figura 9. Análisis de las trayectorias de dinámica molecular de los 5 sistemas simulados en las condiciones del SELEX. Se muestra el RMSD (desvío cuadrático medio), la flexibilidad local (factores de temperatura) y los 12 sub-estados conformacionales capturados. Las regiones sombreadas se corresponden con el motivo común de secuencia. Los porcentajes representan la cantidad de tiempo en los que se encuentra un sub-estado dado durante la simulación. A) ApVLP4F_1; B) ApVLP4R_1; C) ApVLP4R_2; D) ApVLP11F_1; y E) ApVLP11F_2.

Los resultados de esta etapa fueron un conjunto de estructuras 3D optimizadas y refinadas de acuerdo a las condiciones experimentales que son representativas de los sub-estados conformacionales estables que cada aptámero adoptó durante las simulaciones. Eso dejó 12 conformaciones tridimensionales que caracterizan la flexibilidad que pueden adoptar las tres secuencias iniciales que recibimos del SELEX. Para etiquetarlos, se agregó la letra "c" por "*cluster*" y el número del *clúster*, siendo 0 el más poblado y 2 el menos poblado. De este modo las 12

estructuras finales del análisis de la dinámica se llamaron: ApVLP4F_1c0, ApVLP4F_1c1, ApVLP4R_1c0, ApVLP4R_1c1, ApVLP4R_1c2, ApVLP4R_2c0, ApVLP4R_2c1, ApVLP11F_1c0, ApVLP11F_1c1, ApVLP11F_1c2, ApVLP11F_2c0 y ApVLP11F_2c2 (ver Figura 10). Como puede observarse, se logró generar una amplia diversidad de estructuras tridimensionales, lo que permitió generar múltiples hipótesis sobre la unión aptámero-E1-3, abriendo el abanico de oportunidades de predecirlo correctamente.

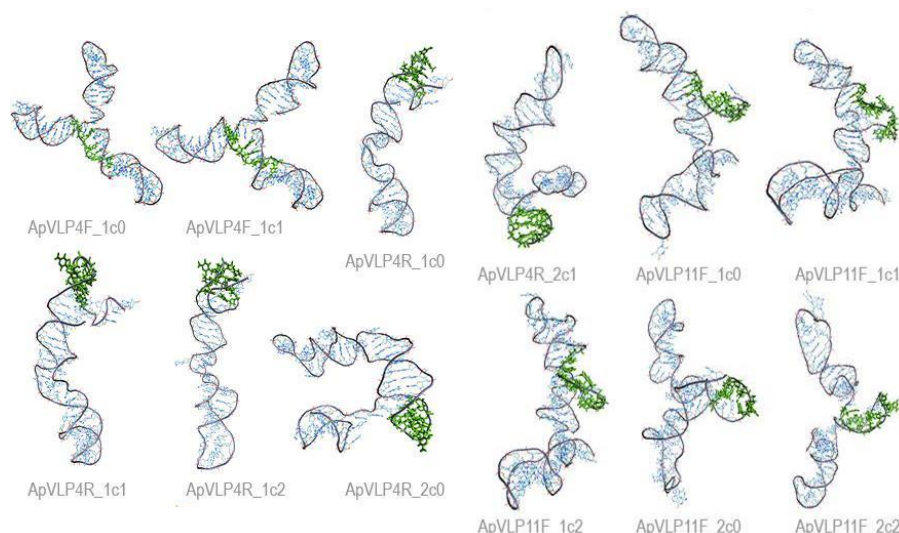


Figura 10. Estructura de los 12 sub-estados conformacionales capturados de las dinámicas moleculares. Se conservaron 2 sub-estados para el aptámero ApVLP4F, 5 para el aptámero ApVLP4R y otros 5 para ApVLP11F. Los residuos que se corresponden con el motivo común encontrado por el algoritmo MEME se muestran en verde.

