



**Centro de Investigación y de Estudios Avanzados  
del Instituto Politécnico Nacional  
Unidad Monterrey**

**Predicción teórica de la estructura molecular y  
determinación del mecanismo de acción de péptidos  
cíclicos inhibidores de la interacción  
proteína-proteína de la cápside del *Cowpea Chlorotic  
Mottle Virus***

Tesis que presenta

**Josué Alejandro Alemán Vilis**

Para obtener el grado de

**Maestro en Ciencias**

En la Especialidad de Ingeniería y Física Biomédica

Director de Tesis:

Dr. Mauricio Carrillo Tripp

Apodaca, Nuevo León

Septiembre 2020



El presente trabajo fue realizado en el Laboratorio de la Diversidad Biomolecular en el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional Unidad Monterrey, bajo la dirección del Dr. Mauricio Carrillo Tripp y evaluado por el comité tutorial conformado por el Dr. Bruno A. Escalante Acosta y la Dra. Roxana G. Gutiérrez Vidal.

## **Agradecimientos**

Esta investigación fue financiada por el Fondo de Ciencia Básica (proyecto número 132376) y el Fondo Sectorial de Investigación para la Educación (proyecto número A1-S-17041), ambos del Consejo Nacional de Ciencia y Tecnología México, así como por fondos fiscales del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional Unidad Monterrey, todos ellos aprobados y asignados a M.C.T. Agradecemos al Centro por dar acceso a su infraestructura a todos los miembros del grupo de investigación del Laboratorio de la Diversidad Biomolecular.

También agradecemos al Consejo Nacional de Ciencia y Tecnología México por haber otorgado una beca de posgrado a J.A.A.V durante los dos años del programa de Maestría en Ingeniería y Física Biomédica del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional Unidad Monterrey.

# Índice

<b>Índice de figuras</b>	<b>8</b>
<b>Índice de tablas</b>	<b>10</b>
<b>Resumen</b>	<b>11</b>
<b>Abstract</b>	<b>12</b>
<b>1. Introducción</b>	<b>13</b>
<b>2. Antecedentes</b>	<b>14</b>
2.1. Interacciones proteína-proteína (IPP) . . . . .	14
2.2. La cápside viral como blanco de inhibición . . . . .	16
2.3. Modelo viral: <i>Cowpea Chlorotic Mottle Virus</i> . . . . .	17
2.4. Inhibidores de la interacción proteína-proteína . . . . .	19
2.5. Péptidos cíclicos (PCs) . . . . .	20
2.6. PCs inhibidores de IPPs . . . . .	21
2.7. Predicción teórico-computacional de la estructura de PCs . . . . .	22
<b>3. Hipótesis</b>	<b>23</b>

<b>4. Objetivo General</b>	<b>23</b>
<b>5. Objetivos Específicos</b>	<b>23</b>
<b>6. Metodología</b>	<b>24</b>
6.1. PCs estudiados . . . . .	24
6.2. Predicción estructural de PCs <i>in-silico</i> . . . . .	25
6.2.1. Modelado molecular . . . . .	25
6.2.2. Dinámica molecular . . . . .	27
6.3. Acoplamiento Molecular . . . . .	29
<b>7. Resultados y Discusión</b>	<b>31</b>
7.1. Modelado Molecular . . . . .	31
7.2. Dinámica molecular . . . . .	35
7.3. Análisis Estructural . . . . .	40
7.3.1. Equilibrio dinámico conformacional: RMSD(t) . . . . .	41
7.3.2. Análisis por clusters . . . . .	44
7.3.3. Matrices de RMSD . . . . .	51
7.4. Mecanismo molecular de inhibición . . . . .	57
7.4.1. Acoplamiento molecular: cavidades . . . . .	57

7.4.2.	Acoplamiento molecular: <i>residuos clave</i> . . . . .	58
7.4.3.	Acomplamiento molecular: resultados . . . . .	62
7.4.4.	Escaneo de Alaninas . . . . .	67
<b>8.</b>	<b>Conclusiones y Perspectivas</b>	<b>72</b>
<b>9.</b>	<b>Referencias</b>	<b>75</b>
<b>10.</b>	<b>Apéndice 1</b>	<b>85</b>
10.1.	Generación de la topología para los péptidos cíclicos . . . . .	85
10.2.	Fase de preparación de la DM . . . . .	88
10.3.	Parámetros de simulación . . . . .	90
10.3.1.	ions.mdp . . . . .	90
10.3.2.	01_minim.mdp . . . . .	91
10.3.3.	02_nvt.mdp . . . . .	92
10.3.4.	03_npt.mdp . . . . .	94
10.3.5.	04_md.mdp . . . . .	96
10.4.	Fase de análisis . . . . .	99
<b>11.</b>	<b>Apéndice 2</b>	<b>103</b>
11.1.	Síntesis, ciclación y purificación de PCs . . . . .	103

11.1.1. Síntesis de péptidos en fase sólida . . . . .	103
11.1.2. Carga de la resina . . . . .	104
11.1.3. Cuantificación de la carga . . . . .	105
11.1.4. Capping . . . . .	105
11.1.5. Acoplamiento . . . . .	105
11.1.6. Desprotección . . . . .	106
11.1.7. Prueba de ninhidrina . . . . .	106
11.1.8. Desanclaje de la resina . . . . .	107
11.2. Ciclación de los péptidos lineales . . . . .	107
11.2.1. Adición del grupo aldritol . . . . .	109
11.2.2. Reacción de ciclación . . . . .	109
11.2.3. Reducción de los enlaces disulfuro con TCEP . . . . .	109
11.3. Purificación por cromatografía de líquidos de alta resolución . . .	110
11.4. Determinación estructural mediante RMN . . . . .	110

## Índice de figuras

1.	Esquema de una IPP . . . . .	15
2.	Ensamblado de partículas virales . . . . .	17
3.	Estructura cuaternaria del <i>Cowpea Chlorotic Mottle Virus</i> (CCMV) . . . . .	18
4.	Pipeline de MM y MD . . . . .	26
5.	Estructura química de los PCs . . . . .	32
6.	Estructura molecular de S1, S2, y S3 de los PCs . . . . .	33
7.	RMSD(t) de las conformaciones S1, S2 y S3 de los PCs. . . . .	41
8.	RMSD(t) de las conformaciones S1, S2 y S3 del PC-C7. . . . .	42
9.	Superficie de energía conformacional . . . . .	43
10.	RMSD(t) y análisis de clusters de las conformaciones C1 . . . . .	45
11.	RMSD(t) y análisis de clusters de las conformaciones C2 . . . . .	45
12.	RMSD(t) y análisis de clusters de las conformaciones C3 . . . . .	46
13.	RMSD(t) y análisis de clusters de las conformaciones C4 . . . . .	46
14.	RMSD(t) y análisis de clusters de las conformaciones C5 . . . . .	47
15.	RMSD(t) y análisis de clusters de las conformaciones C6 . . . . .	47
16.	RMSD(t) y análisis de clusters de las conformaciones C7 . . . . .	48
17.	RMSD(t) y análisis de clusters de las conformaciones C8 . . . . .	48



18.	Cluster de la serie C de PCs, <i>Cut-Off</i> = 0.20nm . . . . .	50
19.	Matriz de RMSD de las diferentes conformaciones del PC-C1 . . .	52
20.	Matriz de RMSD de las diferentes conformaciones del PC-C2 . . .	52
21.	Matriz de RMSD de las diferentes conformaciones del PC-C3 . . .	53
22.	Matriz de RMSD de las diferentes conformaciones del PC-C4 . . .	53
23.	Matriz de RMSD de las diferentes conformaciones del PC-C5 . . .	54
24.	Matriz de RMSD de las diferentes conformaciones del PC-C6 . . .	54
25.	Matriz de RMSD de las diferentes conformaciones del PC-C7 . . .	55
26.	Matriz de RMSD de las diferentes conformaciones del PC-C8 . . .	55
27.	Cápside, homodímero primario y subunidades A2 y B1 del CCMV.	58
28.	Las cinco cavidades más grandes de la subunidad A2 del CCMV. . .	59
29.	Las cinco cavidades más grandes de la subunidad B1 del CCMV. . .	59
30.	Cavidades localizadas en la interfaça A2-B1. . . . .	60
31.	Cavidades y <i>residuos clave</i> de la subunidad B1 del CCMV . . . . .	62
32.	Apocinina . . . . .	66
34.	Secuencia se síntesis de péptidos en fase sólida . . . . .	104
35.	Tipos de ciclación de péptidos. . . . .	108

## Índice de tablas

1.	Estructura primaria de PCs inhibidores del CCMV . . . . .	24
2.	TM-score entre S1 y S2 de los PCs . . . . .	34
3.	TM-score entre S1 y S3 de los PCs . . . . .	34
4.	Condiciones químicas y termodinámicas de simulación . . . . .	37
5.	Cavidades como sitio de unión . . . . .	60
6.	Acoplamiento de los PCs con la CP-CCMV . . . . .	64
7.	Péptidos cíclicos control E1, E2 y E3 . . . . .	65
8.	Acoplamiento CP-CCMV con Apocinina . . . . .	66
9.	Acoplamiento de mutantes PC C2 con el CCMV . . . . .	69
10.	Acoplamiento de mutantes PC C8 con el CCMV . . . . .	70

## Resumen

Los péptidos cíclicos (PCs) combinan diversas propiedades biológicas y fisico-químicas que hacen de ellos una modalidad atractiva para el desarrollo de nuevos fármacos. En particular, nuestro grupo ha mostrado con anterioridad que los PCs pueden funcionar como inhibidores específicos de la interacción proteína-proteína (IPP) de cápsides virales *in-vitro*. Sin embargo, todavía carecemos de un entendimiento molecular del proceso de inhibición de la IPP. Considerando lo anterior, es importante desarrollar una estrategia teórico-computacional eficiente y precisa para predecir la estructura tridimensional de PCs. En este trabajo acoplamos el Modelado Molecular con simulaciones numéricas de Dinámica Molecular para predecir y caracterizar las estructuras tridimensionales y el comportamiento en solución de PCs. Se estudiaron ocho diferentes secuencias de PCs inhibidores de la formación de homodímeros de la proteína de cápside (CP) del *Cowpea Chlorotic Mottle Virus*, previamente identificados experimentalmente en nuestro grupo. Los resultados muestran la existencia de una alta diversidad conformacional de los PCs en solución fisiológica. Dichas predicciones estructurales se utilizaron en cálculos de acoplamiento molecular con la CP del CCMV. Se identificaron los cinco sitios de unión más probables en la superficie molecular de la CP y se estimó la afinidad de los PCs en cada uno de ellos. Nuestros resultados permiten proponer un posible mecanismo de inhibición molecular en el que se bloquea la participación del *residuo clave* E176 para la formación del complejo CP-CP. Esto concuerda con la evidencia experimental reportada previamente por nuestro grupo. Por último, también se mostró que la presencia de residuos aromáticos en sitios específicos de la secuencia de los PCs es esencial para que haya una buena interacción con la CP.

## Abstract

Cyclic peptides (CP) combine diverse biological and physicochemical properties making them an attractive modality for the development of new drugs. In particular, CP can function as specific inhibitors of protein-protein interaction (PPI). However, we still lack a molecular understanding of the inhibition process of PPI. Considering the above, develop an efficient and precise theoretical-computational strategy to predict the three-dimensional structure of PCs, describe the molecular mechanisms underlying their inhibitory property of protein complexes. In this work, we propose to integrate Molecular Modeling with numerical simulations of Molecular Dynamics to predict and characterize the three-dimensional structures of CPs. We studied 8 different CPs sequences, each of them able to inhibit the formation of homodimers of the Capsid Protein of *Cowpea Chlorotic Mottle Virus*, these CPs were previously experimentally identified in our group. The results of this prediction process show the existence of a high conformational diversity of CPs in physiological solution. The previous predictions were used in molecular docking experiments, using the Capsid Proteins of CCMV as receptors. We identified 5 binding sites, with that evidence we propose a possible molecular inhibition mechanism where the participation of a hot-spot residue is blocked. Fact that agrees with the experimental evidence previously reported by our group. Finally, it was also demonstrated that the presence of aromatic residues in the sequence of the CPs is essential for a good interaction with the capsid proteins.

# 1. Introducción

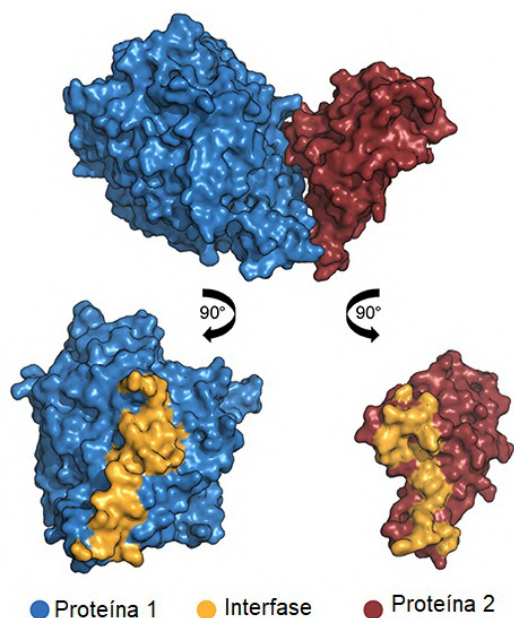
Aproximadamente el 75 % de las proteínas humanas relevantes en el desarrollo de patologías, incluidas aquellas relacionadas con interacciones proteína-proteína (IPP), son intratables con las modalidades farmacéuticas actuales. Los péptidos cíclicos (PCs) pueden ser una solución potencial a este problema, pues sus características fisicoquímicas y conformacionales les dan la capacidad de unirse específicamente a la interfase de una IPP con gran afinidad. Recientemente, y gracias al uso de tecnologías basadas en el diseño racional y uso de bibliotecas moleculares, se han descubierto nuevos PCs inhibidores de las IPPs para diversas patologías [1]. A pesar de estos avances, aún se desconoce el mecanismo molecular por el cual actúan la mayoría de estos PCs, pues para ello se requiere conocer su estructura tridimensional. A la fecha, existen diversos algoritmos computacionales para predecir estas estructuras, pero siempre será necesaria la validación comparando resultados con datos experimentales. En este trabajo se realizó la predicción de la estructura molecular de un conjunto de PCs a partir de su secuencia primaria por métodos teóricos, con la perspectiva de ser validada posteriormente con métodos experimentales. Así, con esa información se estudió el mecanismo molecular de inhibición de la formación de la cápside de un virus de topología icosaédrica.

## 2. Antecedentes

### 2.1. Interacciones proteína-proteína (IPP)

Las interacciones proteína-proteína (IPP) pueden ser descritas como interacciones débiles que toman lugar entre dos superficies proteínicas diferentes pero complementarias. Generalmente, las IPPs se pueden clasificar como interacciones lipofílicas,  $\pi - \pi$ ,  $\pi$  - catión, puentes de hidrógeno o interacciones iónicas [2, 3]. Recientemente, se han propuesto a las IPPs como blancos terapéuticos. Esto se debe a que las interacciones proteínicas tienen un rol importante en muchos procesos biológicos y patológicos, incluyendo al cáncer, Alzheimer y problemas cardiovasculares [4, 5]. Se ha estimado teóricamente que el interactoma humano contiene hasta 650,000 IPPs [6], de las cuales sólo una pequeña fracción han sido estudiados como blancos farmacológicos. En la última década, poco más de 40 IPPs se han caracterizado como blancos terapéuticos y algunos inhibidores de éstas han pasado a la fase clínica [7].

Los análisis estructurales de complejos IPP realizados hasta la fecha muestran que la superficie de contacto entre las proteínas va de los 1,000 a los 4,000 Å<sup>2</sup>, con una geometría más plana y mucho más grande que la superficie de interacción de los sitios de unión de receptores proteicos conocidos [8]. Una interfase típica IPP (Figura 1) contiene dos regiones específicas: el núcleo y el borde. Se presume que la mayoría de las IPP ocurren debido a la presencia de residuos hidrofóbicos en los núcleos, los cuales se encuentran más conservados y son menos móviles que los residuos hidrofílicos de los bordes [7, 3, 9].



**Figura 1:** Esquema de una IPP. La zona amarilla representa la región de la interfase P-P, en cuyo centro se haya el núcleo hidrofóbico, donde se presume ocurren las interacciones que estabilizan el complejo.

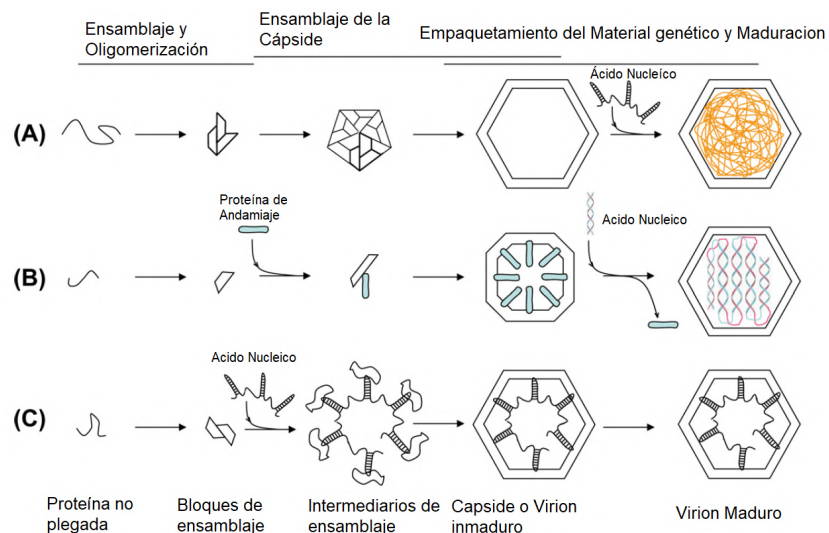
Mientras que moléculas pequeñas destacan en la interacción con proteínas que funcionan como receptores, la interfase de las IPP demanda de moléculas más grandes. Por lo anterior, diversos grupos de investigación han enfocado su atención en moléculas macrocíclicas. Estas suelen tener tamaños de tres a cinco veces más grandes que los de los fármacos convencionales, y con una flexibilidad y rigidez conformacional balanceada. Los macrociclos son capaces de interactuar con la interfase plana de las IPP con una afinidad similar a la de los anticuerpos y al mismo tiempo retienen algunas de las particularidades de las moléculas pequeñas, como la estabilidad bioquímica de su estructura [10, 11].

## 2.2. La cápside viral como blanco de inhibición

Las partículas víricas infecciosas son diversas en tamaño, forma, composición molecular, organización estructural y complejidad. Pero a pesar de dichas diferencias, la cápside viral puede ser descrita de la misma manera para todos los virus, como un complejo oligomérico y simétrico hueco, formado a partir de muchas copias de una o de pocos tipos de proteínas de cápside (CP) referidas como subunidades [12]. La mayoría de las cápsides tienen una simetría helicoidal o icosaédrica. Esta última está presente en aproximadamente la mitad de las familias de virus conocidas [13].

Los agentes antivirales disponibles actualmente tienen como objetivo interferir en alguna etapa del ciclo de vida de los virus, particularmente inhibiendo blancos enzimáticos. Dicho ciclo comienza con la unión del virus a la célula hospedera al interactuar con receptores específicos, pasando por la replicación del material genético viral una vez adentro, hasta el ensamblaje de la cápside con el material genético y la liberación de los viriones recién sintetizado. [14]. Considerando los diversos tipos de proteínas virales de acuerdo a su función, la CP es donde se encuentra el mayor porcentaje de residuos conservados dentro de un mismo género o familia, incluso con respecto a proteínas celulares [9]. Tales regiones conservadas corresponden a las interfaces CP-CP. Aún más, existe evidencia que en esas regiones también se encuentran residuos clave para la correcta formación de la estructura macromolecular [15]. El ensamblaje de la cápside se destaca como un paso común a todos los virus en el ciclo de vida viral. Así, recientemente hemos comenzado a explorar la hipótesis que plantea que la cápside es un blanco alternativo como un enfoque novedoso para la inhibición de infecciones virales [16].



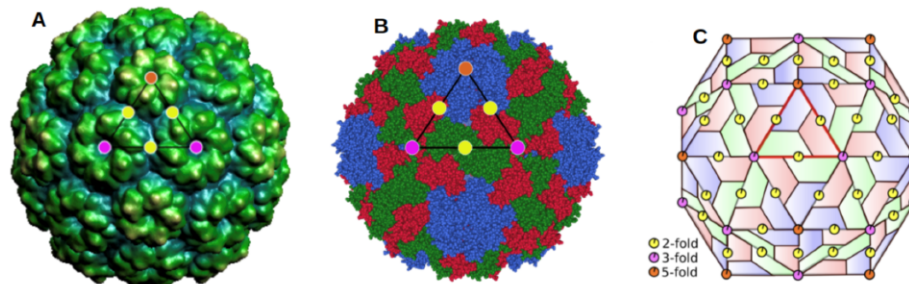


**Figura 2:** Esquema simplificado de las tres estrategias de ensamblado de las partículas virales (Imagen modificada de [13]).

Actualmente se reconocen tres estrategias de ensamblaje de las cápsides virales (Figura 2): (A) Autoensamblaje de las proteínas de cápside; (B) Ensamblaje auxiliado por proteínas de andamiaje; (C) Ensamblaje asistido por la condensación de ácidos nucleicos víricos [13]. En este proceso de ensamblaje, las interacciones intermoleculares entre proteínas toma el papel principal. Por lo tanto, si se quiere inhibir la formación de la cápside viral es natural pensar en intentar interrumpir dichas IPPs CP-CP.

### 2.3. Modelo viral: *Cowpea Chlorotic Mottle Virus*

El *Cowpea Chlorotic Mottle Virus* (CCMV) fue el primer virus icosaédrico ensamblado *in-vitro*. Se utilizaron proteínas purificadas de la cápside y ARN pa-



**Figura 3:** Estructura cuaternaria de una cápside típica de un virus icosaédrico. Tres representaciones del *Cowpea Chlorotic Mottle Virus* (CCMV). **A)** Representación de la superficie molecular de la cápside. Los tonos de verde indican la distancia radial al centro de la partícula. **B)** Localización tridimensional de todos los átomos que componen la cápside. Los diferentes colores indican distintos entornos locales. Las subunidades del mismo color tienen posiciones equivalentes en la cápside viral. **C)** Cada trapezoide representa una subunidad individual ensamblada en el arreglo icosaédrico. También se muestra la ubicación de tres tipos de ejes de simetría que definen dicha geometría; *2-fold* en amarillo, *3-fold* en magenta y *5-fold* en naranja. Esta arquitectura se caracteriza por tener 60 caras triangulares equivalentes. Una de ellas está resaltada (rojo), comúnmente conocida como la unidad asimétrica icosaédrica central (CIAU, en inglés) en la convención VIPERdb [20].

ra formar partículas infecciosas [17]. Desde entonces, el CCMV ha sido sujeto a diversos estudios y caracterizaciones, lo que lo ha convertido en el arquetipo de los virus icosaédricos. Se trata de un fitopatógeno, con un genoma de ARN+ de una sola cadena, envuelto en una cápside formada por 180 copias idénticas de una proteína de cápside (CP) de 20.3 kDa [16]. En 1995 se determinó la estructura cristalográfica del CCMV y para el 2002 se pudieron formar las primeras partículas virales no infecciosas o VLPs (*Virus Like Particles*) [18, 19]. En la Figura 3 se puede apreciar la cápside del CCMV, su arquitectura cuaternaria y elementos topológicos de simetría. Dadas estas características su uso experimental en el laboratorio es seguro.

