



**Universitat**  
de les Illes Balears

# Glicación y estrés oxidativo en el plegamiento y desplegamiento de proteínas: aplicación a Trp-Cage.

MELINA SALAZAR OSORIO

**Memoria del Trabajo de Fin de Máster**

Máster Universitario en Ciencia y Tecnología Química  
(Especialidad en Química Biológica)  
de la  
UNIVERSITAT DE LES ILLES BALEARS

Curso Académico 2018-2019

*Septiembre de 2019*

*Dr. Rodrigo Casasnovas Perera*

*Dr. Juan Frau Munar*



## **AGRADECIMIENTOS**

Quiero agradecer a la Fundación Carolina por permitirme vivir esta experiencia, a la Universidad de las Islas Baleares y los profesores de la maestría por acogerme de una muy buena manera, y especialmente a mi grupo de investigación, el cual desde el primer día me hizo sentir muy a gusto; admiro a cada uno de sus integrantes por su calidad humana y por el trabajo que realizan día a día.

Para mí fue todo un reto llevar a cabo este tema de investigación porque nunca había tenido una experiencia directa, pero gracias a mi tutor Rodrigo Casanovas quien me enseñó con paciencia y fundamentos, hoy puedo finalizar con orgullo esta etapa. Gracias Rodrigo por tus valiosas enseñanzas, por estar siempre atento y por ser un excelente maestro.

Gracias a mi familia en Colombia por estar siempre pendiente de este importante proceso académico y personal. A mi madre Liliana Osorio, quien estuvo en todas las fases de esta experiencia; a mi padre Fernando, mi hermana Angélica y nuestro Salvador.

Como siempre en todos mis momentos Nubia Ramírez.

España me llenó de conocimientos, aquí se aumentó mi pasión por la ciencia y mis ojos conocieron hermosos e inolvidables lugares.

Todo empezó con una solicitud para una beca que hoy se materializa en las hojas de este proyecto. Y de aquí en adelante, abierta a todas las oportunidades y bendiciones que traerá el universo.

Gracias a todos los que hicieron realidad este sueño.

## RESUMEN

Las reacciones de glicación entre azúcares y lípidos o proteínas son necesarias para ciertas funciones metabólicas y están reguladas enzimáticamente. Cuando la concentración de azúcares en el organismo es elevada y prolongada en el tiempo, estas reacciones ocurren sin la participación de enzimas. Esto implica la glicación indiscriminada de biomoléculas que en su forma nativa no deben estarlo. La degradación de los productos de glicación modifica irreversiblemente la estructura de las biomoléculas glicadas. Además, incrementa el estrés oxidativo del organismo y promueve la degradación de biomoléculas, incluso aunque no estén glicadas. En el caso de las proteínas, la glicación puede alterar significativamente su normal plegamiento y dinámica, ambos importantes para la función biológica.

En este trabajo se emplearon simulaciones de dinámica molecular de la miniproteína Trp-Cage a su temperatura de fusión. También se estudiaron dos variantes de Trp-Cage modificando el residuo Trp6 por N-formilquinurenina (proteína NFK), y modificando en el residuo Arg16 por carboximetilarginina (proteína CMA). Múltiples simulaciones comenzando desde los estados plegados y desplegados permitieron obtener las superficies de energía libre de los procesos de plegamiento/desplegamiento para caracterizar termodinámicamente el proceso y las conformaciones más relevantes. El análisis de las trayectorias con modelos de estados Markovianos permitió caracterizar la cinética y mecanismo de plegamiento/desplegamiento de Trp-Cage y el efecto de las modificaciones de degradación a la temperatura de estudio.