4.3. Hipótesis del modo de unión y proposición de mutantes

Con el objetivo último de generar una serie de pautas (mutantes) que permitieran por un lado contrastar el enfoque teórico y por otro permitir comprender la estructura y modo de unión de los 3 aptámeros, se llevó a cabo experimentos de *docking* molecular usando los 12 sub-estados capturados durante las dinámicas (ver Figura 10). Esas 12 estructuras se enfrentaron al complejo E1-3, sesgando la búsqueda de posibles complejos a la región de secuencia común identificada por el programa MEME (ver Materiales y métodos). De este modo, se predijo la formación de 120 complejos (10 posibles poses por 12 sub-estados) usando el programa HDock. El mismo ordena las soluciones que brinda de acuerdo a la estabilidad relativa del complejo predicho del 1 al 10. De los 120 complejos se enfocó el análisis en las 10 poses que surgieron de enfrentar los 5 sub-estados mayoritarios (etiquetados como ApVLP4F_1c0, ApVLP4R_1c0, ApVLP4R_2c0, ApVLP11F_1c0 y ApVLP11F_2c0) contra E1-3. En la Tabla III se muestra un análisis de las orientaciones que adoptaron los aptámeros en su unión con las

proteínas de las cápsides. Solo aquellos sub-estados capaces de unirse a la superficie exterior de la cápside resultan de interés para el uso de los aptámeros como sondas moleculares para el diagnóstico. De la inspección visual de los 50 complejos, se seleccionó 16 que mostraron una orientación favorable (ver los complejos marcados con un asterisco en la Tabla III). Esas 16 hipótesis de unión se muestran en la Figura 11. De forma de identificar los 16 complejos seleccionados, se agregó a las etiquetas una “p” por pose y el número de la pose seleccionada (ver Figura 11). De este modo, los 16 complejos quedaron etiquetados como: ApVLP4F_1c0p10, ApVLP4R_1c0p4, ApVLP4R_1c0p5, ApVLP4R_1c0p7, ApVLP4R_1c0p8, ApVLP4R_1c0p9, ApVLP4R_2c0p2, ApVLP4R_2c0p3, ApVLP4R_2c0p9, ApVLP4R_2c0p10, ApVLP11F_1c0p1, ApVLP11F_1c0p4, ApVLP11F_1c0p5, ApVLP11F_1c0p6, ApVLP11F_1c0p9, ApVLP11F_1c0p10, ApVLP11F_2c0p1 y ApVLP11F_2c0p5.

Tabla III. Análisis de las proteínas de la cápside del virus en contacto con el aptámero y su orientación relativa. Las poses marcadas con una estrella son aquellas compatibles con una interacción favorable para el uso de los aptámeros como herramienta de diagnóstico.

Pose	ApVLP4R_1c0	ApVLP4R_2c0	ApVLP11F_1c0	ApVLP11F_2c0	ApVLP4F_1c0
1	E1-E2	E1-E2	E1-E2-E3 *	E1-E2 *	E2
2	E1-E2	E1*	E2	E2-E3	E2
3	E1-E2	E1-E2-E3 *	E1-E2	E2	E2-E3
4	E1-E2 *	E1-E3	E1-E2 *	E1-E2	E2-E3
5	E1-E2 *	E1-E2	E2-E3	E1-E2	E1
6	E1-E2-E3	E1-E2	E1-E3 *	E1-E2	E2
7	E2 *	E1-E2	E1-E2	E1-E2	E1-E2
8	E1-E2 *	E1-E2	E1-E2	E2-E3	E2-E3
9	E1-E2-E3 *	E1-E2-E3 *	E1-E2-E3 *	E1-E3	E2
10	E2	E1-E2-E3 *	E1-E2-E3 *	E2	E1 *

* Sub-estados con orientación favorable, capaces de unirse a la superficie exterior de la cápside