## 2.4. Inhibidores de la interacción proteína-proteína

Dado que las IPPs tienen características fisicoquímicas y geométricas diferentes a los sitios de unión de los receptores proteicos, las moléculas que inhiban estas IPP también deben de tener características diferentes a aquellos fármacos tradicionales. Morelli y colaboradores [21] analizaron estadísticamente las características químicas de 39 inhibidores no redundantes unidos a sus correspondientes interfaces de interacción. Dicho trabajo se realizó con información disponible en la literatura y en el *Protein Data Bank*. La plataforma recolecta datos cristalográficos de inhibidores de IPP (*IIPP*) y permite generar una idea más clara de las características que hacen mejor a un inhibidor de IPPs. Con base en esta plataforma, Morelli propuso la “Rule of 4 (RO4)”, con la cual se pretende definir el perfil general que debe de cumplir un compuesto inhibidor de IPPs: peso molecular (PM) > 400 Da, hidrofobicidad molecular (AlogP) > 4, número de anillos (#anillos) > 4, y el número de aceptores de hidrógeno (#AH) > 4.

Por otro lado, Villoutreix *et al.* [22] analizaron las mismas propiedades fisicoquímicas que Morelli en una base de datos, *i-PPI*, que contiene la información estructural, propiedades fisicoquímicas, datos farmacológicos y el perfil de interacción de varios cientos de moduladores de IPPs. Sus resultados muestran congruencia con los de Morelli. Los ligandos en *i-PPI* son más pesados (PM promedio: 421 Da) y más hidrofóbicos (AlogP promedio: 3.58) que los fármacos tradicionales (PM promedio: 341 Da, AlogP promedio: 2.61).

Los péptidos cíclicos toman protagonismo como candidatos a inhibidores de las IPPs, pues pueden cumplir con la regla “RO4” de Morelli. Además, son sintéticamente accesibles y generalmente menos tóxicos que los fármacos tradicionales, pues no se acumulan en órganos y los productos de degradación son aminoácidos [23]

## **2.5. Péptidos cíclicos (PCs)**

Los péptidos lineales son conocidos por su capacidad de participar de manera imprescindible en muchos aspectos de la fisiología celular. Su gran diversidad y baja toxicidad hacen de ellos una plataforma prometedora en diversas áreas de la ciencia, la medicina y la tecnología [24]. A pesar de sus prometedores resultados, una de las mayores barreras para el uso terapéuticos de péptidos lineales es su baja estabilidad en condiciones fisiológicas, ya que son fácilmente degradados por enzimas, perdiendo así su actividad biológica [25, 26]. Como solución a este problema se han propuesto diversas alternativas. Una de ellas es la ciclación de la configuración lineal.

Similar a las moléculas pequeñas, los péptidos cíclicos (PCs) son sintéticamente accesibles y pueden ser sujetos a optimización con métodos químicos tradicionales. En este sentido, prácticamente cualquier propiedad fisicoquímica de estos péptidos, por ejemplo, su afinidad de unión, la estabilidad proteolítica y la solubilidad, pueden optimizarse y adaptarse a una función específica [27].

Se ha sugerido que la ciclación de un péptido resulta en una mayor resistencia a endopeptidasas, en comparación con sus análogos lineales [28, 29]. También

quedó demostrado que los PCs incrementan su propensión para formar giros beta, lo cual se cree, aumenta la afinidad de unión [30, 31]. Otra de las ventajas sugeridas es la eliminación de las cargas de los residuos terminales en los PCs, incrementando así la permeabilidad celular. Además, la falta de estos grupos amino y carboxilo terminal los hace menos propensos a la hidrólisis por exopeptidasas [32, 33].

## 2.6. PCs inhibidores de IPPs

Existen grupos de investigación dedicados a desarrollar plataformas experimentales para encontrar PCs inhibidores de IPPs específicas [34] [35]. Una plataforma *in-vitro* que ha mostrado gran potencial es conocida como SICLOPPS (Split-Intein Mediated Circular Ligation Of Peptides and Proteins) [36]. Entre todos los resultados reportados a la fecha, destacan dos trabajos en los que se muestra la identificación y validación *in-vitro* de la actividad inhibitoria de este tipo de moléculas en procesos celulares, por ejemplo, (i) la inhibición de la salida de partículas del Virus de la Inmunodeficiencia Humana (VIH) de células infectadas [37] y (ii) la inhibición del receptor de lipoproteínas de baja densidad [4]. En particular, previamente se utilizó esta plataforma por nuestro grupo de investigación para encontrar PCs que inhibieran de forma específica la IPP de las proteínas de cápside de un conjunto de virus con diferentes topologías (trabajo de Doctorado de Andrade-Medina, M., 2020). A pesar de que dichos trabajos reportan la secuencia de aminoácidos de los PCs identificados (y en pocos casos las constantes de afinidad por los complejos IPP), en ninguno de ellos se ha extendido la investigación hacia el análisis de la estructura terciaria de los PCs. Esta última pieza de información es importante si se desea elucidar el mecanismo molecular de inhibición de la

formación de los complejos P-P. En la actualidad, las estructuras tridimensionales de los PCs pueden ser resueltas utilizando métodos teórico-computacionales como el modelado molecular, o mediante técnicas experimentales como la Cristalografía de Rayos X o Resonancia Magnética Nuclear.

## 2.7. Predicción teórico-computacional de la estructura de PCs

A pesar del gran avance en el desarrollo de métodos de modelado molecular (MM), aún no se cuenta con uno que sea cien por ciento eficaz para la predicción de la estructura terciaria nativa de proteínas. Comparados con los péptidos lineales, los PCs pueden tener grandes barreras de energía para transiciones conformacionales [38]. Para la predicción estructural de proteínas en general, se han desarrollado algoritmos de búsqueda conformacional guiados por mínimos de energía [39, 40]. Una desventaja de estos métodos es que se utiliza una función de energía efectiva con solvatación implícita para evaluar la energía libre de una determinada estructura. Esto limita significativamente la precisión de estos métodos. Por ejemplo, uno de los predictores de estructuras proteínicas más populares por su eficiencia, Rosetta [41], falla al predecir la estructura cristalográfica de la  $\alpha$ -conotoxina GI (una toxina de caracol marino, con 13 residuos y dos puentes de disulfuro) [42]. Sin embargo, con el aumento rápido del poder computacional en los últimos años, la simulación numérica de Dinámica Molecular (DM) de biomoléculas en un solvente explícito en condiciones químicas y termodinámicas fisiológicas se ha vuelto factible [43, 44].

En este trabajo proponemos que la combinación del MM con la DM dará lugar a una predicción certera de la configuración de PC en solución.

### **3. Hipotésis**

La combinación del método teórico de Modelado Molecular con el método computacional de simulaciones numéricas de Dinámica Molecular aumenta el poder predictivo de la determinación de la estructura molecular de péptidos cíclicos en solución acuosa. Esta última, a su vez, permitirá explicar el mecanismo molecular del proceso de inhibición de la formación del homodímero de la proteína de cápside del CCMV.

### **4. Objetivo General**

Predecir la conformación de péptidos cíclicos en solución acuosa en condiciones fisiológicas y utilizar dicha predicción para estudiar el mecanismo de inhibición de la interacción proteína-proteína de la CP del CCMV.

### **5. Objetivos Específicos**

1. Generar una configuración tridimensional de la estructura molecular de un conjunto de PCs a partir de su secuencia de aminoácidos.
2. Realizar un muestreo de la diversidad conformacional de los PCs y encontrar la configuración más estable en condiciones fisiológicas en medio acuoso.
3. Estudiar el posible mecanismo de acción molecular de los PCs en la inhibición de la IPP de las CP del CCMV.

## 6. Metodología

### 6.1. PCs estudiados

En nuestro grupo de investigación se ha trabajado con el *Cowpea Chlorotic Mottle Virus* (CCMV) como modelo experimental. Nuestros hallazgos recientes llevaron a la detección experimental de ocho secuencias de PCs con acción inhibitoria de la interacción CP-CP del CCMV (Tabla 1). Para su identificación se utilizó una biblioteca tipo SICLOPPS en conjunto con la plataforma experimental RTHS [35] (trabajo de Doctorado de M. Andrade-Medina, en colaboración con el grupo de investigación del Dr. Ali Tavassoli, 2020). Dichas secuencias fueron utilizadas como punto de partida en este estudio de predicción molecular teórica.

<b>PC.C1:</b> c-SPSFTF	<b>PC.C5:</b> c-SFVVMLY
<b>PC.C2:</b> c-SGYTVP	<b>PC.C6:</b> c-SNSFVY
<b>PC.C3:</b> c-SFRIYVF	<b>PC.C7:</b> c-SRDLLM
<b>PC.C4:</b> c-SFWVLAV	<b>PC.C8:</b> c-STWWVT

**Tabla 1:** Secuencia de los péptidos cíclicos inhibidores de la interacción de proteínas de cápside del CCMV previamente identificados experimentalmente (trabajo de Doctorado de Andrade-Medina, M., 2020).



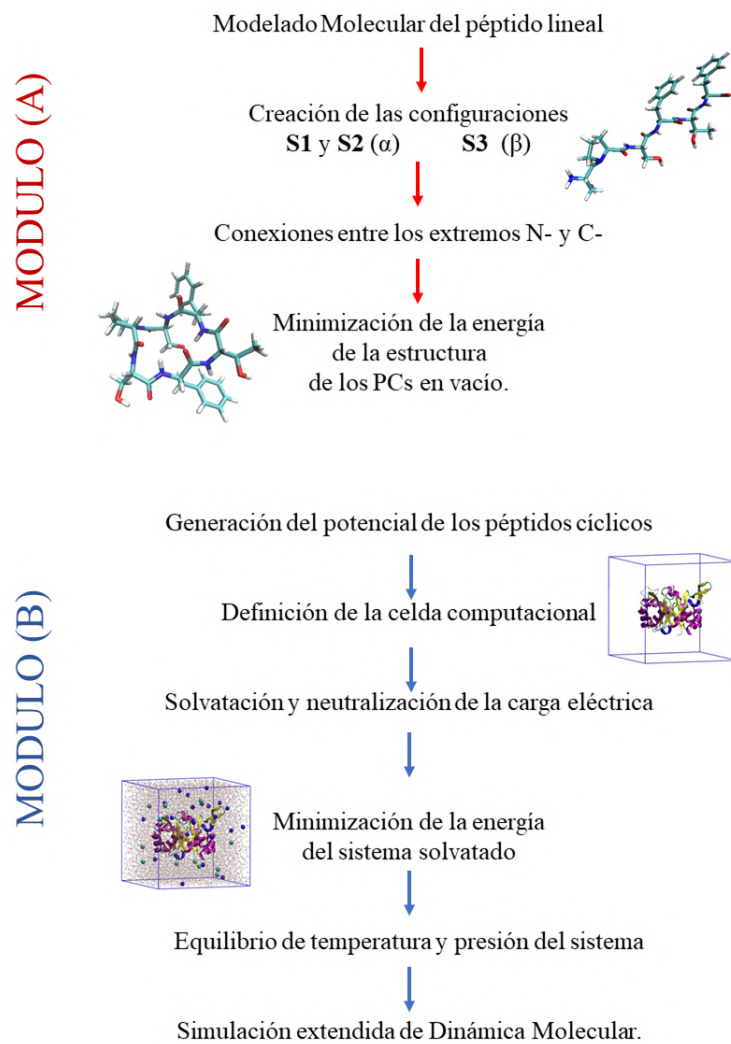
## 6.2. Predicción estructural de PCs *in-silico*

Para lograr predecir una estructura molecular de los PCs, estable en condiciones fisiológicas, se propuso usar un protocolo de dos módulos que integre el modelado molecular (MM) y las simulaciones extendidas de dinámica molecular (DM). Dicha propuesta se puede observar en la Fig. 4. Así, para cada secuencia de aminoácidos descrita en la Tabla 1 se construyó la estructura tridimensional para ser equilibrada en condiciones cercanas a las fisiológicas (solvente acuoso con NaCl, temperatura ambiente, y presión de 1 atmósfera).

### 6.2.1. Modelado molecular

El Modelado Molecular se realizó utilizando el software *UCSF-Chimera* [45]. El diseño experimental original contempló dos conformaciones estructurales para cada PC. La diferencia entre ambas conformaciones son los valores de ángulos peptídicos  $\Phi$  y  $\Psi$ , congruentes con una estructura tipo  $\alpha$ -hélice en un caso, o de  $\beta$ -laminar en el otro. Accidentalmente, al inicio del trabajo se crearon y simularon las dos conformaciones con una definición de  $\alpha$ -hélice (*S1* y *S2*). El error se corrigió al generar una tercer conformación (*S3*) con la definición de ángulos peptídicos  $\Phi$  y  $\Psi$  congruentes con una estructura tipo  $\beta$ -laminar. Se decidió incluir en el análisis los datos que ya se habían generado para *S2*, aprovechando esa información para extender el estudio. Los detalles metodológicos del modelado se describen a continuación.

*UCSF-Chimera* es una aplicación de escritorio. A través de su interfaz gráfica se construyó la versión lineal de cada péptido desde la sección *Structure Editing*



**Figura 4:** Propuesta teórico-computacional para la predicción con alta precisión de la estructura tridimensional de péptidos cíclicos en la cual se integra el modelado molecular (A) y la dinámica molecular (B).

del menú *Tools*. Ahí se seleccionó la opción *Build Structure* y posteriormente *Peptide*. Se indicó la secuencia del péptido a construir y al hacer clic en el botón *Apply* se abrió una ventana donde se indicaron los ángulos diedros de cada uno de los enlaces peptídicos. Es en esta última etapa donde se les asignan las características  $\alpha$ -hélice o  $\beta$ -lamina a los péptidos. Para ciclar a los péptidos, primero se seleccionaron a los átomos entre los cuales se creará el enlace peptídico. Esto se hizo en la sección *Atom Specifier* del menú *Select*. Ahí seleccionamos al N y C terminales. Posteriormente, en el submenú *Build Structure* se cambió a la opción *Adjust Bonds* en la parte superior de la ventana. Ahí se seleccionó *All Possible* seguido de *Add*. Existe un enlace extra en el sitio de la ciclación, el cual se debe eliminar. Esto se logró desde la sección *Atom Specifier* del menú *Select*. Cuando se tuvo listo el péptido cíclico, se procedió a agregar hidrógenos, carga eléctrica y a minimizar la energía desde el submenú *Structure Editing*. Todas las conformaciones se guardaron en archivos en formato PDB.

### 6.2.2. Dinámica molecular

Las simulaciones de DM se realizaron usando GROMACS 2018.2 [46, 47] y el potencial atómico OPLS-AA [48] para describir las interacciones intra- e intermoleculares de los PCs. Se utilizó el potencial TIP4P [49] para describir explícitamente a las moléculas de agua. Se siguió un procedimiento estándar con valores de parámetros reportados previamente en la literatura. Cada una de las estructuras, *S1*, *S2* y *S3*, se colocaron por separado en el centro de una celda computacional cúbica que posteriormente se llenó con moléculas de agua explícita. Las dimensiones de la caja fueron tales que la distancia mínima entre cualquier átomo del PC

y las paredes de la celda fue no menor a 1.0 nm. Se agregaron iones  $Na^+$  y  $Cl^-$  en proporciones tales que neutralizaron la carga total del sistema con una concentración final de 100 mM. Para eliminar los malos contactos entre cualquier átomo se usó el algoritmo *steepest descent*. La equilibración de moléculas de solvente a una temperatura de 300K se hizo utilizando condiciones isocóricas-isobáricas (NTV) durante 100 ps. Posteriormente, la densidad del sistema se equilibró en condiciones isobáricas-isotérmicas (NPT) durante 100 ps a 1 bar. Las simulaciones de producción se hicieron en el ensamble NPT a 300K y 1 bar por 5  $\mu s$  (5,000 ns). En el estándar actual, este tiempo de simulación es de 50 a 500 veces más extenso que lo que se acepta como razonable, pero viable en la infraestructura de cómputo de alto rendimiento de nuestro grupo de investigación. Para mantener la temperatura del sistema simulado constante se usó el termostato *V-reescale* con un acoplamiento de 0.1 ps. Para evitar problemas de tipo solvente caliente - soluto frío, el PC y el solvente fueron acoplados a termostatos por separado. La presión fue regulada con el barostato isotrópico *Parrinello-Rahman* con un tiempo de acoplamiento de 2.0 ps y una compresibilidad de  $4.5 \times 10^{-5} \text{bar}^{-1}$ . El algoritmo LINCS [50] se usó para restringir los enlaces que involucran hidrógeno a sus valores de equilibrio. Las interacciones no covalentes (Lennard-Jones y electrostáticas) se truncaron a 1.0 nm. Las interacciones electrostáticas de largo alcance más allá de la distancia de corte se calcularon utilizando el método Ewald de malla de partículas (PME) [51] con un espaciado de Fourier de 0.16 nm y una interpolación cúbica de orden 4. La dinámica del sistema dependiente del tiempo se desarrolló utilizando el integrador *Leap-Frog* con un paso de tiempo de 2 fs (0.000002 ns). Las trayectorias de simulación se guardaron cada 50 ps (0.05 ns) para realizar los análisis posteriores. Al final de cada simulación de DM, se usaron las trayectorias para realizar un análisis

de la evolución temporal de las estructuras  $S1$ ,  $S2$  y  $S3$  de cada PC mostrado en la Tabla 1. Se considerará que la convergencia estructural se logra cuando las simulaciones de DM para las diversas conformaciones de PCs proporcionaron resultados con una similitud menor a un umbral dado.

Este protocolo se realiza a través de comandos en línea capturados manualmente dentro de una terminal de Linux. Los programas de dinámica molecular se ejecutaron utilizando el paradigma de cómputo en paralelo en un cluster de alto rendimiento remoto. En el *Apéndice* se describen los comandos del pipeline necesarios para la preparación, producción, postprocesamiento y análisis de las trayectorias resultantes.

### 6.3. Acoplamiento Molecular

Los cálculos de Acoplamiento Molecular, o *Docking*, se realizaron en la plataforma en línea *CB-DOCK* (Cavity-based Blind Docking) [52]. Dicha plataforma hace uso de un método que identifica las cavidades en la superficie molecular y los prueba como sitios de unión potenciales, para posteriormente realizar el acoplamiento molecular entre los dos elementos del complejo usando la función de energía definida por AutoDock Vina [53].

CB-Dock realiza un acoplamiento ciego sólo en sitios identificados (cavidades), en lugar de hacer una búsqueda sobre toda la superficie de la proteína. Por lo tanto, el primer paso es detectar los posibles sitios de unión o cavidades. El paso que sigue es clasificar a dichas cavidades de acuerdo a su tamaño, pues los sitios de unión al ligando generalmente son las cavidades mas grandes. Posteriormente,

se establece automáticamente el centro de acoplamiento y se ajusta el tamaño de la caja de muestreo. Estos parámetros son necesarios para el acoplamiento molecular con AutoDock Vina (centro y tamaño). Una vez finalizado el proceso de acoplamiento, se reclasifican las conformaciones de acuerdo con la puntuación energética del acoplamiento y el tamaño de la cavidad (Dock y Rerank).

Esta plataforma necesita dos archivos de entrada para realizar el acoplamiento. La estructura molecular de la proteína, que debe de estar en formato PDB y la estructura molecular del ligando que puede estar en diversos formatos (MOL2, MOL, SDF, o PDB).

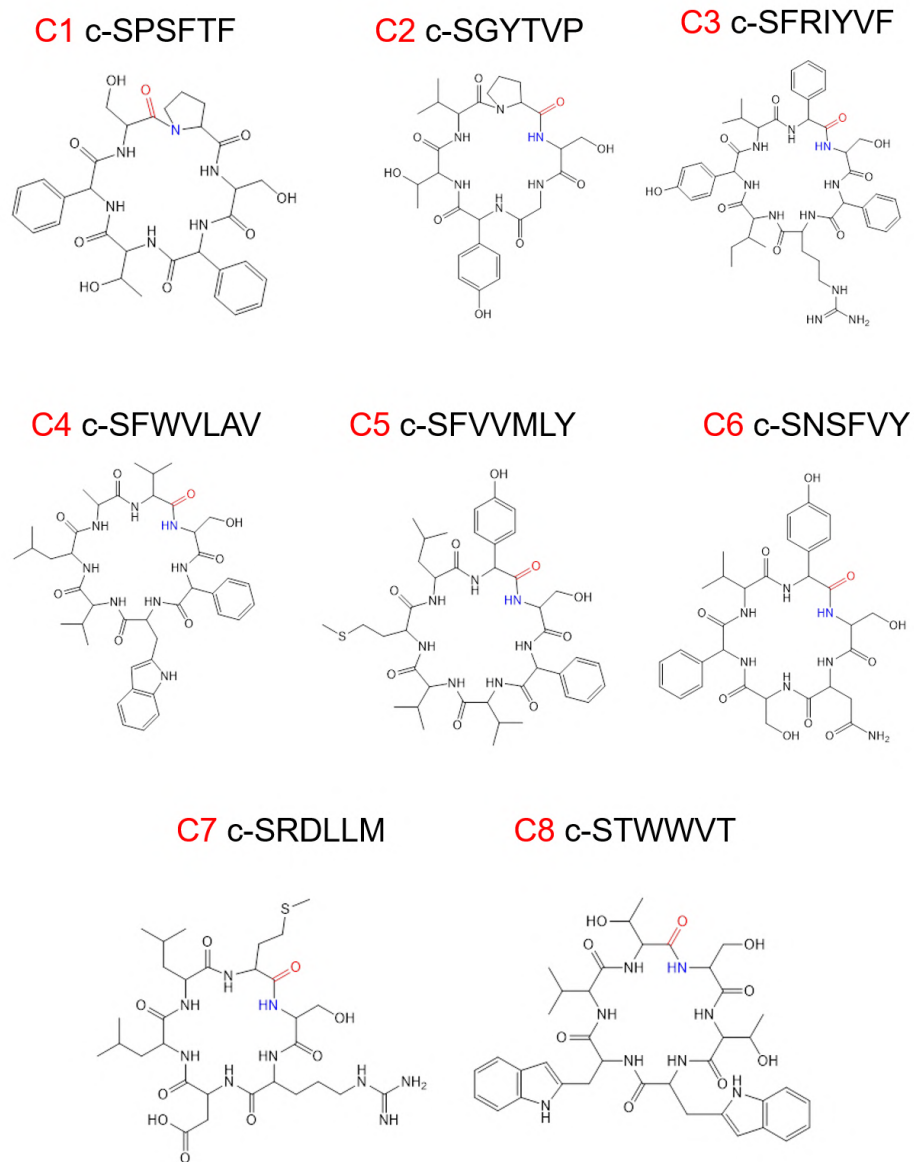
## 7. Resultados y Discusión

Como se mencionó en la sección 6.1 de la metodología, como primer fase se trabajó en la predicción de la estructura molecular de 8 PCs con actividad inhibitoria de la interacción de proteínas de cápside del CCMV. La secuencia y la estructura química de dichas moléculas se muestran en la Figura 5. Las 8 estructuras péptidicas se componen de una secuencia de 6 o 7 residuos. Todos ellos contienen al menos una Serina en su estructura, y a excepción del PC-C7, todos tienen, al menos, un residuo aromático en la secuencia.