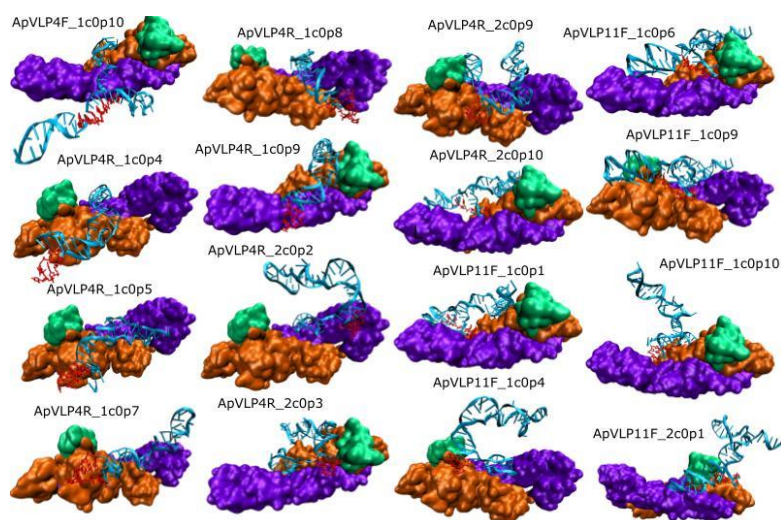


Figura 11. Representación de los 16 complejos seleccionados en los experimentos numéricos de docking molecular. Las proteínas se muestran como superficies de color violeta (E1), naranja (E2) y verde (E1).

Cada uno de los complejos representados en la Figura 11 fueron analizados uno a uno usando el visualizador VMD de forma de clasificar las interacciones a nivel atómico formadas entre los aptámeros y las proteínas de la cápside E1-3 (una versión de estas imágenes como anaglifs están disponibles en la Figura 13 en anexos). En base a ese análisis, se caracterizó las interacciones en específicas, inespecíficas y artefactos (ver Materiales y métodos). Los pocos artefactos hallados son una consecuencia de llevar a cabo experimentos de unión considerando ambos interactuantes (aptámero y complejo E1-3) cómo rígidos. Si bien existe la posibilidad de hacer *docking* moleculares semi-flexibles, los mismos son más costosos desde el punto de vista computacional y se consideraron más allá del presente trabajo. Los resultados del análisis de las interacciones moleculares se muestran para el complejo ApVLP11F_1c0p1 en la Tabla IV y en los anexos (Tabla V-VIII) para los complejos ApVLP11F_2c0p1, ApVLP4R_2c0p2, ApVLP4R_1c0p4 y ApVLP4F_1c0p10 respectivamente.

Se hizo enfoque en las interacciones específicas, por qué son aquellas que forman una interacción directa de las proteínas de la cápside con las bases nitrogenadas de los residuos del aptámero (ver Tabla IV). En base a los átomos involucrados en dichas interacciones, es posible proponer mutantes, que deberían, si las hipótesis de unión generadas son correctas, disminuir significativamente la afinidad de los aptámeros por el blanco. Se propuso para el aptámero ApVLP11F la mutación T36>A y A37>m3A (3-metilcitosina) y para el ApVLP4R los mutantes

A52>T y T53>G. Los resultados no fueron concluyentes para el aptámero ApVLP4F, por lo que no fue posible proponer mutantes en este caso.

Finalmente, mientras se elaboraba el proyecto que dio lugar a este trabajo de investigación, la empresa Google y su IA Deepmind publicaron el último servidor llamado Alphafold 3 (19) que permite predecir la estructura e interacción entre cualquier biomolécula de la cual se conoce la secuencia. Se tomó la secuencia de los tres aptámeros y del complejo E1-3 y se pidió a Alphafold 3 que predijera su estructura y la interacción entre ambos. En pocos segundos, se obtuvo el resultado mostrado en la Figura 12 (sólo se ilustra el resultado para el aptámero ApVLP4R, ya que los resultados fueron análogos para el caso ApVLP4F y ApVLP11F). Cómo puede verse, Alphafold 3 falla completamente a la hora de predecir la estructura del aptámero e incluso tiene problemas para representar correctamente la estructura de las proteínas de la cápside. Este resultado destaca la relevancia e importancia del enfoque teórico elaborado en el presente trabajo como una alternativa válida dónde la IA más avanzada falla.

Tabla IV. Análisis de las interacciones moleculares entre el aptámero ApVLP11F_1c0p1 y las proteínas de la cápside E1-3. En el caso de las interacciones específicas se indica las distancias involucradas en la interacción.

Tipo de interacción	Residuo de proteína	Residuo de ADN	Distancia (Å)
Inespecífico	Ser 744:OG	DC 21:OP1	--
Específica	Glu 739:OE2	DG 19:O6	2.65
Específica	Gln 747:OE1	DC 21:N4	3.41
Específica	Asn 534:OD1	DG 15:N7	1.80
Inespecífico	Asn 534:ND2	DG 15:OP2	--
Inespecífico	Asn 535:O	DG 14:OP2	--
Específico	Asn 535:N2	DG 23:N1	3.06
Específico	Asn 535:ND2	DG 23:O6	2.62
Inespecífico	Asn 658:ND2	DC 40:OP2	--
Específico	Asn 658:ND2	DT 36:O2	2.16
Específico	Asn 658:O	DT 36:N3	2.90
Específico	Thr 657:O	DT 36:O4	3.45
Específico	Arg 661:NH2	DT 36:O4	2.75
Inespecífico	Asn 658:OD1	DA 38:OP2	--
Inespecífico	Glu 201:O	DT 36:OP2	--
Inespecífico	Gln 209:O	DA 73:OP1	--
Específico	Thr 145:O	DT 72:O2	3.35
Específico	Hie 113:O	DG 52:O6	2.63
Específica	Asp 114:OD2	DG 53:N1	2.91

Específica	Asp 114:OD1	DG 53:O6	3.23
Específica	Thr 116:OG1	DT 54:O2	3.41
Artefacto	Trp 115:CZ2	DT 54:O4	--
Inespecífica	Glu 130:OE1	DG 56:OP1	--
Específica	Asn 56:N	DC 78:N4	2.74
Específica	Asn 56:OD1	DA 79:N7	3.23
Artefacto	Thr 52:CB	DC 81:OP1	--
Inespecífico	Ala 49:O	DC 81:OP1	--
Artefacto	Pro 20:CB	DC 81:O3'	--
Específico	Pro 55:OXT	DC 81:O2	2.89

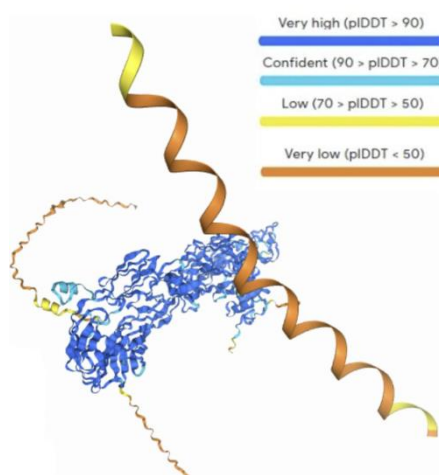


Figura 12. Resultado de la interacción entre el aptámero ApVLP4R y las proteínas E1-3 obtenida usando la IA AlphaFold 3 desarrollada por Google. El color de las moléculas indica la confianza de la IA en su predicción.

5. CONCLUSIONES Y PERSPECTIVAS

Se puso a punto una combinación de métodos teóricos que permitieron pasar de la secuencia de 3 aptámeros a 120 hipótesis de unión con la cápside del virus. La combinación de métodos de la bioinformática estructural y química/biofísica computacional dan una respuesta a la estructura 3D y a las interacciones moleculares aptámero-blanco, dónde la IA más avanzada (AlphaFold 3) no puede. Para contrastar experimentalmente nuestros resultados, se le propuso al IRYSIS repetir los ensayos de unión con los mutantes 36T>A y 37A>m3A para ApVLP11F; 52A>T y 53T>G para ApVLP4R. Estas mutaciones propuestas deberían bajar significativamente la afinidad de los aptámeros por el blanco, confirmando nuestras hipótesis de unión y posibilitando un ajuste fino y racional de los mismos para ser usados como sondas moleculares accesibles y ultra específicas para el diagnóstico de Chikungunya.