### 7.1. Modelado Molecular

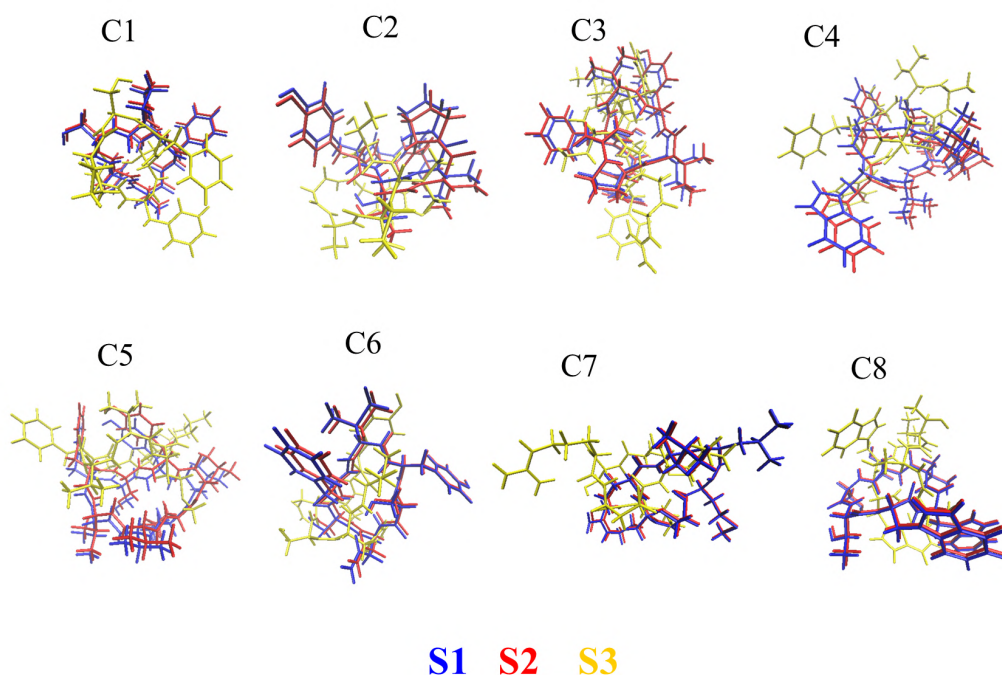
Siguiendo la metodología propuesta en el punto 6.2.1 se obtuvieron las conformaciones  $\alpha$  (S1, S2) y  $\beta$  (S3) de cada uno de los 8 PCs, dando como resultado un total de 24 conformaciones. La información estructural se almacenó en el formato propuesto por el *Protein Data Bank* conocido como *PDB*, el cual almacena las coordenadas tridimensionales en el sistema Cartesiano de cada uno de los átomos que forman parte de la molécula. En la Figura 6 se muestra el trío de conformaciones *S1*, *S2* y *S3* de los 8 PCs obtenidas por Modelado Molecular. En dicha figura se superponen las tres conformaciones de cada péptido para la mejor apreciación de las similitudes o diferencias entre ellas.

Para cuantificar la similitud estructural entre *S1*, *S2* y *S3* se hizo uso de una métrica conocida como *TM-score*. Esta métrica compara dos estructuras moleculares entre sí y les asigna valores entre (0-1], donde 1 indica que se trata de la misma estructura, valores  $> 0.5$  indican el mismo plegado y  $< 0.17$  indican que las



**Figura 5:** Estructura química de los 8 PCs con propiedad inhibitoria de la IPP de proteínas de cápside del CCMV.





**Figura 6:** Estructura molecular de las conformaciones S1, S2 y S3 de los 8 PCs con propiedad inhibitoria de la IPP de proteínas de cápside del CCMV.

estructuras comparadas no se parecen en absoluto [54]. En la Tabla 2 se muestran los valores del *TM-score* para los pares de conformaciones S1 y S2.

Dado que los resultados del *TM-score* nos confirman que las conformaciones S1 y S2 tienen el mismo plegado  $\alpha$ , la comparación con la conformación S3 se hizo sólo con una de ellas. En la Tabla 3 se muestran los valores para los pares de conformaciones entre S1 y S3.

<b>conformaciones Comparadas</b>	<b>TM-score</b>
C1-S1 vs. C1-S2	0.99
C2-S1 vs. C2-S2	1.00
C3-S1 vs. C3-S2	1.00
C4-S1 vs. C4-S2	0.99
C5-S1 vs. C5-S2	0.99
C6-S1 vs. C6-S2	0.99
C7-S1 vs. C7-S2	1.00
C8-S1 vs. C8-S2	1.00

**Tabla 2:** Valores de *TM-score* entre las conformaciones S1 y S2 de los 8 péptidos cíclicos (C1 - C8).

<b>conformaciones Comparadas</b>	<b>TM-score</b>
C1-S1 vs. C1-S3	0.18
C2-S1 vs. C2-S3	0.30
C3-S1 vs. C3-S3	0.26
C4-S1 vs. C4-S3	0.26
C5-S1 vs. C5-S3	0.29
C6-S1 vs. C6-S3	0.39
C7-S1 vs. C7-S3	0.50
C8-S1 vs. C8-S3	0.47

**Tabla 3:** Valores de *TM-score* entre las conformaciones S1 y S3 de los 8 péptidos cíclicos (C1 - C8).

Como se puede apreciar en los valores de *TM-score* de la Tabla 3, la conformación S1, y en consecuencia S2, son diferentes a la conformación S3. A través de tal análisis comparativo verificamos que efectivamente se tienen dos conformaciones diferentes entre sí, esto es, distantes una de otra en el espacio conformacional. Habiendo confirmado dichas similitudes y diferencias se procedió con la fase de dinámica molecular.

## 7.2. Dinámica molecular

Con la información estructural en formato *PDB* obtenida por Modelado Molecular se construyó una topología independiente para cada una de las tres conformaciones cíclicas de los 8 PCs. Dicha topología, por definición, contiene la información necesaria que representa a una molécula dada en una simulación de dinámica molecular (DM), por ejemplo, ángulos de enlace, tipos y cargas de átomos, así como los valores particulares de los parámetros de interacción intra- e inter-moleculares.

Para construir la topología de los PCs partiendo de una descripción lineal se siguió el siguiente procedimiento. El primer paso es modificar el archivo PDB que contiene la información de cada PC. Utilizando un editor de texto, se agregó un residuo extra en cada uno de los extremos terminales del péptido. El residuo adicional agregado al inicio de la secuencia se toma del final, y el que es agregado al final debe de ser tomado del inicio de la secuencia. Así, si la secuencia original es SPSFTF, después de agregar los residuos adicionales quedará de la siguiente manera **F**SPSFT**F**S (en negritas se resaltan los residuos adicionales). El siguiente paso es reenumerar los residuos, el primero será el 0 y el último continuará con la numeración ascendente. A continuación se ejecuta el siguiente comando

```
gmx pdb2gmx -f PCx.pdb -o PCx_mod.pdb
```

usando el nombre del archivo que contine al PC modificado. Con esto se crea un archivo de topología con la extensión .top

El siguiente paso es ejecutar el script en lenguaje Perl

```
gmx.makecyclictopy.pl topfile
```

el cual puede ser descargado de internet, donde *topfile* es el archivo de topología creado en el paso anterior. El script creará una nueva topología, *topol\_cyc.top*. El paso que resta es eliminar los residuos agregados al PDB en el primer paso. Es importante también modificar el número de átomos totales en el mismo archivo PDB.

Con el archivo de topología de las estructuras cíclicas se continuó con la fase de preparación de la DM. Dicha fase tiene como objetivo definir el volumen del sistema, la concentración de iones, así como la temperatura y presión en las que se llevará a cabo la simulación computacional. Dichas condiciones se muestran en la Tabla 4.

**Tabla 4:** Condiciones químicas y termodinámicas de simulación para la Dinámica Molecular de los 8 PCs en tres conformaciones iniciales (*S1*, *S2* y *S3*). En la columna **Soluto** se indica el número de amino ácidos y átomos totales de cada péptido cíclico. En la columna **Solvente** se indica el número de moléculas de agua - átomos de Na<sup>+</sup> - átomos de Cl<sup>-</sup> / total de átomos en la solución salina. Las unidades del volumen de la caja de simulación son nm<sup>3</sup>. La eficiencia computacional, HPC, se mide en ns/día. Se realizaron 5,000 ns en total para cada sistema. En todos los casos la temperatura se estableció en 300 °K y la presión en 1 atm.

PC	Soluto	Solvente	Volumen	HPC
C1-S1	6-90	1,373-3-3 / 5,588	3.48 <sup>3</sup>	137.9
C1-S2	6-90	1,373-3-3 / 5,588	3.48 <sup>3</sup>	137.9
C1-S3	6-90	1363-3-3 / 5,548	3.48 <sup>3</sup>	137.9
C2-S1	6-83	1,297-2-2 / 5,275	3.41 <sup>3</sup>	144.5
C2-S2	6-83	1,301-2-2 / 5,291	3.84 <sup>3</sup>	144.3
C2-S3	6-83	1,007-2-2 / 4,115	3.48 <sup>3</sup>	174.7
C3-S1	7-131	1,494-3-4 / 6,114	3.64 <sup>3</sup>	127.5
C3-S2	7-131	1,494-3-4 / 6,114	3.63 <sup>3</sup>	127.6
C3-S3	7-131	2,243-4-5 / 9,112	4.06 <sup>3</sup>	90.2
C4-S1	7-116	1,553-3-3 / 6,334	3.69 <sup>3</sup>	123.9
C4-S2	7-116	1,553-3-3 / 6,334	3.69 <sup>3</sup>	123.8
C4-S3	7-116	1,685-3-3 / 6,862	3.84 <sup>3</sup>	114.6
C5-S1	7-120	1,478-3-3 / 6,038	3.62 <sup>3</sup>	128.7
C5-S2	7-120	1,479-3-3 / 6,042	3.62 <sup>3</sup>	128.7
C5-S3	7-120	1,689-3-3 / 6,882	3.83 <sup>3</sup>	114.3
C6-S1	6-93	1,496-3-3 / 6,083	3.62 <sup>3</sup>	128.4
C6-S2	6-93	1,497-3-3 / 6,087	3.63 <sup>3</sup>	128.3

*(Continúa en la página siguiente)*

(Viene de la página anterior)

<b>PC</b>	<b>Soluto</b>	<b>Solvente</b>	<b>Volumen</b>	<b>HPC</b>
C6-S3	6-93	1,454-3-3 / 5,915	3.58 <sup>3</sup>	130.4
C7-S1	6-102	1,687-3-3 / 6,856	3.74 <sup>3</sup>	115.4
C7-S2	6-102	1,686-3-3 / 6,852	3.74 <sup>3</sup>	115.7
C7-S3	6-102	1,543-3-3 / 6,280	3.68 <sup>3</sup>	123.7
C8-S1	6-103	1,373-3-3 / 5,601	3.50 <sup>3</sup>	137.6
C8-S2	6-103	1,373-3-3 / 5,601	3.49 <sup>3</sup>	137.3
C8-S3	6-103	1,339-3-3 / 5,465	3.47 <sup>3</sup>	139.5

(Fin de Tabla 4)

Los cálculos de la fase de producción se hicieron en dos clústers de cómputo de alto rendimiento en nodos de 24 núcleos de procesamiento cada uno. El primero perteneciente al Laboratorio de la Diversidad Biomolecular del Cinvestav Unidad Monterrey (*bmdhpc*) y el segundo al Centro de Investigación en Matemáticas en Guanajuato (*Insurgente*). La información procedente de dichos cálculos se almacenó en archivos digitales conocidos como *trayectorias*, los cuales contienen las coordenadas atómicas de todas las moléculas del sistema simulado en función del tiempo (para cada sistema de PCs). Estos datos son una serie de conjuntos de coordenadas correspondiente a un instante o *frame* en el tiempo. Dado que el tiempo de simulación para cada sistema fue de 5  $\mu$ s (5,000 ns) y el  $\delta t$  de integración fue de 0.002ps (0.000002 ns), en total se realizaron  $2.5 \times 10^9$  pasos de simulación. Se almacenó la conformación instantánea cada 200,000 pasos de integración, es decir, cada 400ps (0.4 ns) de simulación, lo que da como resultado final 12,500 *frames*

para su posterior análisis. El tiempo real de las simulaciones fue de seis meses, con un costo computacional equivalente a 41,550 días (113.8 años) de cálculo continuo. Esta reducción en el tiempo real se debe al uso del paradigma computacional de cálculo en paralelo.

Para poder analizar y visualizar las trayectorias de los PCs correctamente es necesario hacer un procesamiento de los datos en el que se tome en cuenta la difusión del PC en la solución acuosa. Esto es importante ya que al ser simulaciones de tiempo muy largos, las moléculas van a recorrer distancias mayores a las dimensiones de la caja de simulación (condiciones periódicas de frontera). En dicho post-procesamiento se realiza una traslación de coordenadas de todo el sistema de forma que el centro de masa del PC coincida con en el centro de la caja de simulación en todos los *frames* de la trayectoria. Con ello es posible analizar al PC sin las rotaciones y traslaciones estocásticas causadas por su interacción con el solvente.

Con apoyo del software *VMD* [55] fue posible visualizar la trayectoria dinámica de cada sistema. De esta manera, se pudo estimar cuál podría ser la conformación de los PCs en solución. Pero, para poder predecir de una manera cuantificable cuál es la conformación o conformaciones moleculares más probables, se realizaron análisis estructurales que permitieron comparar la evolución de las conformaciones S1, S2 y S3 de cada PC. Dichos análisis se muestran y discuten en la siguiente sección.

### 7.3. Análisis Estructural

El planteamiento original de este proyecto establece que dos estructuras formadas por los mismos átomos y conectadas por los mismos enlaces pero con dos conformaciones estructurales iniciales diferentes que se someten a las mismas condiciones de simulación eventualmente llegaran a un estado de convergencia. En dicho estado, sería posible observar la conformación más estable, y por ende, la que con mayor probabilidad sería observable en solución. Para corroborar o descartar dicho planteamiento, se practicaron los siguientes análisis estructurales sobre los datos contenidos en las trayectorias obtenidas por dinámica molecular.

- **Desviación Cuadrática Media (RMSD por sus siglas en ingles).** Medida cuantitativa utilizada para comparar la similitud estructural entre un par de moléculas. Mientras más cercano a cero sea el valor del RMSD, más se parecerán las dos estructuras entre sí. Dados dos conjuntos  $v$  y  $w$  cada uno con  $n$  átomos, esta métrica se define como:

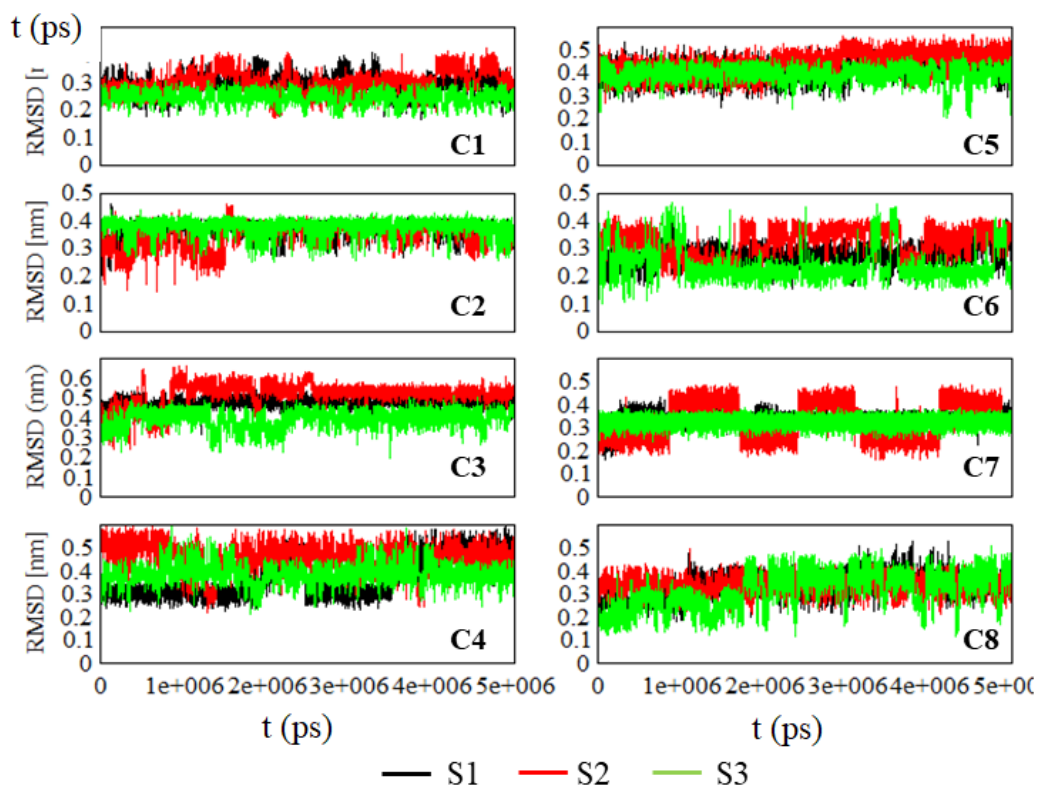
$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n ((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2)} \quad (1)$$

- **Clusterización.** Agrupa y ordena las conformaciones obtenidas en el muestreo de DM de una molécula de un péptido de acuerdo a las similitudes que haya entre ellas (utilizando el RMSD).
- **Matriz de Convergencia.** Compara todos los frames de dos trayectorias diferentes y cuantifica qué tan parecidas son en un instante de tiempo determinado (utilizando el RMSD).



### 7.3.1. Equilibrio dinámico conformacional: RMSD(t)

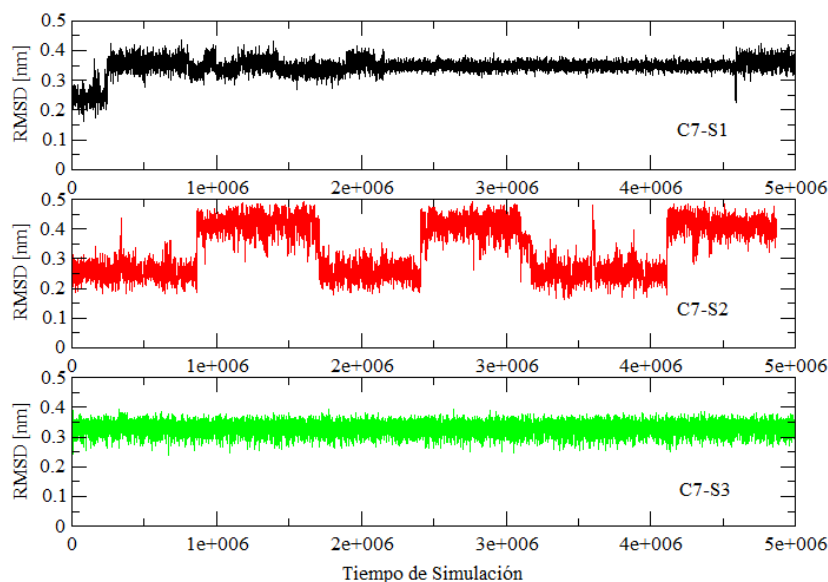
El análisis del RMSD(t) permitió observar las fluctuaciones en las conformaciones individuales S1, S2 y S3 de los 8 PCs a lo largo del tiempo de simulación. Lo que se esperaría observar es que los valores de RMSD(t) de las tres conformaciones permanezcan estables después de un periodo de simulación dado.



**Figura 7:** RMSD(t) de las conformaciones S1, S2 y S3 de los PCs.

En la Figura 7 se pueden observar los valores de RMSD(t) de las trayectorias de las tres conformaciones de los 8 PCs. Si utilizamos la misma notación de la definición de  $RMSD$ , aquí  $v(t)$  es cada una de las conformaciones en la trayectoria

del PC, y  $w$  es la conformación de referencia, que en este caso es la conformación inicial de la simulación. En general, los valores de  $\text{RMSD}(t)$  de cada una de las trayectorias se comportan de dos formas diferentes, una estable y una que no lo es. Tomemos como ejemplo al PC-C7 (Figura 8), en cuyos gráficos de  $\text{RMSD}(t)$  se observan diferencias obvias para cada conformación.

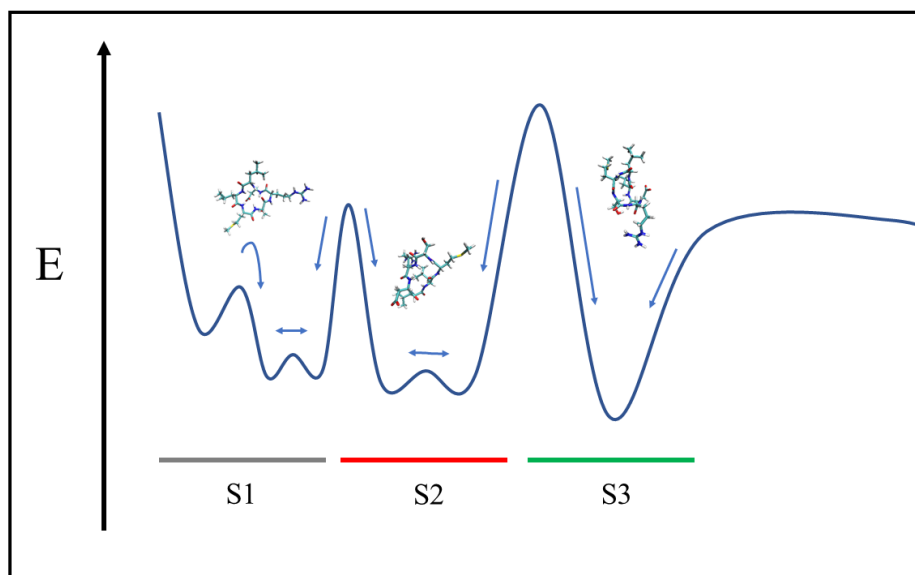


**Figura 8:**  $\text{RMSD}(t)$  de las conformaciones S1, S2 y S3 del PC-C7.

Para C7-S1 y C7-S2 se observan 'brincos' en el comportamiento de la función  $\text{RMSD}(t)$ , mientras que para C7-S3 el comportamiento es estable a lo largo de la simulación. Estos brincos están relacionados con la presencia de diversos mínimos locales de energía. Así, para los casos C7-S1 y C7-S2 se observa que hay al menos dos mínimos locales que los PC visitan. Para el otro caso, C7-S3, la falta de 'saltos'

sugiere que dicha conformación encontró un mínimo de energía más profundo en donde permaneció durante toda la trayectoria (Fig. 9).