6. REFERENCIAS BIBLIOGRÁFICAS

1. Burgueño A, Giovanetti M, Fonseca V, Morel N, Lima M, Castro E, et al. Genomic and eco-epidemiological investigations in Uruguay reveal local Chikungunya virus transmission dynamics during its expansion across the Americas in 2023. *Emerging Microbes Infections*. 2024;13:1-10.
2. European Centre for Disease Prevention and Control. Geographical distribution of chikungunya virus disease cases reported worldwide, 2021. 2021.
3. Rougeron V, Sam IC, Caron M, Nkoghe D, Leroy E, Roques P. Chikungunya, a paradigm of neglected tropical disease that emerged to be a new health global risk. *Journal of Clinical Virology*. 2015;64:144–52. Available from: <https://www.sciencedirect.com/science/article/pii/S1386653214003710>
4. Kohlberger M, Gadermaier G. SELEX: Critical factors and optimization strategies for successful aptamer selection. *Biotechnology App Biochemistry*. 2022;69:1771–92.
5. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 2024 May 8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/38718835>
6. Agu PC, Afiukwa CA, Orji OU, Ezech EM, Ofoke IH, Ogbu CO, et al. Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management. *Sci Rep*. 2023;13(1).
7. Wayment-Steele HK, Kladwang W, Strom AI, Lee J, Treuille A, Becka A, et al. RNA secondary structure packages evaluated and improved by high-throughput experiments. *Nat Methods*. 2022;19(10):1234–42.
8. Sarzynska J, Popenda M, Antczak M, Szachniuk M. RNA tertiary structure prediction using RNAComposer in CASP15. *Proteins*. 2023;91(12):1790–9. Available from: <https://doi.org/10.1002/prot.26578>
9. Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, et al. Automated 3D structure composition for large RNAs. *Nucleic Acids Res*. 2012; 40(14):e112. Available from: <https://doi.org/10.1093/nar/gks339>
10. Zhang Y, Xiong Y, Xiao Y. 3dDNA: A Computational method of building DNA 3D structures. *Molecules*. 2022;27(18):1-10.
11. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Research*. 2000;28:235-42. Available from: <http://www.rcsb.org/pdb/status.html>
12. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37:W202-8.
13. Welcome to the Predict a Secondary Structure Web Server. [cited 2024 Nov 15]. Available from: <https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html>
14. Santa Lucia J, Hicks D. The Thermodynamics of DNA Structural Motifs. *Annu Rev Biophys*. 2004;33:415–40. Available from: <https://www.annualreviews.org/content/journals/10.1146/annurev.biophys.32.110601.141800>
15. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*. 2011;6(1):26. Available from: <https://doi.org/10.1186/1748-7188-6-26>
16. RNAComposer. [cited 2024 Nov 16]. Available from: <https://rnacomposer.cs.put.poznan.pl/>