Para el caso de C7 las diferencias entre las conformaciones S1, S2 y S3 son muy claras, pero hay PCs cuyas variaciones no son tan aparentes pues el rango de valores del RMSD(t) es más amplio. Para analizar dichas fluctuaciones y establecer si hay o no diversos mínimos de energía locales fue necesario hacer uso de otro tipo de estudio conocido como *clusterización*, cuyos resultados se discuten en la siguiente sección.



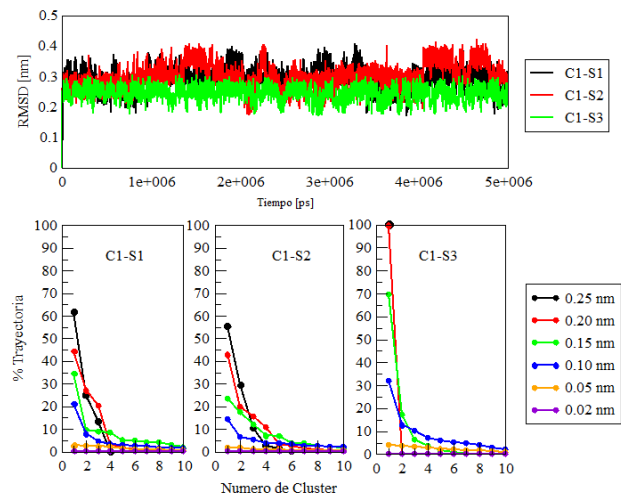
**Figura 9:** Interpretación de los resultados de RMSD(t) del PC-C7 en términos de la superficie de energía conformacional.

### 7.3.2. Análisis por clusters

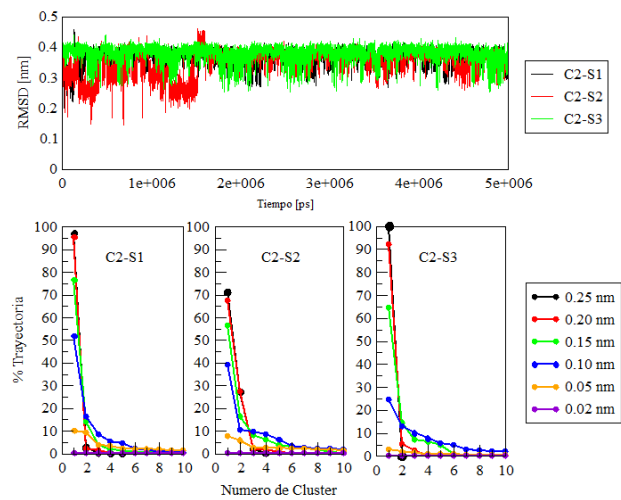
Una de las causas propuestas para explicar los saltos observados en el RMSD(t) de las trayectorias, es la presencia de varios mínimos de energía locales. Para aportar más información y responder esta cuestión, se practicó un análisis por clusters.

Este estudio permitió agrupar las conformaciones de toda la trayectoria en diferentes grupos o clusters, asignando un umbral de semejanza entre ellas. Dicho umbral es arbitrario y se conoce como *Cut-Off*. Si el RMSD entre dos conformaciones es menor al valor del *Cut-Off* definido, entonces esas conformaciones pertenecerán al mismo cluster. Caso contrario, si el RMSD resultara ser mayor, entonces cada una pertenecería a un cluster diferente. Mientras más pequeño sea el valor del *Cut-Off* los clusters contendrán conformaciones más parecidas entre ellas, pero esto hace que el número de clusters aumente considerablemente. En el caso extremo, se obtendrían 12,500 clusters cuando el *Cut-Off*=0 en nuestro caso. Por otro lado, si el valor del *Cut-Off* es demasiado grande habrá menos clusters, pero se corre el riesgo de que la semejanza entre las conformaciones no sea la suficiente. En el caso extremo, se obtendría un solo cluster con 12,500 conformaciones. La finalidad es encontrar un valor de *Cut-Off* razonable, en el que la semejanza entre conformaciones y el número de clusters este balanceada.

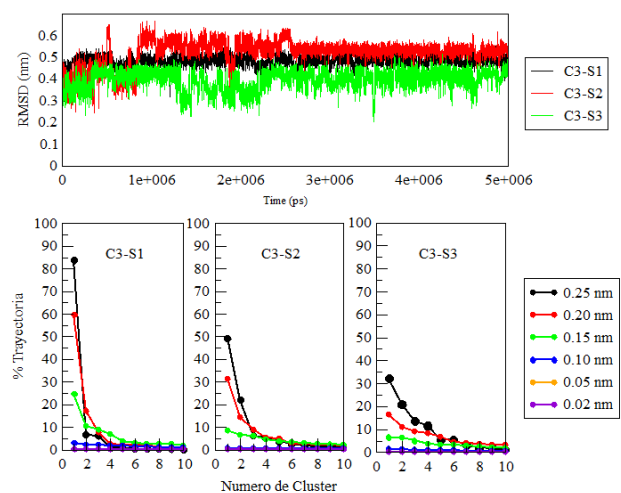
En las Figuras 10 a la 17 se reportan los resultados del análisis de clusters y los RMSDs en función del tiempo de las tres conformaciones de los 8 PCs. Los valores de *Cut-Off* que se analizaron fueron 0.25nm, 0.20nm, 0.15nm, 0.10nm, 0.05nm y 0.002nm. En las figuras se reporta el porcentaje de la trayectoria de los 10 clusters más grandes obtenidos en cada caso. Los clusters se ordenan de mayor a menor, de izquierda a derecha.



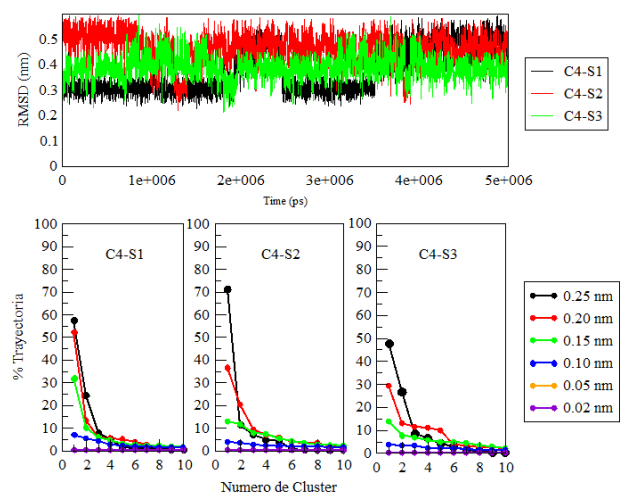
**Figura 10:** RMSD(t) y análisis de clusters de las conformaciones C1



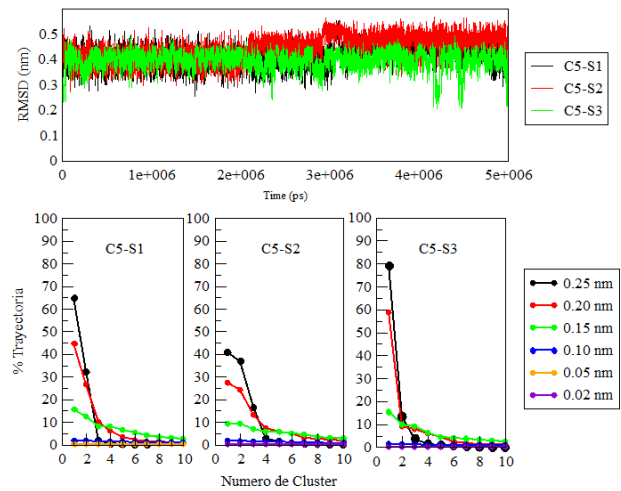
**Figura 11:** RMSD(t) y análisis de clusters de las conformaciones C2



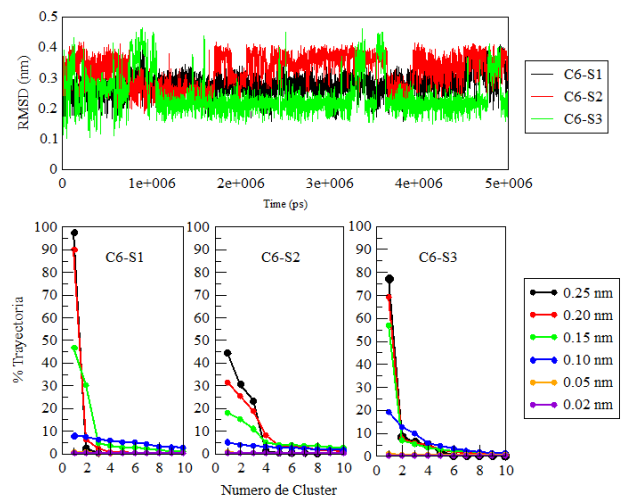
**Figura 12:** RMSD(t) y análisis de clusters de las conformaciones C3



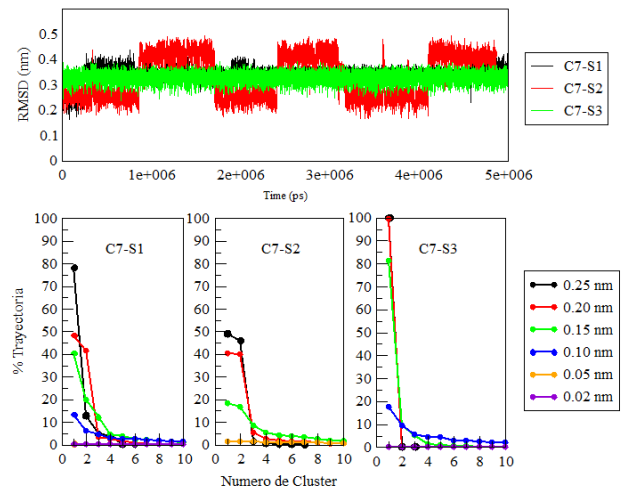
**Figura 13:** RMSD(t) y análisis de clusters de las conformaciones C4



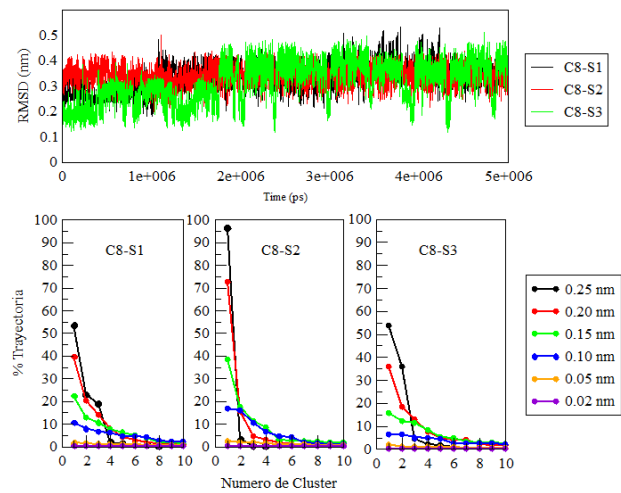
**Figura 14:** RMSD(t) y análisis de clusters de las conformaciones C5



**Figura 15:** RMSD(t) y análisis de clusters de las conformaciones C6



**Figura 16:** RMSD(t) y análisis de clusters de las conformaciones C7



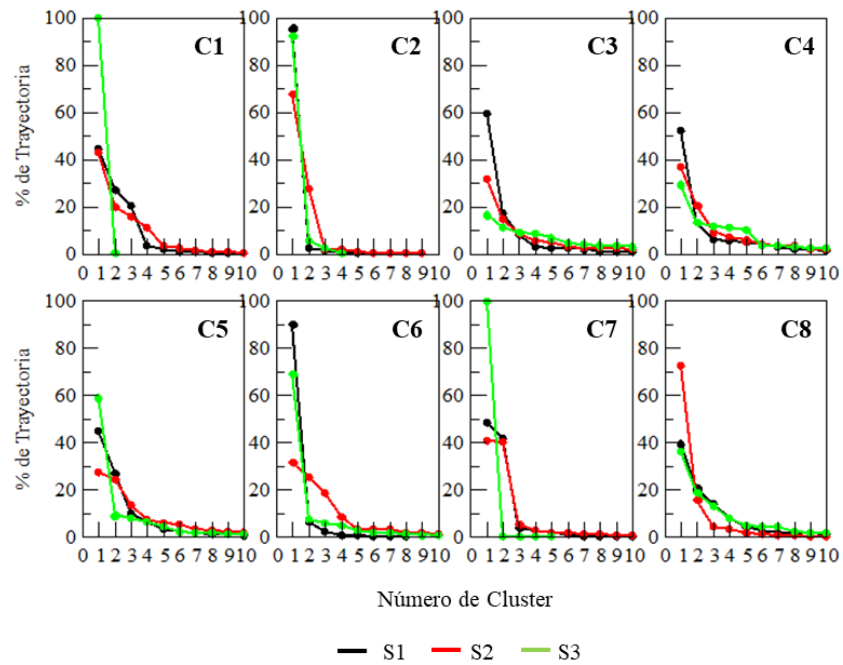
**Figura 17:** RMSD(t) y análisis de clusters de las conformaciones C8



En las figuras (10 - 17) se puede observar que los valores de *Cut-Off* que reportan un porcentaje alto de conformaciones en los primeros tres clusters fueron 0.25nm, 0.20nm y 0.15nm. Sin embargo, existen casos como C1-S3 y C7-S3, donde el uso del *Cut-Off*=0.25nm agrupó al 100 % de las conformaciones de la trayectoria. Por otro lado, cuando se usaron valores de *Cut-Off* de 0.05nm y 0.02nm hubieron tantos clusters como conformaciones en la trayectoria, es decir, cada cluster tenía una sola conformación. Con esto en mente, se determinó que el valor de *Cut-Off* óptimo para trabajar sería 0.20nm. En la Figura 18 se observan los primeros 10 clusters de las trayectorias de los 8 PCs utilizando 0.20nm como valor de *Cut-Off*.

Volvamos a la Figura 18 y tomemos de nuevo al PC C7 como ejemplo. Si observamos la distribución porcentual de las trayectorias por cluster, notaremos que las conformaciones S1 y S2 (datos en color negro y rojo), están divididas en dos clusters principalmente, cada uno de ellos con aproximadamente el 40 % de los datos de su trayectoria respectiva. Estos resultados concuerdan con las fluctuaciones observadas en los gráficos de RMSDs en función del tiempo reportados en la sección anterior (Figura. 8). Para el caso de C7-S3 también hay una concordancia en los resultados, pues si observamos los clusters (datos en color verde) podremos notar que casi el 100 % de las conformaciones está en el cluster 1, aspecto que tiene relación con la nula presencia de saltos en su respectivo RMSD(t). Esto concuerda con la interpretación de los resultados con la imagen de la forma de la superficie energética (Fig. 9).

Si hacemos este mismo análisis para los demás PCs, podremos notar la presencia de diversos clusters, lo que confirma en cierta medida la existencia de diversos

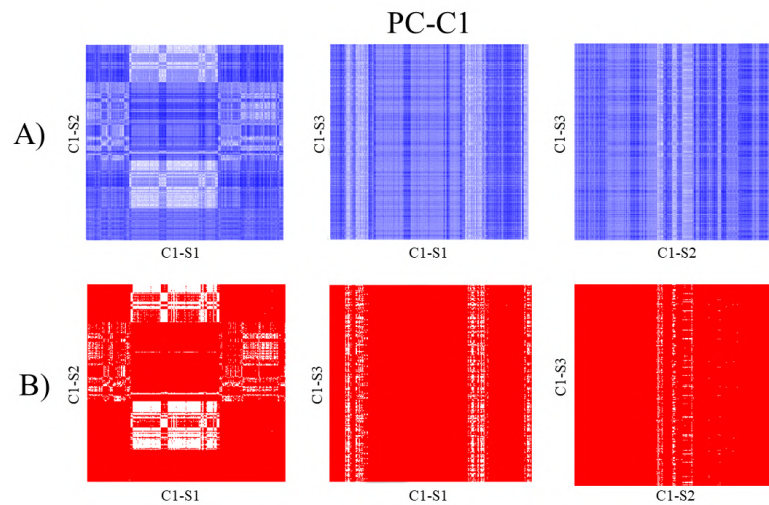


**Figura 18:** Cluster de la serie C de PCs, *Cut-Off* = 0.20nm

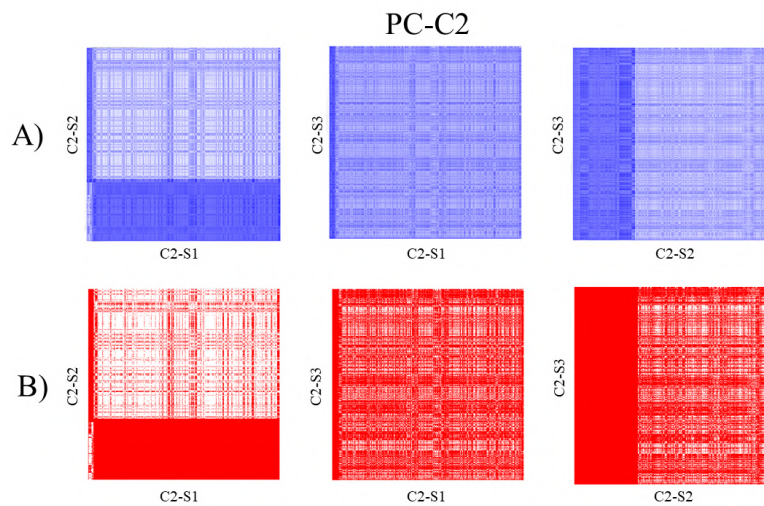
mínimos locales de energía. También resulta interesante observar que las conformaciones  $\beta$  (S3), parecen tener mínimos de energía más profundos (o barreras más altas) que los de las conformaciones  $\alpha$  (S1 y S2). Con los análisis previos podemos decir que los PCs experimentan cambios conformacionales que están determinados por los mínimos de energía locales. Sin embargo, no podemos hablar aún de una posible convergencia estructural entre las conformaciones de cada uno de los PCs. Para ello se debe comparar la evolución de las trayectorias entre sí, esto se logra con un análisis de matrices, el cual se discute a continuación.

### 7.3.3. Matrices de RMSD

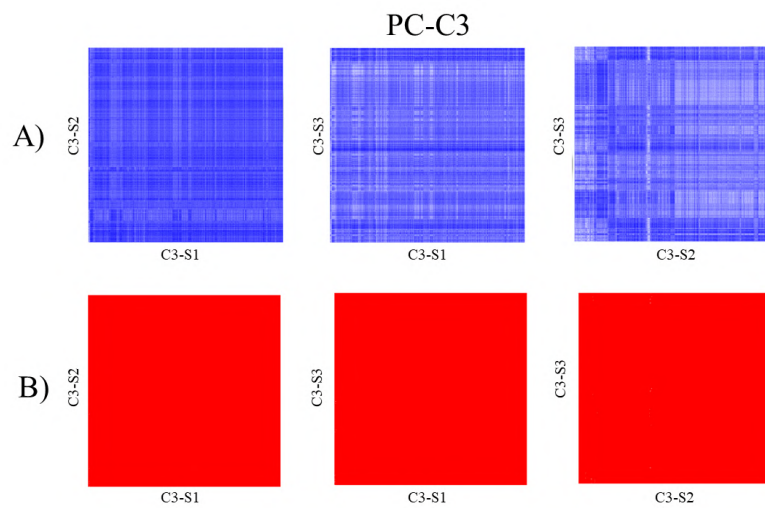
Para establecer si es que hay o no una convergencia estructural entre las diferentes conformaciones de los PCs es necesario comparar las trayectorias de cada conformación entre sí. Esta comparación se puede hacer creando una matriz de valores de RMSD, en la cual de una manera visual se pueden establecer los tiempos en los que hay o no una convergencia. El algoritmo que crea esta matriz compara todas las conformaciones de una trayectoria  $v(t)$  contra todas las conformaciones de una trayectoria  $w(t)$ , calculando el valor de RMSD para cada una de esas comparaciones. Posteriormente se asigna un código de colores y se crea un gráfico tipo *heat map*, donde se puede observar los tiempos de simulación en los cuales las conformaciones se parecen más entre ellas. En este análisis se compararon las conformaciones S1 vs. S2, S1 vs. S3, y S2 vs. S3. Los resultados se aprecian en las Figuras 19 - 26. En dichas imágenes hay dos tipos de matrices, las de tonos azules (A) y las de color rojo (B). En las primeras, los tonos azul claro están asociados a valores mínimos de RMSD en el rango encontrado con una variación continua. Por tanto, zonas claras pueden ser zonas de semejanza estructural. Para poder hacer la distinción estructural, las matrices rojas parten de datos discretizados con un valor mínimo de semejanza de 0.20nm, igual al *Cut-Off* de los clusters de la sección previa. Todas aquellas conformaciones que al ser comparadas entre sí tuvieron un valor de RMSD menor a 0.20nm tendrían asignado un color blanco, mientras que todas aquellas comparaciones con un valor de RMSDs mayor a 0.20nm se colorean de rojo. De esta manera es más sencillo y preciso establecer si se llega a una convergencia o no.



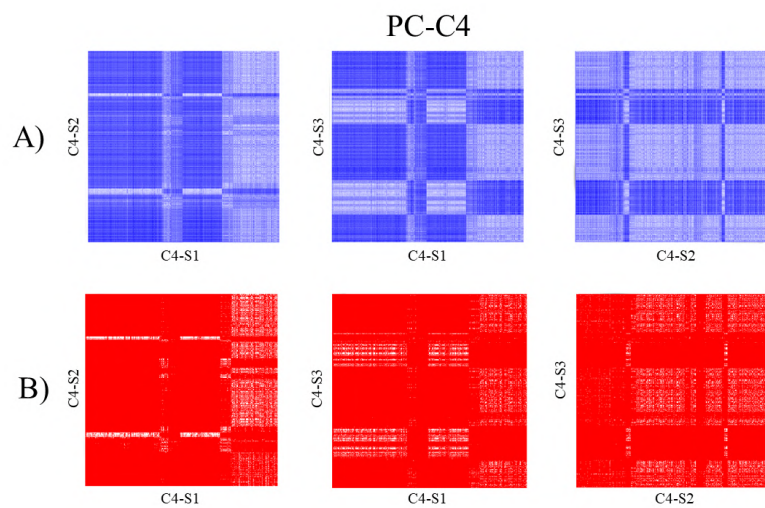
**Figura 19:** Matriz de RMSD de las diferentes conformaciones del PC-C1



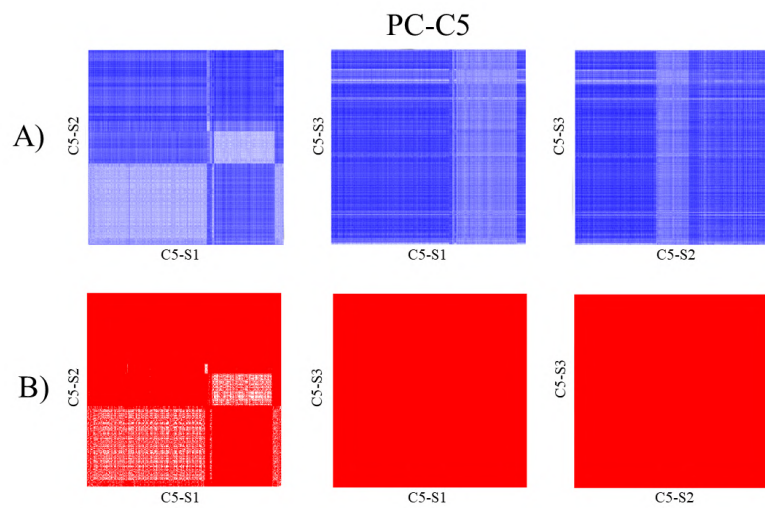
**Figura 20:** Matriz de RMSD de las diferentes conformaciones del PC-C2



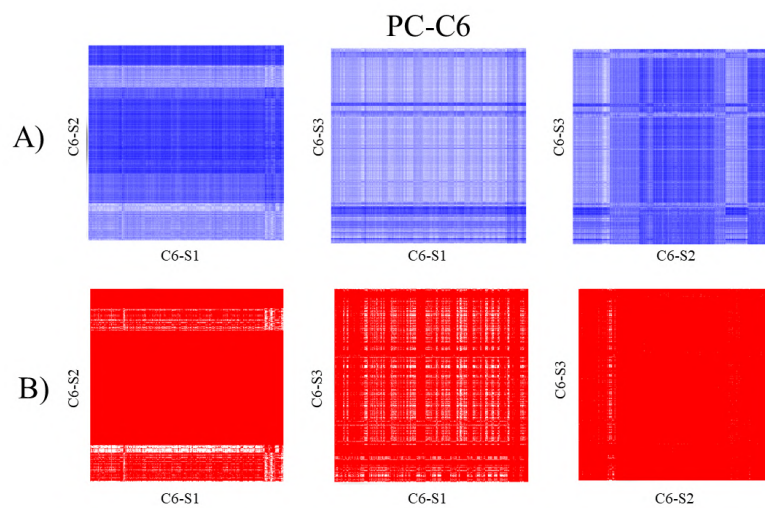
**Figura 21:** Matriz de RMSD de las diferentes conformaciones del PC-C3



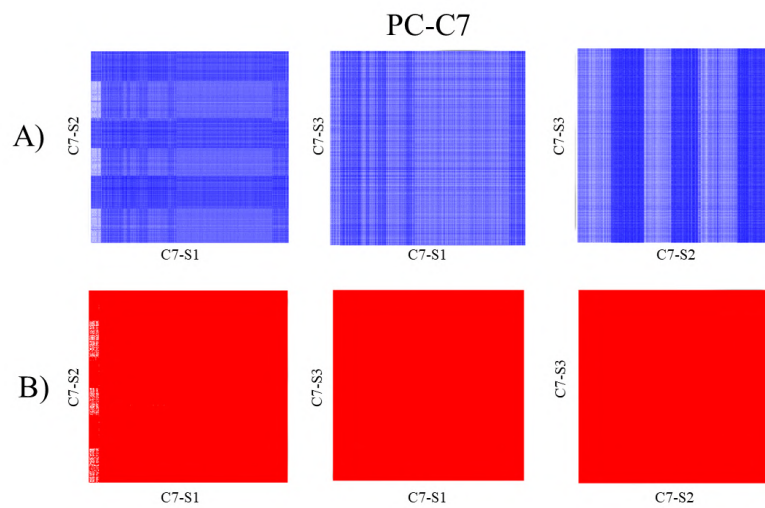
**Figura 22:** Matriz de RMSD de las diferentes conformaciones del PC-C4



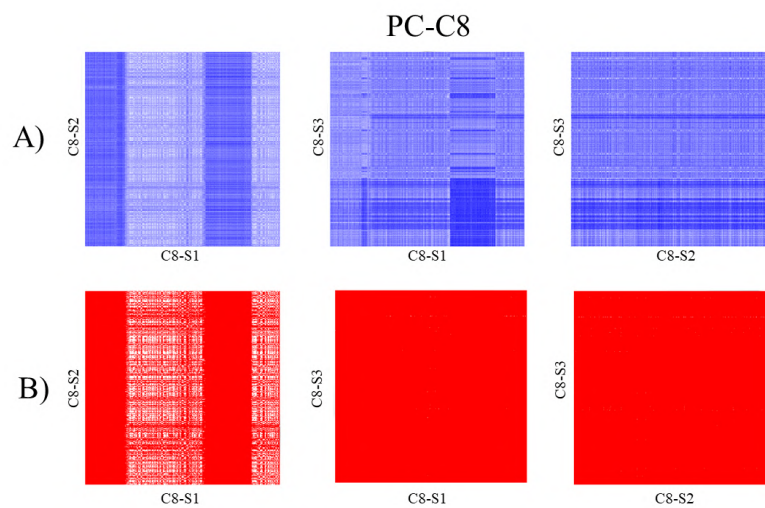
**Figura 23:** Matriz de RMSD de las diferentes conformaciones del PC-C5



**Figura 24:** Matriz de RMSD de las diferentes conformaciones del PC-C6



**Figura 25:** Matriz de RMSD de las diferentes conformaciones del PC-C7



**Figura 26:** Matriz de RMSD de las diferentes conformaciones del PC-C8

El análisis de las matrices, en especial de las rojas, muestra que no se llega a una convergencia estructural en general. El único caso donde se aprecia una gran zona de color blanco, una zona de convergencia, es cuando se comparan las conformaciones S1 y S2 del PC-C2. Existen otros casos, donde se pueden apreciar regiones blancas, pero poco significativas para asegurar que se llegó a una convergencia. Aún así, estos resultados muestran que en la mayoría de los casos ambas trayectorias encontraron los mismos mínimos locales.

Dado que S1 y S2 partieron de una conformación  $\alpha$  hélice, se esperaba que ambas llegaran a una convergencia, pero nuestras simulaciones muestran algo más interesante. Al respecto, se puede decir que las condiciones ambientales rápidamente llevan a las conformaciones iniciales a diferenciarse lo suficiente, explorando diferentes mínimos de energía locales. Esto evita que lleguen a un estado de convergencia, esto es, no existe un mínimo global único.

Para el caso en el que se comparan S1 vs. S3 y S2 vs. S3, la ausencia total o casi total de zonas blancas, deja en evidencia que no hay posibilidad de que haya una convergencia estructural, contrario a lo que se planteó al inicio de este proyecto. En otras palabras, no se observaron cambios permanentes de  $\alpha$  a  $\beta$  o viceversa.

Durante todo el proceso de simulación se generaron muchos datos importantes relacionados con las conformaciones de todos los PCs. Y dado que la conformación de los ligandos influye en el mecanismo de interacción molecular con la proteína de cápside, se decidió ampliar el proyecto haciendo un estudio de *Acoplamiento Molecular*. Las especificaciones y resultados de dicho estudio se discuten en la siguiente sección.

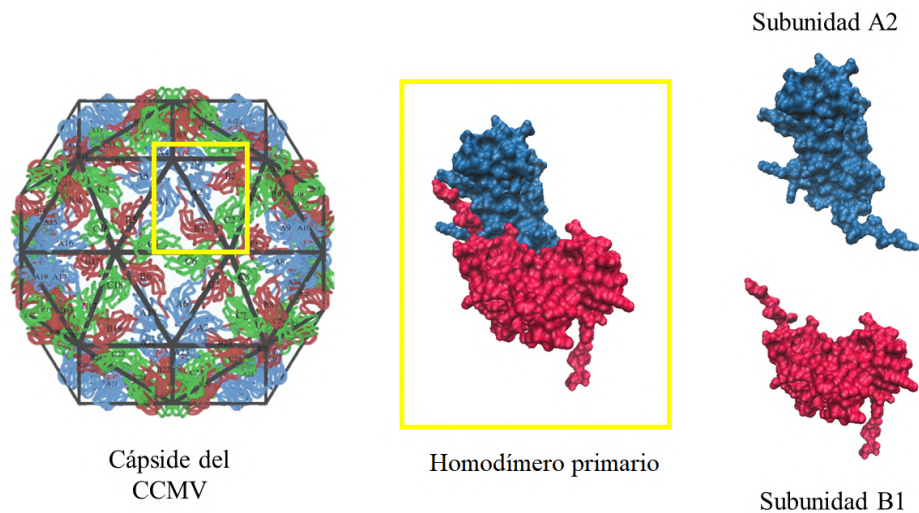


## 7.4. Mecanismo molecular de inhibición

### 7.4.1. Acoplamiento molecular: cavidades

Como se mencionó en la sección 6.3 de la metodología, los ensayos de acoplamiento molecular se realizaron en la plataforma en línea CB-Dock. Como ligandos se utilizaron las estructuras centroides de los Clusters 1, esto es, los más grandes en cada caso, con *Cut-Off* de 0.20 nm. Un centroide,  $w$ , se define como aquella conformación tal que todas las otras conformaciones contenidas dentro del mismo cluster,  $v_i$ , se encuentran a una distancia  $\text{RMSD} < 0.20$  nm con respecto a él. Se acoplaron los centroides de las tres conformaciones (S1, S2 y S3) de los 8 PCs con dos subunidades (A2 y B1) de la cápside viral del CCMV. Dichas subunidades son un ejemplo representativo de los homodímeros que participan en el proceso de autoensamblaje, conocido como dímero primario, el cual es el precursor de la formación de complejos más grandes que eventualmente forman la cápside completa [16]. En la Figura 27 se pueden apreciar las dos subunidades y su localización dentro de la cápside viral. Por convención, se le llama A (en azul) a las subunidades que se agrupan en pentámeros al rededor de un eje de simetría tipo *5-fold* y B (en rojo) a las que participan en la formación de hexámeros junto con las denominadas C (en verde) al rededor de un eje de simetría tipo *3-fold*. Química y conformacionalmente las subunidades A, B y C son idénticas.

CB-Dock, la plataforma donde se hicieron los acoplamientos, se basa en un algoritmo cuya primer tarea es identificar las cavidades o sitios potenciales de interacción en la superficie de la proteína receptora. Para ambas subunidades se analizaron los 5 sitios de interacción más grandes, los cuales se señalan en las Figuras



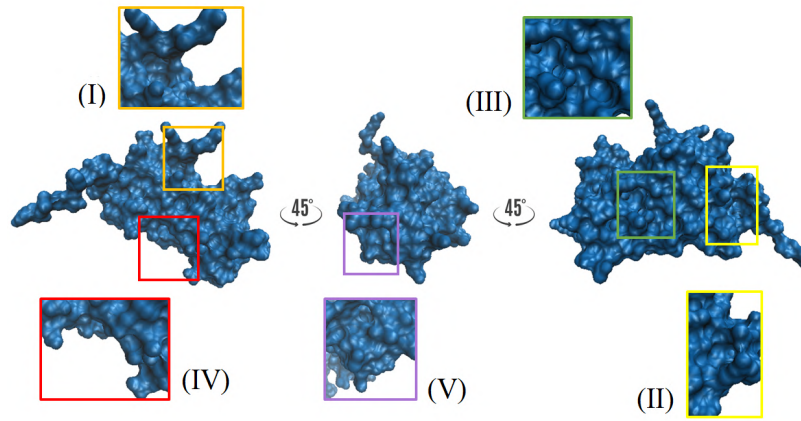
**Figura 27:** Cápside, homodímero primario y subunidades A2 y B1 del CCMV.

28 y 29.

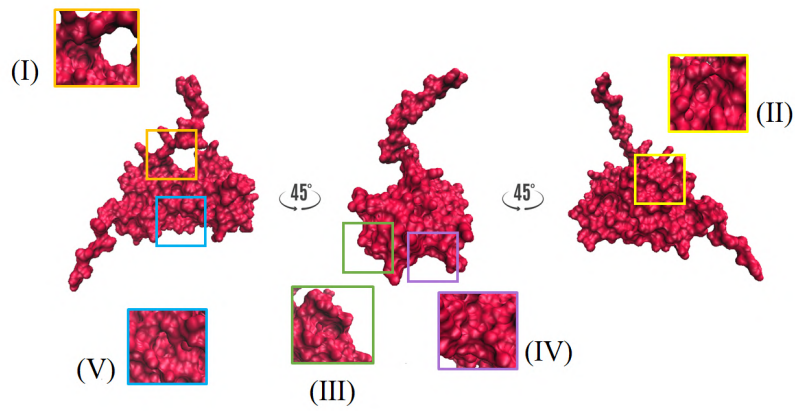
En la Tabla 5 se muestran los tamaños relativos de las cavidades encontradas y señaladas en las Figuras 28 y 29. De todos los sitios que se identificaron, los únicos que están en la interface de interacción A2-B1 son los identificados con el número romano **I**. En la Figura 30 se aprecia el homodímero formado por A2 (azul) y B1 (rojo), y en amarillo y verde PCs unidos a las cavidades **I** respectivas de cada subunidad. Los PCs que se acoplen a este sitio, esencial para la interacción de las subunidades, estarían evitando la formación del dímero.

#### 7.4.2. Acoplamiento molecular: *residuos clave*

En el ámbito general de las IPPs, se ha propuesto la existencia de residuos localizados en las interfases que contribuyen significativamente a la formación y



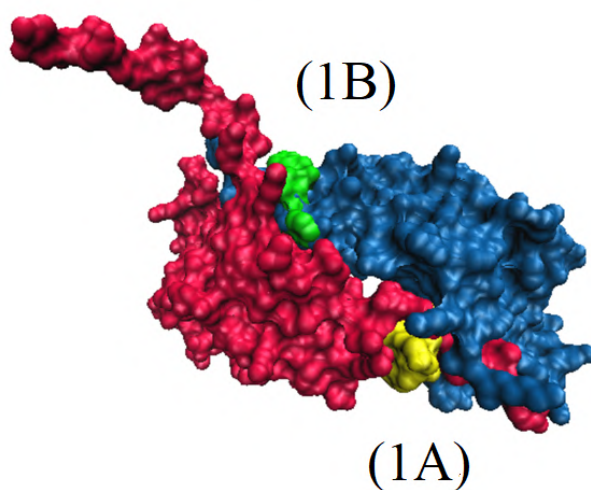
**Figura 28:** Las cinco cavidades más grandes de la subunidad A2 del CCMV.



**Figura 29:** Las cinco cavidades más grandes de la subunidad B1 del CCMV.

Subunidad A2		Subunidad B1	
Cavidad	Tamaño	Cavidad	Tamaño
I	183	I	211
II	118	II	161
III	112	III	96
IV	81	IV	81
V	61	V	72

**Tabla 5:** Tamaño de las cinco cavidades más grandes, en Å<sup>2</sup>, detectadas por CB-Dock en las subunidades A2 y B1 del CCMV.



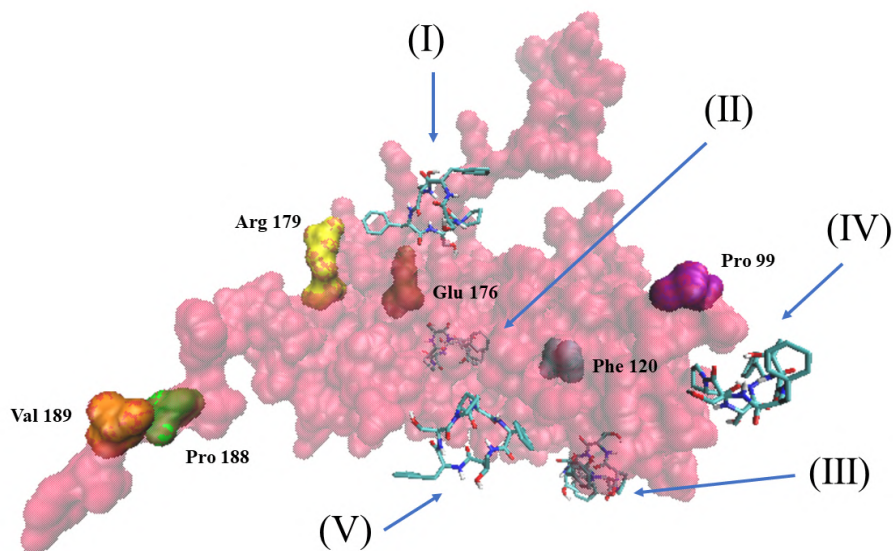
**Figura 30:** Cavidades localizadas en la interfase de interacción. **(1B)** cavidad I de la subunidad B1 (rojo) y **(1A)** cavidad I de la subunidad A2 (azul).

función del complejo proteínico, a los cuales se les ha dado el nombre de *residuos clave*, o *hot-spots* [56]. Nuestro grupo de investigación propuso una metodología sistemática para identificarlos, basada en la conservación estructural desde el ni-

vel primario hasta el cuaternario [57, 15]. Bajo esa definición, los *residuos clave* conforman aquel conjunto de residuos de interfase conservados en secuencia y espacio. Basados en este criterio de conservación estructural, previamente se aplicó tal metodología y se identificaron un total de seis *residuos clave* en la familia de virus *Bromoviridae*, la cual incluye al CCMV. Dichos residuos son: P99, F120, E176, R179, P188, V189.

Los cálculos de acoplamiento molecular nos permiten conocer los sitios de interacción entre los PCs y las subunidades A2 y B1. La pregunta natural que surge es si algunos de los *residuos clave* del CCMV está dentro de alguna de dichas cavidades y si alguno de los PCs interactúa con ellos. En la Figura 31 se observan las cavidades y los respectivos *residuos claves* de la subunidad B1. Las cavidades se señalan con números romanos (I-V) y en cada una de ellas se observa un PC acoplado para su mejor visualización. Las cavidades II y III están por detrás de la subunidad. Como se puede observar en la imagen, la única cavidad que cuenta con *residuos clave* es la cavidad I. En ella se localizan los residuos R179 (carga eléctrica negativa) y E176 (carga eléctrica positiva). Los mismos resultados se observan para la subunidad A2.

La importancia de la relación entre las cavidades y los *residuos clave* recae en la propia naturaleza de estos últimos. Dado que son residuos que están conservados en secuencia y en espacio para toda la familia *Bromoviridae*, es posible que aquellos PCs que interactúen con una gran afinidad en la cavidad I, aquella donde están localizados dos *residuos clave*, también sean candidatos a acoplarse con las subunidades de otros miembros de dicha familia *Bromoviridae*.



**Figura 31:** Relación estructural de los *residuos clave* y los sitios de acoplamiento entre la subunidad B1 del CCMV y los PCs identificados como inhibidores de la formación del homodímero primario.

### 7.4.3. Acoplamiento molecular: resultados

Dado que la búsqueda de la mejor conformación estructural del complejo CP-PC es un proceso estocástico, al aumentar el tamaño del muestreo que se realiza durante el acoplamiento molecular incrementa la probabilidad de encontrar mínimos más profundos. El resumen de los resultados de esta fase se presenta en la Tabla 6. En ella se reporta el mejor valor de afinidad, en kcal/mol, encontrado al realizar tres procesos de acoplamiento molecular independientes siguiendo la misma metodología ya descrita. Este valor es el que interesa, sin embargo, entre paréntesis se reporta el valor promedio y la desviación estándar para dar una mejor idea de la forma del pozo de energía. También se indica, entre corchetes, cual de las cinco cavidades estudiadas fue en la que se acopló el ligando con mayor afinidad.

Los PCs que mostraron consistentemente mayor afinidad por las dos subunidades del CCMV estudiadas en este trabajo fueron C2 y C8, ambos con una secuencia de seis aminoácidos y con al menos un residuo aromático. Por otro lado, el PC con la menor afinidad para las dos subunidades fue el C7, el cual es el único que no tiene anillos aromáticos en su secuencia.

A la par de los cálculos de afinidad previos, se hicieron los ensayos de acoplamiento para moléculas control. Tres de dichas moléculas fueron secuencias cíclicas de 6 y 7 residuos generadas aleatoriamente, las cuales se etiquetaron como E1 (SFAQIP), E2 (SNIAFAF) y E3 (SALGSI). La última de ellas no tiene residuos aromáticos.

Se modelaron las conformaciones  $\alpha$  (S1) y  $\beta$  (S3) de dichas secuencias en UCSF-Chimera siguiendo el protocolo de la sección 6.2.1. Para este caso no se hicieron cálculos de dinámica molecular. Los resultados de dichos acoplamientos se muestran en la Tabla 7.

Los valores de afinidad de las estructuras E1 y E2 se sitúan a la par de los valores de algunas estructuras de la serie C de PCs. Esto muy posiblemente se debe a la presencia de residuos aromáticos en su secuencia. Por otro lado, E3 presenta valores de afinidad por debajo de la mayoría de los PC-C, aparentemente por carecer de anillos aromáticos en su estructura. Este resultado sugiere que los anillos aromáticos en los PCs ayudan a un buen acoplamiento con la CP. Aún así, ninguna conformación del conjunto control E es tan buena como las mejores conformaciones del conjunto C identificadas experimentalmente como inhibidores de la IPP del CCMV.

<b>Subunidad A2</b>			
<b>Péptido cíclico</b>	<b>S1</b>	<b>S2</b>	<b>S3</b>
C1 c-SPSFTF	<b>-6.7</b> (-6.7 ± 0.0) [I]	<b>-7.9</b> (-7.4 ± 0.5) [I]	<b>-7.4</b> (-7.1 ± 0.5) [I]
C2 c-SGYTVP	<b>-7.7</b> (-7.7 ± 0.1) [I]	<b>-8.0</b> (-7.9 ± 0.1) [I]	<b>-7.2</b> (-7.1 ± 0.1) [I]
C3 c-SFRIYVF	<b>-7.4</b> (-7.2 ± 0.2) [I]	<b>-6.8</b> (-6.8 ± 0.1) [I]	<b>-7.2</b> (-7.1 ± 0.1) [I]
C4 c-SFWVLAV	<b>-8.0</b> (-7.3 ± 1.2) [II]	<b>-6.9</b> (-6.9 ± 0.1) [I]	<b>-7.6</b> (-7.5 ± 0.1) [I]
C5 c-SFVVMLY	<b>-6.5</b> (-6.2 ± 0.3) [I]	<b>-6.3</b> (-6.3 ± 0.1) [I]	<b>-6.4</b> (-6.2 ± 0.3) [I]
C6 c-SNSFVY	<b>-7.1</b> (-7.1 ± 0.0) [I]	<b>-6.9</b> (-6.9 ± 0.1) [I]	<b>-7.3</b> (-7.1 ± 0.2) [I]
C7 c-SRDLLM	<b>-5.9</b> (-5.9 ± 0.0) [I]	<b>-6.0</b> (-5.9 ± 0.2) [I]	<b>-6.3</b> (-6.2 ± 0.1) [I]
C8 c-STWWVT	<b>-7.6</b> (-7.5 ± 0.2) [I]	<b>-8.1</b> (-7.8 ± 0.3) [II]	<b>-7.7</b> (-7.7 ± 0.1) [I]
<b>Subunidad B1</b>			
<b>Péptido cíclico</b>	<b>S1</b>	<b>S2</b>	<b>S3</b>
C1 c-SPSFTF	<b>-7.0</b> (-6.9 ± 0.2) [I]	<b>-7.0</b> (-6.9 ± 0.1) [II]	<b>-7.1</b> (-7.1 ± 0.0) [I]
C2 c-SGYTVP	<b>-8.3</b> (-7.6 ± 0.6) [I]	<b>-8.6</b> (-8.6 ± 0.0) [I]	<b>-7.4</b> (-7.3 ± 0.1) [I]
C3 c-SFRIYVF	<b>-7.1</b> (-6.8 ± 0.4) [II]	<b>-7.4</b> (-7.2 ± 0.2) [I]	<b>-6.8</b> (-6.8 ± 0.1) [II]
C4 c-SFWVLAV	<b>-7.5</b> (-7.5 ± 0.0) [I]	<b>-7.7</b> (-7.7 ± 0.0) [II]	<b>-7.7</b> (-7.7 ± 0.0) [I]
C5 c-SFVVMLY	<b>-6.4</b> (-6.3 ± 0.1) [II]	<b>-6.3</b> (-6.3 ± 0.1) [II]	<b>-7.0</b> (-6.9 ± 0.1) [I]
C6 c-SNSFVY	<b>-7.2</b> (-7.1 ± 0.1) [II]	<b>-7.0</b> (-7.0 ± 0.0) [I]	<b>-6.8</b> (-6.8 ± 0.0) [I]
C7 c-SRDLLM	<b>-6.1</b> (-5.8 ± 0.4) [II]	<b>-5.7</b> (-5.4 ± 0.3) [I]	<b>-5.8</b> (-5.3 ± 0.4) [I]
C8 c-STWWVT	<b>-7.7</b> (-7.7 ± 0.0) [I]	<b>-8.3</b> (-8.3 ± 0.0) [I]	<b>-8.5</b> (-8.5 ± 0.0) [I]

**Tabla 6:** Valores de afinidad, en kcal/mol, de los PCs acoplados a la subunidad A2 o B1 de la cápside del CCMV. Se muestra el valor mínimo encontrado en tres repeticiones independientes del proceso de *docking* y entre paréntesis se muestra su promedio con desviación estándar. Entre corchetes se muestra el sitio de unión de la CP, de los cinco estudiados, donde se produjo el acoplamiento molecular con mayor afinidad para formar el complejo.