17. Case DA, Cheatham TE, et al. Amber 2024. [cited 2024 Nov 15]. Available from: <https://ambermd.org/contributors.html>
18. Ivani I, Dans PD, Noy A, Pérez A, Faustino I, Hospital A, et al. Parmbsc1: A refined force field for DNA simulations. *Nat Methods*. 2015;13(1):55–8.
19. Izadi S, Anandakrishnan R, Onufriev A V. Building Water Models: A Different Approach. *J Phys Chem Lett*. 2014;5(21):3863–71. Available from: <https://doi.org/10.1021/jz501780a>
20. Li P, Song LF, Merz KMJr. Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J Chem Theory Comput*. 2015;11(4):1645–57. Available from: <https://doi.org/10.1021/ct500918t>
21. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: An automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res*. 2004;32:W665–7. Available from: <https://doi.org/10.1093/nar/gkh381>
22. AlphaFold Server. Google DeepMind. [cited 2024 Nov 17]. Available from: <https://deepmind.google/technologies/alphafold/alphafold-server/>
23. Atom. [cited 2024 Nov 15]. Available from: <https://atom-editor.cc/>
24. GIMP - GNU Image Manipulation Program. [cited 2024 Nov 15]. Available from: <https://www.gimp.org/>
25. LibreOffice en español - suite ofimática libre, basada en OpenOffice, compatible con Microsoft. [cited 2024 Nov 15]. Available from: <https://es.libreoffice.org/>
26. Grace Home. [cited 2024 Nov 15]. Available from: <https://plasma-gate.weizmann.ac.il/Grace/>
27. Gnuplot. [cited 2024 Nov 15]. Available from: <http://www.gnuplot.info/>
28. Mol*. [cited 2024 Nov 15]. Available from: <https://molstar.org/>
29. Humphrey W, Dalke A, Schulten K. VMD: Visual Molecular Dynamics. 1996;14(1):33–8.
30. da Rosa G, Grille L, Calzada V, Ahmad K, Arcon JP, Battistini F, et al. Sequence-dependent structural properties of B-DNA: what have we learned in 40 years? *Biophysical Reviews*; 2021;13:995–1005.

7. ANEXO

Tabla V. Análisis de las interacciones moleculares entre el aptámero ApVLP11F_2c0p1 y las proteínas de la cápside E1-3. En el caso de las interacciones específicas se indica las distancias involucradas en la interacción.

Tipo de interacción	Residuo proteína	Residuo ADN	Distancia (Å)
Artefacto	Lys 526:CE	DA 37:OP2	--
Inespecífico	Leu 428:O	DA 37:O4´	--
Específico	Asn 664:ND2	DA 37:N3	3.22
Inespecífico	Glu 385:OE2	DG 39:O3´	--
Inespecífico	Lys 321:N2	DC 40:OP1	--
Artefacto	Lys 200:CG	DC 40:OP2	--
Específico	Gln 198:OE1	DT 28:O2	2.51
Artefacto	Hie 198:CB	DG 7: C5´	--
Inespecífico	Gln 197:O	DG 29:OP2	--
Inespecífico	Arg 195:NH2	DT 24: OP2	--
Artefacto	Arg 155:CB	DG 23:OP1	--
Artefacto	Thr 143:CG2	DG 23:C5´	--
Específico	Val 136:O	DT 16:O4	2.67
Específico	Arg 137:NH1	DT 16:O2	2.92
Específico	Hie 135:CB	DT 16:C	2.99
Artefacto	Pro 141:CD	DG 19:OP1	--
Artefacto	Pro 141:CD	DG 19:OP2	--
Específico	Ser 139:O	DT 18:O2	3.03

Tabla VI. Análisis de las interacciones moleculares entre el aptámero ApVLP4R_2c0p2 y las proteínas de la cápside E1-3. En el caso de las interacciones específicas se indica las distancias involucradas en la interacción.

Tipo de interacción	Residuo proteína	Residuo ADN	Distancia (Å)
Artefacto	Glu 737:O	DT 50:C7	--
Específico	Gln 532:O	DT 52:N6	3.05
Específico	Gly 533:O	DT 52:N6	3.01
Artefacto	Leu 530:CD2	DT 50:OP1	--
Específico	Asn 534:ND2	DT 53:O4	3.37
Artefacto	Lys 551:N7	DA 57:C4´	--
Artefacto	Val 350:C61	DC 58:OP1	--
Específico	Lys 526:N7	DC 72:N1	3.12
Artefacto	Thr 539:C62	DC 72:OP2	--
Específica	Ser 524:O6	DA 73:N6	2.17
Inespecífico	Tyr 541:OH	DT 71:O5´	--
Inespecífico	Arg 661:NH2	DT 74:OP2	--
Específico	Glu 432:OE2	DT 70:N3	3.02
Específico	Glu 423:OG1	DT 70:O4	2.82
Específico	Asn 658:OD2	DG 67:N7	2.73
Inespecífico	Leu 436:O	DT 68:OP2	--
Inespecífico	Thr 516:OH	TG 67:OP1	--
Inespecífico	Ser 514:OG	DG 67:OP1	--
Artefacto	Pro 602:CO	DC 64:N4	--
Inespecífico	Leu 438:N	DT 68:OP1	--