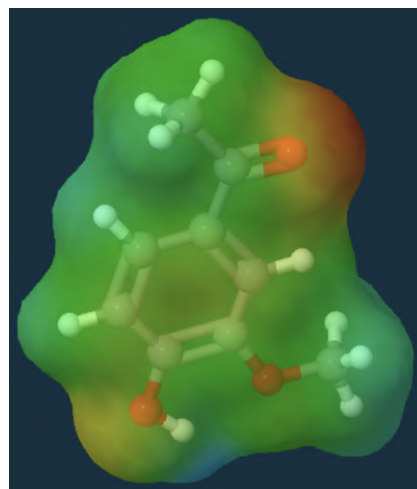
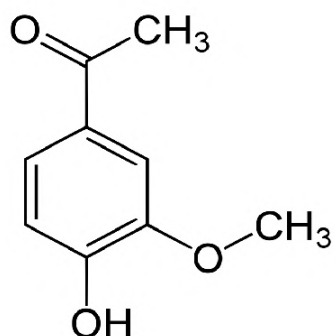


<b>Subunidad A2</b>		
<b>Péptido cíclico</b>	<b>S1</b>	<b>S3</b>
E1 c-SFAQIP	<b>-6.8</b> (-6.4 ± 0.3) [II]	<b>-7.9</b> (-7.4 ± 0.7) [I]
E2 c-SNIAFAF	<b>-6.8</b> (-6.7 ± 0.1) [I]	<b>-7.1</b> (-7.1 ± 0.0) [I]
E3 c-SALGSI	<b>-7.5</b> (-7.5 ± 0.1) [I]	<b>-7.1</b> (-7.1 ± 0.1) [I]
<b>Subunidad B1</b>		
<b>Péptido cíclico</b>	<b>S1</b>	<b>S3</b>
E1 c-SFAQIP	<b>-6.9</b> (-6.9 ± 0.0) [II]	<b>-7.9</b> (-7.4 ± 0.7) [I]
E2 c-SNIAFAF	<b>-7.4</b> (-7.2 ± 0.2) [I]	<b>-7.1</b> (-6.9 ± 0.3) [I]
E3 c-SALGSI	<b>-5.8</b> (-5.8 ± 0.1) [II]	<b>-6.6</b> (-6.6 ± 0.0) [I]

**Tabla 7:** Valores de afinidad entre dos subunidades de la cápside del CCMV y las secuencias peptídicas aleatorias E1, E2 y E3. Se muestra el valor mínimo encontrado en tres repeticiones independientes del proceso de *docking* y entre paréntesis se muestra su promedio con desviación estándar. Entre corchetes se muestra el sitio de unión de la CP, de los cinco estudiados, donde se produjo el acoplamiento molecular con mayor afinidad.

Se utilizó una cuarta molécula como control positivo (*decoy*) para probar el método empleado para realizar los cálculos de acoplamiento molecular. Se eligió un compuesto no peptídico pequeño, la *Apocinina* (Figura 32). En este caso se esperaba que la afinidad hacia la CP fuese baja. Se realizó el acoplamiento usando el mismo procedimiento que con los PCs. Los valores de afinidad entre la *Apocinina* y las subunidades A2 y B1 se muestran en la Tabla 8.

Los resultados de afinidad de *Apocinina* concuerdan con lo esperado. Su estructura molecular limita la afinidad por cualquiera de las subunidades del CCMV, eliminando así, la propiedad inhibitoria de la formación de la cápside.



**Figura 32:** Estructura química y molecular de la Apocinina.

Subunidad	Apocinina
A2	<b>-5.5</b> (-5.5 ± 0.1) [I]
B1	<b>-5.2</b> (-5.2 ± 0.0) [I]

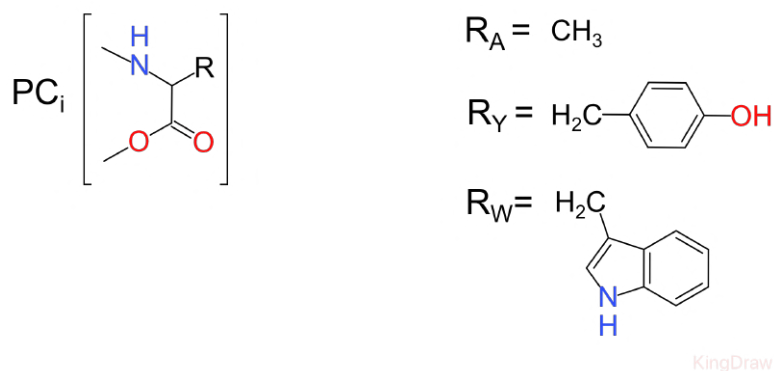
**Tabla 8:** Valores de afinidad, en kcal/mol, entre la Apocinina y dos subunidades de la cápside del CCMV. Se muestra el valor mínimo encontrado en tres repeticiones independientes del proceso de *docking* y entre paréntesis se muestra su promedio con desviación estándar. Entre corchetes se muestra el sitio de unión de la CP, de los cinco estudiados, donde se produjo el acoplamiento molecular con mayor afinidad.

Un resultado interesante que se observó con los experimentos de acoplamiento, fue la relevancia de los residuos aromáticos en la secuencia para generar una buena interacción entre los PC y las proteínas receptoras. Para seguir explorando esta observación se planteó un experimento de escaneo de Alaninas, para estudiar qué sucede cuando se remueven estos residuos aromáticos de las secuencias.

#### **7.4.4. Escaneo de Alaninas**

Esta es una técnica utilizada para determinar la contribución de un residuo específico a la estabilidad o función de una secuencia dada. La alanina se utiliza debido a su grupo funcional no voluminoso y químicamente inerte (Fig. 33). Para este caso se utilizaron como moléculas de estudio a C2 y C8, aquellas secuencias que tuvieron una mayor afinidad por las subunidades A2 o B1 de forma consistente. Se crearon nuevas estructuras desde cero, siguiendo el mismo procedimiento descrito en la sección 6.2.1, donde se sustituyeron uno a uno los residuos de la secuencia por alaninas. Para cada PC se construyeron 6 mutantes simples, a las cuales también se les asignó una conformación  $\alpha$  y  $\beta$ . Las nuevas secuencias creadas, así como los resultados de los experimentos de acoplamiento se muestran en las tablas 9 y 10.

En el caso del PC-C2, su secuencia incluye una Tirosina (Y) como residuo aromático, cuya sustitución por alanina, modificó el valor de afinidad que se obtuvo al acoplarlo con ambas subunidades A2 y B1. Su valor ronda entre las -5.2 y -5.7 kcal/mol, disminuye la afinidad alrededor de 1-2 kcal/mol, con respecto a la secuencia original. Por otro lado, los valores de afinidad de los PCs cuyas sustituciones se hicieron en residuos no aromáticos siempre fueron más negativos, hecho



**Figura 33:** Estructura química de los residuos Tirosina (Y) y Triptófano (W) que se encuentran en la secuencia del péptido cíclico  $PC_i$ , comparada con la de la Alanina (A).

que, en principio, los haría más afines por las proteínas receptoras.

La secuencia del PC-C8 cuenta con dos Triptófanos (W) como residuos aromáticos, situados uno al lado del otro. Para los casos de acoplamiento con A2, el efecto de sustituir solo uno de estos residuos por alanina no provocó cambios significativos en la afinidad. A pesar de esto, se observa que los valores de afinidad de las otras secuencias mutantes son más negativos. En contraste, en los acoplamientos con B1 sí se observa una reducción de la afinidad entre dichas secuencias con sustituciones individuales del W. La reducción de la afinidad, con respecto a la secuencia original esta entre las 0.6 y 1.4 kcal/mol. Por otro lado, cuando se sustituyeron ambos residuos aromáticos por alaninas, el PC-C8 se volvió considerablemente menos afín por cualquiera de las dos subunidades de la cápside del CCMV. La afinidad disminuyó entre 2.0 y 1.6 kcal/mol para el caso de A2 y entre 2.7 y 1.9 kcal/mol para el caso de B1.

Las razones de la dependencia entre la buena afinidad CP-PC se podrían deber a la naturaleza misma de los sustituyentes aromáticos, pues estos poseen una

<b>Subunidad A2</b>		
<b>Variante</b>	<b>S1</b>	<b>S3</b>
SGYTVP	<b>-6.9</b> (-6.9 ± 0.0) [I]	<b>-6.7</b> (-6.0 ± 0.6) [I]
<b>A</b> GYTVP	<b>-7.2</b> (-7.2 ± 0.0) [I]	<b>-6.8</b> (-6.0 ± 0.7) [I]
S <b>A</b> YTVP	<b>-7.1</b> (-7.1 ± 0.0) [I]	<b>-5.6</b> (-5.6 ± 0.0) [I]
<b>SGA</b> TVP	<b>-5.6</b> (-5.5 ± 0.1) [I]	<b>-5.7</b> (-5.7 ± 0.0) [I]
SGY <b>A</b> VP	<b>-6.8</b> (-6.8 ± 0.0) [I]	<b>-7.9</b> (-6.3 ± 1.4) [I]
SGYT <b>A</b> P	<b>-7.2</b> (-6.9 ± 0.4) [I]	<b>-6.7</b> (-6.4 ± 0.6) [I]
SGYTV <b>A</b>	<b>-7.6</b> (-7.6 ± 0.0) [I]	<b>-6.7</b> (-6.7 ± 0.0) [I]
<b>Subunidad B1</b>		
<b>Variante</b>	<b>S1</b>	<b>S3</b>
SGYTVP	<b>-8.0</b> (-6.7 ± 1.1) [I]	<b>-7.6</b> (-7.0 ± 0.5) [I]
<b>A</b> GYTVP	<b>-7.5</b> (-6.7 ± 0.8) [I]	<b>-7.2</b> (-7.0 ± 0.2) [I]
S <b>A</b> YTVP	<b>-7.3</b> (-7.2 ± 0.1) [I]	<b>-6.8</b> (-5.9 ± 1.0) [I]
<b>SGA</b> TVP	<b>-5.6</b> (-5.2 ± 0.5) [I]	<b>-5.8</b> (-5.6 ± 0.2) [I]
SGY <b>A</b> VP	<b>-7.0</b> (-6.3 ± 0.7) [I]	<b>-6.9</b> (-6.4 ± 0.4) [I]
SGYT <b>A</b> P	<b>-7.2</b> (-7.1 ± 0.1) [I]	<b>-6.6</b> (-5.6 ± 0.9) [I]
SGYTV <b>A</b>	<b>-7.4</b> (-6.9 ± 0.8) [I]	<b>-6.9</b> (-6.3 ± 0.5) [I]

**Tabla 9:** Valores de afinidad, en kcal/mol, de las mutantes del PC C2 acopladas a las subunidades A2 y B1 del CCMV. Se muestra el valor mínimo encontrado en tres repeticiones independientes del proceso de *docking* y entre paréntesis se muestra su promedio con desviación estándar. Entre corchetes se muestra el sitio de unión de la CP, de los cinco estudiados, donde se produjo el acoplamiento molecular con mayor afinidad.

<b>Subunidad A2</b>		
<b>Variante</b>	<b>S1</b>	<b>S3</b>
STWWVT	<b>-7.4</b> (-7.4 ± 0.1) [I]	<b>-7.8</b> (-7.8 ± 0.0) [I]
<b>A</b> TWWVT	<b>-8.5</b> (-8.5 ± 0.0) [I]	<b>-8.7</b> (-8.2 ± 0.5) [I]
S <b>A</b> WWVT	<b>-7.9</b> (-7.6 ± 0.3) [I]	<b>-8.2</b> (-8.2 ± 0.0) [I]
<b>ST</b> <b>A</b> WVT	<b>-7.1</b> (-7.1 ± 0.0) [I]	<b>-7.3</b> (-7.3 ± 0.0) [I]
<b>ST</b> <b>A</b> <b>A</b> VT	<b>-5.4</b> (-5.4 ± 0.0) [I]	<b>-6.1</b> (-6.1 ± 0.0) [I]
<b>STW</b> <b>A</b> VT	<b>-6.6</b> (-6.6 ± 0.0) [I]	<b>-8.2</b> (-8.2 ± 0.0) [I]
STWW <b>A</b> T	<b>-8.1</b> (-7.5 ± 0.8) [I]	<b>-8.2</b> (-8.2 ± 0.0) [I]
STWWV <b>A</b>	<b>-8.4</b> (-8.2 ± 0.2) [I]	<b>-8.7</b> (-8.6 ± 0.1) [I]
<b>Subunidad B1</b>		
<b>Variante</b>	<b>S1</b>	<b>S3</b>
STWWVT	<b>-8.5</b> (-8.5 ± 0.0) [I]	<b>-8.0</b> (-8.0 ± 0.0) [I]
<b>A</b> TWWVT	<b>-8.4</b> (-8.4 ± 0.0) [I]	<b>-8.4</b> (-8.4 ± 0.0) [I]
S <b>A</b> WWVT	<b>-9.0</b> (-8.7 ± 0.4) [I]	<b>-8.8</b> (-8.7 ± 0.1) [I]
<b>ST</b> <b>A</b> WVT	<b>-7.1</b> (-7.1 ± 0.0) [I]	<b>-7.3</b> (-7.0 ± 0.3) [I]
<b>ST</b> <b>A</b> <b>A</b> VT	<b>-5.8</b> (-5.6 ± 0.2) [I]	<b>-6.1</b> (-6.0 ± 0.0) [I]
<b>STW</b> <b>A</b> VT	<b>-7.3</b> (-7.2 ± 0.1) [I]	<b>-7.4</b> (-7.4 ± 0.1) [I]
STWW <b>A</b> T	<b>-8.2</b> (-7.5 ± 0.6) [I]	<b>-8.5</b> (-8.4 ± 0.1) [I]
STWWV <b>A</b>	<b>-8.6</b> (-7.9 ± 0.7) [I]	<b>-9.3</b> (-9.1 ± 0.2) [I]

**Tabla 10:** Valores de afinidad, en kcal/mol, de las mutantes del PC C8 acopladas a las subunidades A2 y B1 del CCMV. Se muestra el valor mínimo encontrado en tres repeticiones independientes del proceso de *docking* y entre paréntesis se muestra su promedio con desviación estándar. Entre corchetes se muestra el sitio de unión de la CP, de los cinco estudiados, donde se produjo el acoplamiento molecular con mayor afinidad.

estabilidad química como resultado de la disposición de los electrones  $\pi$  situados encima y debajo del plano del anillo aromático. El sistema  $\pi$  de los anillos aromáticos da lugar a tres tipos de interacciones que involucran restos aromáticos: (i)  $\pi$ - $\pi$ , (ii) catión- $\pi$  y (iii) X-H- $\pi$ . Estas interacciones generan tres tipos diferentes de geometrías, de las cuales, se ha encontrado que la mayoría de las orientaciones atractivas suelen tener forma de T. Se ha estimado que las energías libres de formación de interacciones aromáticas están entre -0,6 y -1,3 kcal/mol [58], lo que concuerda con los cambios en los valores de afinidad cuando se sustituyen los residuos aromáticos en los PCs.

Con los experimentos de escaneo de alaninas se aportó más evidencia que apoya la idea de la dependencia entre la afinidad por la CP del CCMV y la presencia de residuos aromáticos dentro de la secuencia de los PC. Sin embargo, la plataforma experimental que se empleó para identificar péptidos cíclicos con la propiedad de inhibir la formación del dímero primario probó todas las combinaciones posibles para una secuencia de cinco amino ácidos (donde, por cuestiones metodológicas, el sexto siempre fue una serina en el caso de secuencias de seis residuos, o serina y fenilalanina en el caso de secuencias de siete residuos). Con esa plataforma se generaron *in-vitro*  $20^5=3.2 \times 10^6$  de secuencias distintas. únicamente 8 de ellas mostraron la propiedad inhibitoria y son las que se estudiaron en este trabajo. Lo anterior implica que aumentar el número de residuos aromáticos en cada secuencia de PCs no es el camino a seguir. Así, un posible siguiente paso para mejorar la afinidad de los PC hacia la CP del CCMV es modificar los residuos aromáticos actuales por otros no naturales. Existen reportes que indican que el incluir halógenos en la estructura aromática de un ligando aumenta la afinidad por la proteína receptora [59].

## 8. Conclusiones y Perspectivas

En primer lugar, nuestras predicciones teóricas apoyan la evidencia experimental que se ha generado previamente en nuestro grupo de investigación. Hemos mostrado que cuando E176, uno de los seis *residuos clave* del CCMV, se cambia por alanina (mutante E176A) se inhibe la formación de la cápside [16]. En concordancia, los hallazgos de este trabajo muestran que dicho residuo está localizado dentro de la cavidad donde todos los PCs identificados experimentalmente se acoplan con mayor afinidad. Con estos datos proponemos que el posible mecanismo de inhibición molecular funciona de la siguiente manera. Los PCs se unen a la CP en la cavidad donde se localiza el E176, bloqueando así la participación de este *residuo clave* en la IPP necesaria para la formación del homodímero primario, lo que posteriormente frustra la formación de la cápside del CCMV. Un estudio posterior analizará las interacciones específicas de los PCs con el residuo E176.

Por otro lado, no hay que olvidar que la elección de las ocho secuencias de PCs (conjunto C) que se estudiaron en este trabajo no fue al azar, sino de una identificación experimental rigurosa realizada con anterioridad (trabajo de Doctorado de M. Andrade-Medina, 2020). El diseño experimental del trabajo que aquí se presenta incluyó tres secuencias peptídicas adicionales en cuyo caso sí fueron generadas aleatoriamente. Los resultados de los cálculos de acoplamiento molecular muestran que ninguna de esas tres secuencias control da valores de afinidad tan buenos como las mejores conformaciones del conjunto C identificadas *in-vitro*. Adicionalmente, los resultados de los cálculos de escaneo de alanina sugieren que los anillos aromáticos localizados en posiciones específicas en la secuencia de los PCs identificados son importantes para maximizar la afinidad entre estos y la CP.



Se debe recordar que la contribución individual de cada residuo en la secuencia también depende del ambiente local en el que se encuentren, principalmente de los residuos vecinos. Por esta razón es importante el no verlos como algo aislado. La interacción intermolecular total es producto de una combinación de efectos no aditivos.

En lo que concierne a la predicción de la estructura molecular, el haber hecho un muestreo conformacional muy extenso en condiciones termodinámicas compatibles a un ambiente fisiológico y no haber observado una convergencia permanente entre las conformaciones iniciales, sugiere que la superficie energética en la que se encuentran los PCs está caracterizada por tener muchos mínimos locales. Esta es una diferencia importante con respecto a las proteínas globulares, las cuales presentan un mínimo global profundo (hipótesis del embudo). Estas observaciones concuerdan con algunos reportes de la literatura [60, 61], y ponen en manifiesto, una vez más, el gran reto que representa la predicción de la conformación espacial, incluso de biomoléculas pequeñas como los PCs. La metodología que propusimos fue exitosa para encontrar dichas diferencias gracias a la integración del Modelado Molecular y la Dinámica Molecular, sumado a la realización de las simulaciones de forma rigurosa y en infraestructura de cómputo de alto rendimiento, lo cual permitió reducir los tiempos reales de cientos de años a varios meses.

Por último, dentro de las perspectivas de este proyecto están la síntesis química de PCs y la determinación experimental de su estructura molecular con RMN. Además, se plantea la idea de evaluar la inclusión de residuos aromáticos no naturales en la secuencia de los PCs con la finalidad de aumentar la afinidad por la CP. Con respecto al protocolo computacional, en estudios posteriores se pueden aplicar

una serie de métodos de muestreos mejorados para acelerar el proceso de evaluación conformacional de los PCs. Los más populares son la dinámica molecular de intercambio de replicas usando la temperatura como coordenada de reacción o la metadinámica.

## 9. Referencias

- [1] Alessandro Zorzi, Kaycie Deyle, and Christian Heinis. Cyclic peptide therapeutics: past, present and future. *Current Opinion in Chemical Biology*, 38:24–29, June 2017.
- [2] Dario Valenti, Stanimira Hristeva, Dimitrios Tzalis, and Christian Ottmann. Clinical candidates modulating protein-protein interactions: The fragment-based experience. *European Journal of Medicinal Chemistry*, 167:76–95, April 2019.
- [3] Michael G. Wuo, Andrew B. Mahon, and Paramjit S. Arora. An effective strategy for stabilizing minimal coiled coil mimetics. *Journal of the American Chemical Society*, 137(36):11618–11621, September 2015.
- [4] Eilidh K. Leitch, Nagarajan Elumalai, Maria Fridén-Saxin, Göran Dahl, Paul Wan, Paul Clarkson, Eric Valeur, Garry Pairaudeau, Helen Boyd, and Ali Tavassoli. Inhibition of low-density lipoprotein receptor degradation with a cyclic peptide that disrupts the homodimerization of IDOL e3 ubiquitin ligase. *Chemical Science*, 9(27):5957–5966, 2018.
- [5] Lech-Gustav Milroy, Tom N. Grossmann, Sven Hennig, Luc Brunsveld, and Christian Ottmann. Modulators of protein–protein interactions. *Chemical Reviews*, 114(9):4695–4748, April 2014.
- [6] M. P. H. Stumpf, T. Thorne, E. de Silva, R. Stewart, H. J. An, M. Lappe, and C. Wiuf. Estimating the size of the human interactome. *Proceedings of the National Academy of Sciences*, 105(19):6959–6964, May 2008.

- [7] Woong-Hee Shin, Charles W. Christoffer, and Daisuke Kihara. In silico structure-based approaches to discover protein-protein interaction-targeting drugs. *Methods*, 131:22–32, December 2017.
- [8] Rushikesh Sable and Seetharama Jois. Surfing the protein-protein interaction surface using docking methods: Application to the design of PPI inhibitors. *Molecules*, 20(6):11569–11603, June 2015.
- [9] Daniel J. Montiel-García, Ranjan V. Mannige, Vijay S. Reddy, and Mauricio Carrillo-Tripp. Structure based sequence analysis of viral and cellular protein assemblies. *Journal of Structural Biology*, 196(3):299–308, December 2016.
- [10] William M. Hewitt, Siegfried S. F. Leung, Cameron R. Pye, Alexandra R. Ponkey, Maria Bednarek, Matthew P. Jacobson, and R. Scott Lokey. Cell-permeable cyclic peptides from synthetic libraries inspired by natural products. *Journal of the American Chemical Society*, 137(2):715–721, January 2015.
- [11] Pär Matsson, Bradley C. Doak, Björn Over, and Jan Kihlberg. Cell permeability beyond the rule of 5. *Advanced Drug Delivery Reviews*, 101:42–61, June 2016.
- [12] B. V. Venkataram Prasad and Michael F. Schmid. Principles of virus structural organization. In *Viral Molecular Machines*, pages 17–47. Springer US, November 2011.
- [13] Mauricio G. Mateu. Assembly, stability and dynamics of virus capsids. *Archives of Biochemistry and Biophysics*, 531(1-2):65–79, March 2013.

- [14] E. Paintsil and Yung-Chi Cheng. Antiviral agents. In *Encyclopedia of Microbiology*, pages 223–257. Elsevier, 2009.
- [15] Armando Díaz-Valle, José Marcos Falcón-González, and Mauricio Carrillo-Tripp. Hot spots and their contribution to the self-assembly of the viral capsid: In silico prediction and analysis. *International Journal of Molecular Sciences*, 20(23):5966, November 2019.
- [16] Alejandra Gabriela Valdez-Lara, Mariana Andrade-Medina, Josué Alejandro Alemán-Vilis, Aldo Adrián Pérez-Montoya, Nayely Pineda-Aguilar, Eduardo Martínez-Guerra, Abigail Roldán-Salgado, Paul Gaytán, and Mauricio Carrillo-Tripp. Hot-spots and their contribution to the self-assembly of the viral capsid: in-vitro validation. *bioRxiv*, August 2019.
- [17] J.B. Bancroft and Ernest Hiebert. Formation of an infectious nucleoprotein from protein and nucleic acid isolated from a small spherical virus. *Virology*, 32(2):354–356, June 1967.
- [18] Jeffrey A Speir, Sanjeev Munshi, Guoji Wang, Timothy S Baker, and John E Johnson. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by x-ray crystallography and cryo-electron microscopy. *Structure*, 3(1):63–78, January 1995.
- [19] Florence Tama and Charles L. Brooks. The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus. *Journal of Molecular Biology*, 318(3):733–747, May 2002.
- [20] Mauricio Carrillo-Tripp, Craig Shepherd, Ian Borelli, Sangita Venkataraman, Gabriel Lander, Padmaja Natarajan, John Johnson, Charles Brooks, and Vijay

- Reddy. Viperdb2: An enhanced and web api enabled relational database for structural virology. *Nucleic acids research*, 37:D436–D442, 2008.
- [21] Xavier Morelli, Raphaël Bourgeas, and Philippe Roche. Chemical and structural lessons from recent successes in protein–protein interaction inhibition (2p2i). *Current Opinion in Chemical Biology*, 15(4):475–481, August 2011.
- [22] Olivier Sperandio, Christelle H. Reynès, Anne-Claude Camproux, and Bruno O. Villoutreix. Rationalizing the chemical space of protein–protein interaction inhibitors. *Drug Discovery Today*, 15(5-6):220–229, March 2010.
- [23] Albert Loffet. Peptides as drugs: Is there a market? *Journal of Peptide Science*, 8(1):1–7, 2001.
- [24] Norbert Sewald and Hans-Dieter Jakubke. *Peptides: Chemistry and Biology*. Wiley, June 2002.
- [25] Aikaterini A Zompra, Athanassios S Galanis, Oleg Werbitzky, and Fernando Albericio. Manufacturing peptides as active pharmaceutical ingredients. *Future Medicinal Chemistry*, 1(2):361–377, May 2009.
- [26] Laszlo Otvos. Peptide-based drug design: Here and now. In *Peptide-Based Drug Design*, pages 1–8. Humana Press, 2008.
- [27] Alexander A. Vinogradov, Yizhen Yin, and Hiroaki Suga. Macrocyclic peptides as drug candidates: Recent progress and remaining challenges. *Journal of the American Chemical Society*, 141(10):4167–4181, February 2019.
- [28] P Edman. Chemistry of amino acids and peptides. *Annual Review of Biochemistry*, 28(1):69–96, June 1959.

- [29] Kenji Shibata, Toshiyuki Suzawa, Shiro Soga, Tamio Mizukami, Koji Yamada, Nobuo Hanai, and Motoo Yamasaki. Improvement of biological activity and proteolytic stability of peptides by coupling with a cyclic peptide. *Bioorganic & Medicinal Chemistry Letters*, 13(15):2583–2586, August 2003.
- [30] Min S. Lee, Benjamin Gardner, Michael Kahn, and Hiroshi Nakanishi. The three-dimensional solution structure of a constrained peptidomimetic in water and in chloroform observation of solvent induced hydrophobic cluster. *FEBS Letters*, 359(2-3):113–118, February 1995.
- [31] Derick Lau, Linlang Guo, Ruiwu Liu, Jan Marik, and Kit Lam. Peptide ligands targeting integrin  $\alpha3\beta1$  in non-small cell lung cancer. *Lung Cancer*, 52(3):291–297, June 2006.
- [32] Taha Rezai, Bin Yu, Glenn L. Millhauser, Matthew P. Jacobson, and R. Scott Lokey. Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers. *Journal of the American Chemical Society*, 128(8):2510–2511, March 2006.
- [33] Greg Hussack, Tomoko Hiramata, Wen Ding, Roger MacKenzie, and Jamshid Tanha. Engineered single-domain antibodies with high protease resistance and thermal stability. *PLoS ONE*, 6(11):e28218, November 2011.
- [34] Steven W. Millward, Stephen Fiacco, Ryan J. Austin, and Richard W. Roberts. Design of cyclic peptides that bind protein surfaces with antibody-like affinity. *ACS Chemical Biology*, 2(9):625–634, September 2007.
- [35] Catrin Sohrabi, Andrew Foster, and Ali Tavassoli. Methods for generating

- and screening libraries of genetically encoded cyclic peptides in drug discovery. *Nature Reviews Chemistry*, 4(2):90–101, January 2020.
- [36] Ali Tavassoli and Stephen J Benkovic. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in *e. coli*. *Nature Protocols*, 2(5):1126–1133, May 2007.
- [37] Ali Tavassoli, Quan Lu, Jongsik Gam, Hui Pan, Stephen J. Benkovic, and Stanley N. Cohen. Inhibition of HIV budding by a genetically selected cyclic peptide targeting the gag-TSG101 interaction. *ACS Chemical Biology*, 3(12):757–764, November 2008.
- [38] Fan Jiang and Hao Geng. Computational methods for studying conformational behaviors of cyclic peptides. In *Methods in Molecular Biology*, pages 61–71. Springer New York, 2019.
- [39] Anwar Rayan, Hanoeh Senderowitz, and Amiram Goldblum. Exploring the conformational space of cyclic peptides by a stochastic search method. *Journal of Molecular Graphics and Modelling*, 22(5):319–333, May 2004.
- [40] Jérôme Beaufays, Laurence Lins, Annick Thomas, and Robert Brasseur. In silico predictions of 3d structures of linear and cyclic peptides with natural and non-proteinogenic residues. *Journal of Peptide Science*, 18(1):17–24, 2012.
- [41] Yifan Song, Frank DiMaio, Ray Yu-Ruei Wang, David Kim, Chris Miles, TJ Brunette, James Thompson, and David Baker. High-resolution comparative modeling with RosettaCM. *Structure*, 21(10):1735–1742, October 2013.



- [42] Rhiju Das. Four small puzzles that rosetta doesnt solve. *PLoS ONE*, 6(5):e20044, May 2011.
- [43] João M. Damas, Luís C.S. Filipe, Sara R.R. Campos, Diana Lousa, Bruno L. Victor, António M. Baptista, and Cláudio M. Soares. Predicting the thermodynamics and kinetics of helix formation in a cyclic peptide model. *Journal of Chemical Theory and Computation*, 9(11):5148–5157, October 2013.
- [44] Hongtao Yu and Yu-Shan Lin. Toward structure prediction of cyclic peptides. *Physical Chemistry Chemical Physics*, 17(6):4210–4219, 2015.
- [45] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF chimera? a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.
- [46] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, September 1995.
- [47] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, September 2015.
- [48] Michael J. Robertson, Julian Tirado-Rives, and William L. Jorgensen. Improved peptide and protein torsional energetics with the OPLS-AA force field. *Journal of Chemical Theory and Computation*, 11(7):3499–3509, June 2015.

- [49] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffrey D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, July 1983.
- [50] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, September 1997.
- [51] Masaaki Kawata and Umpei Nagashima. Particle mesh ewald method for three-dimensional systems with two-dimensional periodicity. *Chemical Physics Letters*, 340(1-2):165–172, May 2001.
- [52] Yang Liu, Maximilian Grimm, Wen tao Dai, Mu chun Hou, Zhi-Xiong Xiao, and Yang Cao. CB-dock: a web server for cavity detection-guided protein–ligand blind docking. *Acta Pharmacologica Sinica*, 41(1):138–144, July 2019.
- [53] Oleg Trott and Arthur J. Olson. AutoDock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, pages NA–NA, 2009.
- [54] Jinrui Xu and Yang Zhang. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics*, 26(7):889–895, February 2010.
- [55] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.
- [56] Brandon S. Zerbe, David R. Hall, Sandor Vajda, Adrian Whitty, and Dima Kozakov. Relationship between hot spot residues and ligand binding hot

- spots in protein–protein interfaces. *Journal of Chemical Information and Modeling*, 52(8):2236–2244, July 2012.
- [57] Armando Díaz-Valle, Gabriela Chávez-Calvillo, and Mauricio Carrillo-Tripp. *in silico* binding free energy characterization of cowpea chlorotic mottle virus coat protein homodimer variants. In *Advances in Intelligent Systems and Computing*, pages 21–28. Springer International Publishing, 2014.
- [58] Ramnath Anjana, , Marthandan Kirti Vaishnavi, Durairaj Sherlin, Surapaneni Pavan Kumar, Kora Naveen, Pasam Sandeep Kanth, and Kanagaraj Sekar. Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: a study based on orientation and distance. *Bioinformatics*, 8(24):1220–1224, December 2012.
- [59] Leo A. Hardegger, Bernd Kuhn, Beat Spinnler, Lilli Anselm, Robert Ecabert, Martine Stihle, Bernard Gsell, Ralf Thoma, Joachim Diez, Jörg Benz, Jean-Marc Plancher, Guido Hartmann, David W. Banner, Wolfgang Haap, and François Diederich. Systematic investigation of halogen bonding in protein-ligand interactions. *Angewandte Chemie International Edition*, 50(1):314–318, December 2010.
- [60] Anna S. Kamenik, Uta Lessel, Julian E. Fuchs, Thomas Fox, and Klaus R. Liedl. Peptidic macrocycles - conformational sampling and thermodynamic characterization. *Journal of Chemical Information and Modeling*, 58(5):982–992, April 2018.
- [61] Sean M. McHugh, Julia R. Rogers, Hongtao Yu, and Yu-Shan Lin. Insights

into how cyclic peptides switch conformations. *Journal of Chemical Theory and Computation*, 12(5):2480–2488, April 2016.

- [62] Sharandip K. Nijjar. *Targeting C-terminal binding proteins (CtBPs) using genetic selection*. PhD thesis, University of Southampton, September 2011.
- [63] Irene Coin, Michael Beyermann, and Michael Bienert. Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nature Protocols*, 2(12):3247–3256, December 2007.
- [64] Muriel Amblard, Jean-Alain Fehrentz, Jean Martinez, and Gilles Subra. Methods and protocols of modern solid phase peptide synthesis. *Molecular Biotechnology*, 33(3):239–254, 2006.
- [65] Shima A.H. Abdel Monaim, Gerardo A. Acosta, Miriam Royo, Ayman El-Faham, Beatriz G. de la Torre, and Fernando Albericio. Solid-phase synthesis of homodetic cyclic peptides from fmoc-MeDbz-resin. *Tetrahedron Letters*, 59(18):1779–1782, May 2018.
- [66] Christopher J. White and Andrei K. Yudin. Contemporary strategies for peptide macrocyclization. *Nature Chemistry*, 3(7):509–524, June 2011.
- [67] Jose M. Palomo. Solid-phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides. *RSC Adv.*, 4(62):32658–32672, 2014.
- [68] John M. Humphrey and A. Richard Chamberlin. Chemical synthesis of natural product peptides: coupling methods for the incorporation of noncoded amino acids into peptides. *Chemical Reviews*, 97(6):2243–2266, October 1997.

## 10. Apéndice 1

A continuación se describen los *pipelines* utilizados en el desarrollo de este proyecto. Aquellas líneas que comienzan con `##` son comentarios, mientras que aquellas que comienzan con `#` son comandos.

### 10.1. Generación de la topología para los péptidos cíclicos

```
##Nombre del sistema
```

```
PEPNM=cpep
```

```
##GROMACS
```

```
GMX=gmx
```

`cyclic.pl` es un script que genera una topología para los peptidos ciclicos a partir de su configuracion lineal. La herramienta puede ser descargada en la siguiente liga:

```
https://github.com/visvaldask/gmx\$\_-\$makecyclictop
```

```
CYC=cyclic.pl
```

```
##Como primer paso se debe crear un archivo '.top'
```

```
usando pdb2gmx utilizando como entrada el PDB creado con UCSF-Chimera. Al archivo original
```

se le agregan dos residuos extras

*##Se copia el ultimo residuo original al inicio*  
de la secuencia y se reenumera usando al 0

*##2: El primer residuo original se copia al*  
final de la secuencia y se reenumera usando n+1

*##Como primer paso se debe crear un archivo*  
'*.top*' usando pdb2gmx utilizando el PDB creado con  
UCSF-Chimera. Al archivo original se le agregan  
dos residuos extras.

*##Por ejemplo si se quiere crear el peptido FAIR,*  
se debera reareglar la secuencia a RFAIRF

*##con este comando se creara una copia del*  
PDB original , para poder trabajar sobre el

*#cp \${PEPNM}.pdb \${PEPNM}\_augmented.pdb*

*## Al ejecutar este comando se deberan hacer*  
modificaciones manuales explicadas en  
las notas de compatibilidad

*#vi \${PEPNM}\_augmented.pdb*

*##Notas de compatibilidad: cambiar el*  
numero de los siguientes hidrogenos

*##HA2 > HA1*

*##HA3 > HA2*

##HB2 > HB1

##HB3 > HB2

##HD2 > HD1

##HD3 > HD2

##HG2 > HG1

##HG3 > HG2

##Ejecutar el comando *pdb2gmx* sobre el

archivo resultante del paso anterior

##Usar el campo de fuerza *OPLS-AA/L (15)*

con *TIP4P (1)* como modelo de agua

```
##$GMX pdb2gmx -f ${PEPNM}_augmented.pdb -o ${PEPNM}_augmented.gro
```

##Ejecutar el *perl script* en ese

archivo de topologia

```
##${CYC} topol.top
```

##El archivo de topologia del peptido ciclico esta hecho.

NO OLVIDAR remover los residuos aumentados

```
##cp ${PEPNM}_augmented.gro ${PEPNM}.gro
```

```
##vi ${PEPNM}.gro
```

```
##Se definen los archivos de TOPOLOGIA
```

## 10.2. Fase de preparación de la DM

```
TOP=topol\_cyc.top
```

```
POS=posre\_cyc.itp
```

```
## Definicion de la celda computacional
```

```
#$GMX editconf -f ${PEPNM}.gro -o ${PEPNM}_box.gro  
-c -d 1.0 -bt cubic
```

```
##Se crea un archivo PosRe para el PC
```

```
#echo "[ position_restraints ]" > ${POS}  
#echo "; atom type fx fy fz" >> ${POS}  
#awk '{ if( (NF==6) && ($2!="H") ) printf("%6d 1 1000  
1000 1000\n", $3) }' ${PEPNM}_box.gro >> ${POS}
```

```
##Actualizar el archivo de topologia
```

```
#sed 's/posre/posre_cyc/g' ${TOP} > tmp
```

```
#mv tmp ${TOP}
```

```
##Se Agrega solvente
```

```
#$GMX solvate -cp ${PEPNM}_box.gro -cs tip4p
```



```
-o ${PEPNM}_solv.gro -p ${TOP}
```

```
##Se agregan iones
```

```
#$GMX grompp -f ions.mdp -c ${PEPNM}_solv.gro -p ${TOP} -o ions.tpr  
#echo 13 | $GMX genion -s ions.tpr -o ${PEPNM}_solv_ions.gro  
-p ${TOP} -pname NA -nname CL -neutral -conc 0.1
```

```
##Se visualiza para cerciorarse de que todo se vea  
correctamente
```

```
# vmd193 ${PEPNM}_solv_ions.gro
```

```
##Minimizacion de la energia
```

```
#$GMX grompp -f 01_minim.mdp -c ${PEPNM}_solv_ions.gro  
-p ${TOP} -o 01_minim.tpr  
#$GMX mdrun -v -deffnm 01_minim
```

```
##Equilibracion NVT
```

```
#$GMX grompp -f 02_nvt.mdp -c 01_minim.gro -r 01_minim.gro  
-p ${TOP} -o 02_nvt.tpr  
#$GMX mdrun -v -deffnm 02_nvt
```

```
##Equilibracion NPT
```

```
#$GMX grompp -f 03_npt.mdp -c 02_nvt.gro -r 02_nvt.gro  
-t 02_nvt.cpt -p ${TOP} -o 03_npt.tpr
```

```
#$GMX mdrun -v -deffnm 03_npt
```

```
## Creacion del archivo .TPR
```

```
#$GMX grompp -f 04_md.mdp -c 03_npt.gro -t 03_npt.cpt
```

```
-p 04_md.tpr
```

### 10.3. Parámetros de simulación

A continuación se muestra el contenido de los archivos de configuración *.mdp* utilizados en la fase de preparación de la Dinámica Molecular. El punto y coma ';' indica que se trata de un comentario.

#### 10.3.1. ions.mdp

```
; ions.mdp – used as input into grompp to generate ions.tpr  
; Parameters describing what to do, when to stop and what to save  
integrator = steep ; Algorithm  
(steep = steepest descent minimization)  
emtol = 1000.0 ; Stop minimization when the maximum force  
< 1000.0 kJ/mol/nm  
emstep = 0.01 ; Energy step size  
nsteps = 50000 ; Maximum number of (minimization) steps  
to perform  
  
; Parameters describing how to find the neighbors of each  
atom and how to calculate the interactions
```

```
nstlist = 1 ; Frequency to update the neighbor list
and long range forces
cutoff-scheme = Verlet
ns_type = grid ; Method to determine neighbor list
(simple , grid)
coulombtype = PME ; Treatment of long range electrostatic
interactions
rcoulomb = 1.0 ; Short-range electrostatic cut-off
rvdw= 1.0 ; Short-range Van der Waals cut-off
pbc = xyz ; Periodic Boundary Conditions (yes/no)
```

### 10.3.2. 01\_minim.mdp

```
; minim.mdp - used as input into grompp to generate em.tpr
integrator = steep ; Algorithm (steep = steepest descent
minimization)
emtol = 1000.0 ; Stop minimization when the maximum
force < 1000.0 kJ/mol/nm
emstep = 0.01 ; Energy step size
nsteps = 50000 ; Maximum number of (minimization)
steps to perform

; Parameters describing how to find the neighbors of
each atom and how to calculate the interactions
nstlist = 1 ; Frequency to update the neighbor list
and long range forces
```

```

cutoff-scheme = Verlet
ns_type = grid ; Method to determine neighbor list
(simple , grid)
coulombtype = PME ; Treatment of long range
electrostatic interactions
rcoulomb = 1.0 ; Short-range electrostatic cut-off
rvdw = 1.0 ; Short-range Van der Waals cut-off
pbc = xyz ; Periodic Boundary Conditions (yes/no)

```

### 10.3.3. 02\_nvt.mdp

```

title = OPLS Lysozyme NVT equilibration
define = -DPOSRES ; position restrain the protein
; Run parameters
integrator = md ; leap-frog integrator
nsteps = 50000 ; 2 * 50000 = 100 ps
dt = 0.002 ; 2 fs
; Output control
nstxout = 500 ; save coordinates every 1.0 ps
nstvout = 500 ; save velocities every 1.0 ps
nstenergy = 500 ; save energies every 1.0 ps
nstlog = 500 ; update log file every 1.0 ps
; Bond parameters
continuation = no ; first dynamics run
constraint_algorithm = lincs ; holonomic constraints
constraints = all-bonds ; all bonds

```

```

(even heavy atom-H bonds) constrained
lincs_iter      = 1      ; accuracy of LINCS
lincs_order     = 4      ; also related to accuracy
; Neighborsearching
cutoff-scheme   = Verlet
ns_type         = grid   ; search neighboring grid cells
nstlist        = 10     ; 20 fs, largely irrelevant with Verlet
rcoulomb       = 1.0    ; short-range electrostatic cutoff (in nm)
rvdw          = 1.0    ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype     = PME    ; Particle Mesh Ewald for long-range
electrostatics
pme_order      = 4      ; cubic interpolation
fourierspacing = 0.16   ; grid spacing for FFT
; Temperature coupling is on
tcoupl         = V-rescale ; modified Berendsen thermostat
tc-grps       = Protein Non-Protein ; two coupling groups
- more accurate
tau_t         = 0.1     0.1 ; time constant, in ps
ref_t         = 300     300 ; reference temperature, one
for each group, in K
; Pressure coupling is off
pcoupl        = no     ; no pressure coupling in NVT
; Periodic boundary conditions
pbc = xyz     ; 3-D PBC

```

```

; Dispersion correction
DispCorr = EnerPres ; account for cut-off vdW scheme
; Velocity generation
gen_vel = yes ; assign velocities from Maxwell distribution
gen_temp = 300 ; temperature for Maxwell distribution
gen_seed = -1 ; generate a random seed