Tabla VII. Análisis de las interacciones moleculares entre el aptámero ApVLP4R_1c0p4 y las proteínas de la cápside E1-3. En el caso de las interacciones específicas se indica las distancias involucradas en la interacción.

Tipo de interacción	Residuo proteína	Residuo ADN	Distancia (Å)
Inespecífico	Leu 438:O	DT 26:OP1	--
Inespecífico	Arg 600:O	DT 26:OP1	--
Inespecífico	Tir 516:OH	DA 27:OP1	--
Inespecífico	Glu 603:E2	DG 17:OP2	--
Inespecífico	Lys 509:O	DG 16:OP2	--
Inespecífico	Asn 314:ND2	DG 16:OP1	--
Inespecífico	Glu 201:OE1	DT 12:O4'	--
Específico	Glu 201:OE2	Dt 12:O2	3.32
Inespecífico	Gln 297:OE1	DA 10:O5'	--
Inespecífico	Arg 195:NH2	DG 8:OP2	--
Inespecífico	Gln 209:O	DG 41:OP1	--
Artefacto	Fe 301:CR1	DG 41:OP1	--
Inespecífico	Arg 131:NH2	DC 42:OP1	--
Inespecífico	Arg 131:NH2	DC 42:OP2	--
Inespecífico	The 147:OG1	DC 42:OP2	--
Artefacto	Thr 145:O	DG 41:C5'	--
Específico	Trp 115:CZ3	Dc 42:C5	2.90
Inespecífico	Asp 114:OD2	DA 57:OP1	--
Inespecífico	Glu 130:OE2	DC 55:OP2	--
Artefacto	Glu 117:CE	DG 56:OP2	--
Inespecífico	Lys 108:NZ	DG 56:OP1	--

Tabla VIII. Análisis de las interacciones moleculares entre el aptámero ApVLP4F_1c0p10 y las proteínas de la cápside E1-3. En el caso de las interacciones específicas se indica las distancias involucradas en la interacción.

Tipo de interacción	Residuo proteína	Residuo ADN	Distancia (Å)
Específico	Arg 600:NH1	DG 34:N7	3,42
Específico	Arg 600:NH2	DG 34:O6	3,82
Específico	Asn 658:ND2	DT 60:O2	2,03
Específico	Asn 658:OD1	DG 59:N3	3,39
Específico	Glu 201:OE1	DC 64:N4	3,32
Específico	Glu 201:OE1	DC 64:N4	2,58
Específico	Arg 318:NH2	DA 70:N6	2,46
Inespecífico	Tyr 608:OH	DG 44:OP1	--
Inespecífico	Asp 438:OD1	DC 32:OP1	--
Inespecífico	Asp 438:OD1	DC 32:OP2	--
Inespecífico	Asp 438:OD1	DC 32:OP5'	--
Inespecífico	Arg 517:NH1	DA 31:OP1	--
Inespecífico	Asn 569:OD1	DA 31:OP1	--
Inespecífico	Asn 569:ND2	DA 31:OP1	--
Inespecífico	Glu 603:OE1	DG 59:OP1	--
Inespecífico	Lys 575:O	DG 61:OP1	--
Inespecífico	Arg 195:NH1	DA 70:OP2	--
Inespecífico	Arg 195:NH2	DA 70:OP2	--
Inespecífico	Gln 209:OE1	DT 69:O3'	--
Artefacto	Phe 512:CZ	DT 60:OP1	--
Artefacto	Thr 316:C62	DC 71:N4	--
Artefacto	Val 315:C62	DC 71:OP2	--
Artefacto	Val 315:C61	DC 71:OP2	--
Artefacto	Thr 206:C62	DC 71:OP1	--
Artefacto	Gln 209:CG	DT 69:O2	--