```

#### 10.3.4. 03\_npt.mdp

```

title = OPLS Lysozyme NPT equilibration
define = -DPOSRES ; position restrain the protein
; Run parameters
integrator = md ; leap-frog integrator
nsteps = 50000 ; 2 * 50000 = 100 ps
dt = 0.002 ; 2 fs
; Output control
nstxout = 500 ; save coordinates every 1.0 ps
nstvout = 500 ; save velocities every 1.0 ps
nstenergy = 500 ; save energies every 1.0 ps
nstlog = 500 ; update log file every 1.0 ps
; Bond parameters
continuation = yes ; Restarting after NVT
constraint_algorithm = lincs ; holonomic constraints
constraints = all-bonds ; all bonds
(even heavy atom-H bonds) constrained
lincs_iter = 1 ; accuracy of LINCS

```

```

lincs_order          = 4          ; also related to accuracy
; Neighborsearching
cutoff-scheme       = Verlet
ns_type              = grid        ; search neighboring
grid cells
nstlist              = 10          ; 20 fs, largely
irrelevant with Verlet scheme
rcoulomb              = 1.0         ; short-range electrostatic
cutoff (in nm)
rvdw                  = 1.0         ; short-range van der Waals
cutoff (in nm)
; Electrostatics
coulombtype          = PME          ; Particle Mesh Ewald for
long-range electrostatics
pme_order             = 4           ; cubic interpolation
fourierspacing       = 0.16        ; grid spacing for FFT
; Temperature coupling is on
tcoupl                = V-rescale    ; modified Berendsen thermostat
tc-grps              = Protein Non-Protein ; two coupling groups
- more accurate
tau_t                 = 0.1         0.1          ; time constant, in ps
ref_t                 = 300         300          ; reference temperature,
one for each group, in K
; Pressure coupling is on
pcoupl                = Parrinello-Rahman ; Pressure coupling on

```

```

in NPT
pcoupltype      = isotropic      ; uniform scaling of
box vectors
tau_p           = 2.0            ; time constant, in ps
ref_p           = 1.0            ; reference pressure, in bar
compressibility = 4.5e-5         ; isothermal
compressibility of water, bar^-1
refcoord_scaling = com
; Periodic boundary conditions
pbc             = xyz            ; 3-D PBC
; Dispersion correction
DispCorr        = EnerPres      ; account for cut-off
vdW scheme
; Velocity generation
gen_vel         = no            ; Velocity generation is off

```

### 10.3.5. 04\_md.mdp

```

title           = OPLS Lysozyme MD simulation
; Run parameters
integrator      = md            ; leap-frog integrator
nsteps         = 500000        ; 2 * 500000 = 1000 ps (1 ns)
dt             = 0.002         ; 2 fs
; Output control
nstxout        = 5000          ; save coordinates every 10.0 ps
nstvout        = 5000          ; save velocities every 10.0 ps

```



```

nstenergy      = 5000 ; save energies every 10.0 ps
nstlog         = 5000 ; update log file every 10.0 ps
nstxout-compressed = 5000 ; save compressed coordinates
every 10.0 ps
; nstxout-compressed replaces nstxtcout
compressed-x-grps = System ; replaces xtc-grps
; Bond parameters
continuation      = yes ; Restarting after NPT
constraint_algorithm = lincs ; holonomic constraints
constraints        = all-bonds ; all bonds (
even heavy atom-H bonds) constrained
lincs_iter        = 1 ; accuracy of LINCS
lincs_order        = 4 ; also related to accuracy
; Neighborsearching
cutoff-scheme     = Verlet
ns_type           = grid ; search neighboring grid cells
nstlist           = 10 ; 20 fs, largely irrelevant with Verlet scheme
rcoulomb          = 1.0 ; short-range electrostatic cutoff (in nm)
rvdw              = 1.0 ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype       = PME ; Particle Mesh Ewald for long-range
electrostatics
pme_order         = 4 ; cubic interpolation
fourierspacing    = 0.16 ; grid spacing for FFT
; Temperature coupling is on

```

```

tcoupl          = V-rescale      ; modified Berendsen thermostat
tc-grps        = Protein Non-Protein ; two coupling groups
- more accurate
tau_t          = 0.1      0.1      ; time constant , in ps
ref_t          = 300      300      ; reference temperature ,
one for each group , in K
; Pressure coupling is on
pcoupl         = Parrinello-Rahman ; Pressure c
oupling on in NPT
pcoupltype     = isotropic       ; uniform
scaling of box vectors
tau_p          = 2.0          ; time constant ,
in ps
ref_p          = 1.0          ; reference
pressure , in bar
compressibility = 4.5e-5      ; isothermal
compressibility of water , bar^-1
; Periodic boundary conditions
pbc            = xyz          ; 3-D PBC
; Dispersion correction
DispCorr       = EnerPres     ; account for cut-off
vdW scheme
; Velocity generation
gen_vel        = no          ; Velocity generation
is off

```

## 10.4. Fase de análisis

```
## NOMBRE DEL SISTEMA
```

```
PEPNM=C1S1
```

```
#mv ARCHIVO_DE_SALIDA_DE_DM.pdb ${PEPNM}_prod_TOT.pdb
```

```
#mv ARCHIVO_DE_SALIDA_DE_DM.xtc ${PEPNM}_prod_TOT.xtc
```

```
#mv prod.tpr ${PEPNM}_prod_TOT.tpr
```

```
## EXES
```

```
GMX=gmx
```

```
CYC=cyclic.pl
```

```
## TOPOLOGIA
```

```
TOP=topol_cyc.top
```

```
POS=posre_cyc.itp
```

```
MD=${PEPNM}_prod_TOT
```

```
## CONCATENAR MAS DE UNA TRAYECTORIA EN UNA
```

```
#${GMX} trjcat -f 05_md.xtc ${MD}.part0002.xtc
```

```
-o ${MD}.xtc
```

```
## CENTRAR LA MOLECULA, PASO IMPORTANTE PARA QUE
```

```
TODOS LOS DEMAS ANALISIS SEAN CONFIABLES
```

```
#echo 1 1 | ${GMX} trjconv -f ${MD}.xtc -s ${MD}.tpr
```

```
-o ${MD}_center.xtc -pbc mol -center
```

```

#echo 1 1 | ${GMX} trjconv -f ${MD}_center.xtc
-s ${MD}.tpr -o ${MD}_fit.xtc -fit rot+trans

## EXTRACCION DE TRAYECTORIAS EN DIFERENTES TIEMPOS.
PARA ELEGIR EL TIEMPO DE EXTRACCION SE MODIFICA
EL VALOR DE -dump EN PICOSEGUNDOS
#echo 1 1 | ${GMX} trjconv -f ${MD}_fit.xtc -s ${MD}.tpr
-o ${MD}_fit_0200ns.pdb -dump 200000

## COMANDO PARA EXTRAER EL RMSD DE LA TRAYECTORIA
#echo 1 1 | ${GMX} rms -s ${MD}.tpr -f ${MD}_fit.xtc
-o rmsd_fit.svg -xvg none

##MATRICES DE RMSD. EL ARCHIVO DE ENTRADA
"-f2" ES TAMBIEN UN ARCHIVO DE TRAYECTORIA,
PERO DE OTRA CONFIGURACION, EN ESTE CASO DE C1-S3.
#echo 1 1 | ${GMX} rms -s ${MD}.tpr -f ${MD}_fit.xtc
-f2 ../C1-S3/C1S3_prod_TOT_fit.xtc -o
rmsd_S1_S3_skip10_mtx.svg -m rmsd_S1_S3_skip10_mtx.xpm
-skip 10 -skip2 10
#${GMX} xpm2ps -o rmsd_S1_S3_skip10_mtx.eps -f
rmsd_S1_S3_skip10_mtx.xpm -title none -gradient 0 0 1

## MATRICES DE CONVERGENCIA, DATOS CONTINUOS.
#pair=rmsd_S1_S2_skip10_mtx

```

```

#././././ xpm_extract_mtx.awk $pair
#gmx xpm2ps -legend none -frame -o $pair.eps
-f $pair.xpm -title none -gradient 0 0 1
#convert $pair.eps -background white -flatten
$pair.png

## MATRICES DE CONVERGENCIA, DATOS DISCRETOS
UTILIZANDO UN CUT-OFF ESPECIFICO, EL CUAL SE
PUEDE MODIFICAR
#cutoff=0.20
#././././ xpm_extract_mtx.awk $pair $cutoff
#gmx xpm2ps -legend none -frame -o ${pair}_cutoff-${cutoff}.eps
-f ${pair}_cutoff-${cutoff}.xpm -title none -gradient 1 0 0
#convert ${pair}_cutoff-${cutoff}.eps -background
white -flatten ${pair}_cutoff-${cutoff}.png

## ANALISIS DE LOS ANGULOS RAMAJAN. EN "-b" SE INDICA
EL FRAME DE INICIO DE MUESTREO Y EN "-e" EL FRAME
FINAL DEL MUESTREO
# ${GMX} rama -f ${MD}_fit.xtc -s ${MD}.tpr
-o rama-${MD}_0000_0100.xvg
-b 0 -e 100000 -xvg none

##ANALISIS ANGULOS DIHEDROS CHI
# ${GMX} chi -f ${MD}_fit.xtc -s ${MD}.tpr -phi -psi

```

```
-all -xvg none
```

```
##ANALISIS DE CLUSTERIZACION, MODIFICANDO EL VALOR  
DE COff, SE ALTERA EL NUMERO DE CLUSTERS  
Y DE CONFIGURACIONES POR GRUPO.
```

```
COff=20
```

```
#rm \cluster_fit*  
#echo 1 1 | ${GMX} cluster -s ${MD}.pdb -f ${MD}_fit.xtc  
-o cluster_fit.xpm -minstruct 2 -g cluster_fit.log  
-dist cluster_fit_dist.svg -sz cluster_fit_size.svg  
-clid cluster_fit_id.svg -xvg none -cl cluster_fit_avg.pdb  
-method gromos -b 000000  
-om cluster_raw.xpm -skip 10 -cutoff 0.${COff}  
#display cluster_noh2o_fit.xpm
```

```
## COMANDOS PARA CAMBIAR DE NOMBRE LOS ARCHIVOS ORIGINALES,  
UTILES CUANDO SE ESTAN PROBANDO DIFERENTES COff.
```

```
#mv cluster_fit_dist.svg cluster_fit_dist_0${COff}.svg  
#mv cluster_fit_id.svg cluster_fit_id_0${COff}.svg  
#mv cluster_fit_size.svg cluster_fit_size_0${COff}.svg  
#mv cluster_fit.log cluster_fit_0${COff}.log  
#mv cluster_fit_avg.pdb cluster_fit_avg_0${COff}.pdb  
#mv cluster_fit.xpm cluster_fit_0${COff}.xpm
```

## 11. Apéndice 2

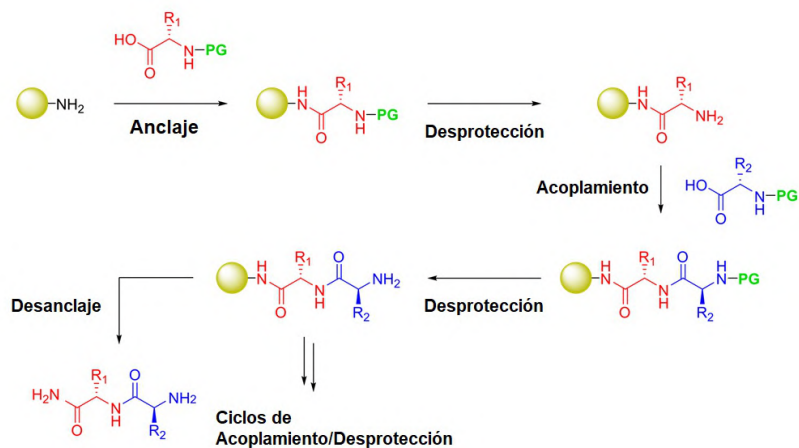
### 11.1. Síntesis, ciclación y purificación de PCs

Todos los péptidos serán sintetizados utilizando la metodología de Síntesis de Péptidos en Fase Sólida, utilizando columnas de polipropileno con un filtro de  $30\mu\text{m}$  de polietileno como contenedores de las reacciones. Este protocolo se obtuvo de [62].

#### 11.1.1. Síntesis de péptidos en fase sólida

La síntesis de péptidos en fase sólida (SPFS) es la técnica por excelencia para la construcción química de estas moléculas (Fig. 34) [63]. En este método, los péptidos se ensamblan en una fase sólida, lo que trae consigo varias ventajas en comparación a la síntesis en solución. Con este sistema, se aumenta el rendimiento y la velocidad de síntesis y se permite la eliminación del exceso de reactivos y los productos secundarios con simples lavados. Además, todo el proceso se puede llevar a cabo en un solo contenedor.

El aminoácido protegido se ancla vía carboxilo a una resina. Después del anclaje del primer aminoácido, los siguientes aminoácidos en la secuencia peptídica deseada se ensamblan de forma lineal desde el extremo carboxilo al extremo amino. Los grupos funcionales de las cadenas laterales deberán de permanecer protegidos durante todo el proceso de elongación. Después del acoplamiento, el exceso de reactivos es removido por filtración y lavados. En general, el proceso de síntesis se resume en una serie de acoplamientos y desprotecciones que se repiten



**Figura 34:** Secuencia de pasos en la técnica *in-vitro* SPFS. La esfera amarilla representa a la resina y *PG* es el grupo protector del aminoácido en turno de acoplamiento.

hasta que se obtiene la secuencia deseada. En el paso final, el péptido es removido de la resina y los grupos protectores de las cadenas laterales se remueve de la misma manera [64]. En los siguientes párrafos se explicará con detalle cada uno de los procedimientos señalados previamente.

### 11.1.2. Carga de la resina

Se disolverá la resina Wang (500 mg, 0.55mmol, 1eq) y el aminoácido protegido por Fmoc (Fmoc-aminoácido) en Dimetilformamida (DMF) (3 mL). Se agitará a temperatura ambiente durante 15 min. Se adicionará Piridina (1.47  $\mu$ L, 1.82 mmol, 3.3 eq) y cloruro de 2,6-diclorobenzoilo (158  $\mu$ L, 1.1 mmol, 2 eq). Se agitará durante 15h a temperatura ambiente. Filtrar la reacción y lavar con DMF (15mL), Diclorometano (DCM) (15mL) y éter dietílico (15 mL). Se secará el crudo de



reacción al vacío.

### 11.1.3. Cuantificación de la carga

La resina cargada se disolverá (6.1 mg) en una disolución de piperidina en DMF al 20 % (10mL), reposará durante 15 min. Se disolverá la mezcla en la misma solución de piperidina/DMF hasta llegar a un volumen de 50 mL. La Absorbancia se medirá a 302 nm. Para calcular la carga se usa la ecuación 2,

$$Carga = \frac{A \times V}{\epsilon \times W} \times 1000, \quad (2)$$

donde A es la absorbancia a 302nm, V es el volumen de la disolución, W es la masa de la resina, y  $\epsilon$  es el coeficiente de extinción molar del aducto a 302nm cuyo valor es  $7,800 m^{-1}$ .

### 11.1.4. Capping

El *capping* se realiza para proteger cualquier hidroxilo remanente en la resina y así evitar productos no deseados. Se agregará cloruro de benzoílo (150  $\mu$ L), piperidina (150  $\mu$ L) y DCM (4 mL) y se incubará por 2 horas.

### 11.1.5. Acoplamiento

El aminoácido-Fmoc (3 eq) se disolverá en una solución de DCM/DMF (4:1, 1.5 mL). Se agregará el activador, hexafluorofosfato de benzotriazol-1-il- oxitri-

pirrolidinofosfonio (PyBop) (3 eq, 0.702 mmol, 365 mg) seguido de N,N- Diisopropylethylamine (DIPEA), un barredor de iones hidrógeno. Se agregará la mezcla a la resina y se agitará durante 2h. Se lavará con DMF (15mL), DCM (15mL), éter dietílico (15 mL). La adición del aminoácido se verificará vía prueba de ninhidrina cuantitativa.

#### **11.1.6. Desprotección**

Se agregará una solución 20 % piperidina/DMF (10 mL) al medio de reacción donde esta el péptido protegido. Se agitará durante 30 minutos y después se lavará con DMF (15mL), DCM (15mL), éter dietílico (15 mL). La remoción del grupo Fmoc se confirma mediante la prueba de ninhidrina cuantitativa (cambio al color purpura).

#### **11.1.7. Prueba de ninhidrina**

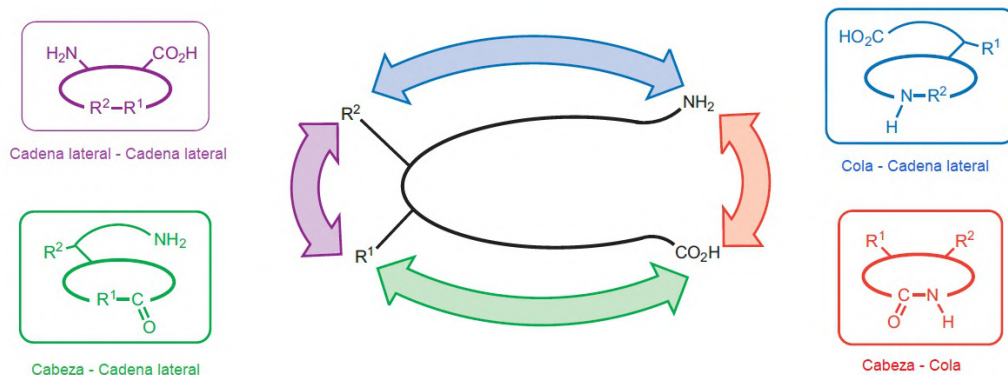
Solución A: 0.5g de ninhidrina en 10 mL de EtOH. Solución B: 0.4 mL de una solución 0.001 M de KCN en 20 mL de piridina. Se agregará una pequeña cantidad de la resina a un vial de reacción y se añadirá 100  $\mu$ L de la solución A y 25  $\mu$ L de la solución B. Se calentará la muestra a 150°C hasta que haya un cambio en el color de la solución.

### **11.1.8. Desanclaje de la resina**

Al finalizar, la resina se colocará en un matraz y se añadirá una mezcla de 94.0 % de Ácido Trifluoroacético (TFA): 2.5 % de agua: 2.5 % de (1,2-ditioetano) EDT: 1 % de Triisopropil silano (TIS) (10-25 mL/g de resina). El EDT sólo se agrega en presencia de Cys y Met. La reacción se agitará durante 3h a temperatura ambiente. La resina se elimina por filtración y se lava con TFA. Los filtrados se combinan y se concentran al vacío, para disolverse nuevamente en la cantidad mínima de TFA. La mezcla se precipita con éter mediante la adición de éter frío (volumen 8-10 veces). La reacción se lleva a cabo en un tubo falcón. Para finalizar, la solución se somete a centrifugación y la solución residual se decanta dejando el sedimento peptídico.

## **11.2. Ciclación de los péptidos lineales**

La ciclación es un paso que puede llevarse a cabo en tres momentos distintos de la SPFS: (i) en la resina antes del desanclado, (ii) en solución después del desanclado y (iii) al mismo tiempo que el desanclado, proceso conocido como Desanclaje Ciclativo [65]. Por otro lado, dependiendo de los grupos funcionales, un péptido puede ser ciclado en cuatro formas distintas: (i) de cabeza a cola (C-terminal a N-terminal), (ii) de cabeza a una cadena lateral, (iii) de una cadena lateral a cola, y (iv) de una cadena lateral a otra cadena lateral (Fig. 35) [66]. El tamaño del anillo es el factor más importante que determina el éxito de una ciclación tipo cabeza-cola, que para los intereses de este proyecto es de la que se hablara de aquí en adelante. La ciclación de péptidos que contienen más de siete residuos no es pro-



**Figura 35:** Tipos de ciclación de péptidos.

blemática y es generalmente sencilla. Sin embargo, péptidos más pequeños pueden resultar problemáticos, o hasta imposibles de ciclar. Es importante mencionar que las ciclaciones de péptidos después del desanclado se realizan mejor bajo condiciones de alta disolución (típicamente submilimolares) para minimizar procesos intermoleculares no deseados, como oligo y polimerizaciones [66, 67].

Una buena ruta de desconexión del péptido cíclico es la clave para una buena ciclación. Es decir, elegir qué aminoácidos de la secuencia serán la cabeza y la cola del precursor lineal, y cuáles serán los que estarán al centro de esta no es irrelevante. Para ello, se han desarrollado diversas guías [68]. Por ejemplo, el rendimiento puede optimizarse si el sitio de ciclación no está estéricamente impedido por aminoácidos N-alkilo,  $\alpha,\alpha$ -sustituidos o  $\beta$ -ramificados (es decir, Val o Ile), o bien, si la macrociclización ocurre entre dos residuos de configuración estereoquímica opuesta.

### **11.2.1. Adición del grupo alditriol**

Para los casos donde en la secuencia de aminoácidos hay Cys, el péptido lineal se disuelve en DMF y se añade alditriol (10 eq) a la solución para evitar la formación de puentes disulfuro entre péptidos. Se agita la reacción durante 24 horas (aparecerá un color verde lima). Concentrar al vacío y precipitar con éter para eliminar el exceso de alditriol.

### **11.2.2. Reacción de ciclación**

El péptido lineal (1eq) se hará reaccionar con 1-etil-3-(3-dimetilaminopropil)carbodiimida (EDC) (3eq) y 1-Hidroxibenzotriazol (HOBt) (6eq) en DMF anhidra. La concentración final del péptido lineal deberá estar en el orden submilimolar para evitar oligomerizaciones. Se agitará durante 24 h, se concentrará al vacío y se precipitará con éter.

### **11.2.3. Reducción de los enlaces disulfuro con TCEP**

Se disolverá el péptido cíclico (1eq) en DMF (10 ml) y se añadirá una solución de tris-(2-carboxietil)-fosfina (TCEP) (1eq) en agua (0.5 mL). Se agitará la mezcla durante 1 hora bajo atmósfera inerte.

### **11.3. Purificación por cromatografía de líquidos de alta resolución**

El PC se purificará por medio de cromatografía de líquidos de alta resolución (HPLC) usando una mezcla de solventes en una columna C18 de fase reversa, utilizando agua y acetonitrilo como solventes. Las proporciones de los solventes dependerán de cada PC.

### **11.4. Determinación estructural mediante RMN**

Las muestras de los PCs se prepararán con agua deuterada a una concentración no menor de 1mM. Los experimentos solicitados a un servicio externo de RMN serán  $^1H$  y  $^{13}C$  para el análisis en una dimensión; COSY, TOCSY y NOESY para el análisis tridimensional. Utilizando los experimentos COSY podremos establecer la relación entre protones separados por no más de tres enlaces, gracias a las constantes de acoplamiento escalares. Con el experimento TOCSY podremos diferenciar protones de diferentes aminoácidos, debido a que los protones de cada residuo pertenecen a diferentes sistemas de espín ya que no hay un acoplamiento escalar a través del enlace peptídico. En un espectro NOESY se podrá identificar una diagonal, que corresponde a un espectro unidimensional de RMN, y donde las señales que queden fuera de la diagonal identifican pares de protones que están separados por menos de 5 Å. Con el uso de tales relaciones de proximidad se podrá construir un modelo de la estructura tridimensional de los PCs en solución.