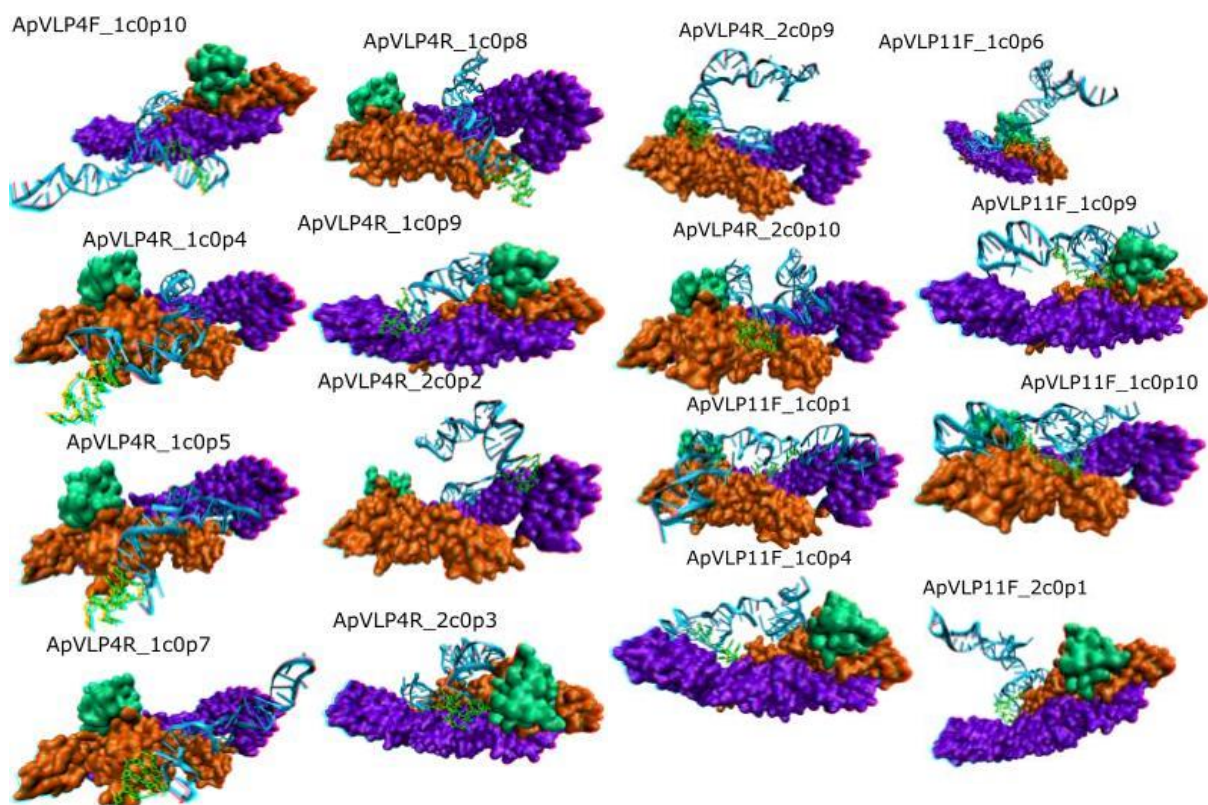


Figura 13. Anaglifos de los 16 complejos seleccionados en los experimentos numéricos de docking molecular. Las proteínas se muestran como superficies de color violeta (E1), naranja (E2) y verde (E1). Para poder visualizar estos complejos es necesario utilizar lentes 3D.