La transformación de Arg16 en carboximetilarginina no generó un efecto negativo en el plegamiento nativo de la proteína. El estado plegado de CMA muestra valores de RMSD bajos respecto a Trp-Cage, similar contenido en estructura secundaria y superficie accesible al disolvente total y del residuo Trp6 que forma el núcleo hidrofóbico de la proteína. En CMA se observó una estabilización del estado plegado comparado con el de la proteína sin modificar y mayores barreras energéticas para las transiciones entre el estado plegado y los desplegados. Caso contrario al que se dio en la transformación del residuo Trp6 a N-formilquinurenina, en donde el mínimo absoluto en la superficie de energía muestra diferencias estructurales significativas con el plegamiento nativo. Además, la modificación de Trp6NFK disminuye la diferencia de energía libre entre el estado plegado y desplegado de 7 kJ/mol a apenas 1 kJ/mol, así como las barreras de energía entre todos los estados. En NFK, la disminución de contenido en estructura secundaria, la destrucción del núcleo hidrofóbico y el mal plegamiento se atribuye a la interacción de un oxígeno carbonílico del residuo modificado Trp6NFK con su propio grupo amida del backbone.

Las barreras energéticas están relacionadas con la dinámica obtenida con los modelos de estados Markovianos. Para Trp-Cage, la transición entre el estado plegado y todos los estados desplegados se da en un tiempo medio de  $400 \pm 19$  ns. La modificación Arg16CMA incrementa este tiempo en casi el doble ( $789 \pm 60$  ns), mientras que la modificación del residuo crítico del núcleo hidrofóbico Trp6 en NFK disminuye en 1 orden de magnitud el tiempo medio de desplegamiento ( $77 \pm 4$  ns).

Estos resultados muestran que los efectos de degradación de las proteínas pueden tanto favorecer como desfavorecer el plegamiento nativo y la cinética de plegamiento/desplegamiento en diferente medida dependiendo del residuo modificado y de la localización de éste en la estructura proteica.

## ÍNDICE

1.	INTRODUCCIÓN .....	6
1.1.	Estructura de proteínas .....	6
1.2.	Relación entre la estructura y función de las proteínas .....	7
1.3.	Plegamiento y desplegamiento de proteínas .....	8
1.3.1.	Interacciones hidrofóbicas en el plegamiento .....	8
1.3.2.	Mecanismos de plegamiento .....	9
1.4.	Modificaciones postraduccionales de proteínas.....	10
1.4.1.	Glicación y formación de AGEs.....	11
1.4.3.	Degradación de algunos residuos por radicales.....	12
1.4.4.	Degradación del triptófano .....	13
1.5.	Trp-Cage.....	14
2.	OBJETIVOS.....	16
3.	MATERIALES Y MÉTODOS.....	17
3.1.	Dinámica molecular.....	17
3.2.	Detalles computacionales.....	18
3.3.	Parametrización de NFK y CMA .....	20
3.4.	Parámetros de orden.....	21
3.4.1.	RMSD .....	21
3.4.2.	Radio de giro .....	21
3.4.3.	Superficie accesible al disolvente (SASA).....	22
3.4.4.	Medida de la estructura secundaria (DSSP) .....	22
3.5.	Superficie de energía libre .....	22
3.6.	Modelos de estados de Markov para la estimación de cinéticas de plegamiento y desplegamiento.....	23
4.	RESULTADOS Y DISCUSIÓN.....	25
4.1.	Variables Colectivas y muestreo de las trayectorias .....	25
4.2.	Termodinámica de plegamiento.....	28
4.3.	Estado plegado.....	29
4.3.1.	Características estructurales .....	30
4.3.2.	Núcleo Hidrofóbico .....	31
4.3.3.	Superficie accesible al disolvente .....	32
4.4.	Estados desplegados .....	34
4.5.	Estudios Cinéticos.....	37
4.5.1.	Trp-Cage .....	37
4.5.2.	Trp-Cage CMA.....	40
4.5.3.	Trp-Cage NFK.....	43
4.6.	Comparación bibliográfica .....	46
4.7.	Implicaciones Biológicas .....	48
5.	CONCLUSIONES .....	50
6.	BIBLIOGRAFIA.....	51

# 1. INTRODUCCIÓN

## 1.1. Estructura de proteínas

Las proteínas son las macromoléculas más versátiles en los sistemas vivos y cumplen funciones cruciales en el metabolismo. Dicha versatilidad está relacionada con la gran diversidad química de las unidades básicas que las componen: los aminoácidos (Feher, 2017). Hay 20 diferentes aminoácidos estándar utilizados por las células para la construcción de proteínas. Todos ellos presentan un carbono alfa ( $C_\alpha$ ) al que se unen un grupo amino, un grupo carboxilo, un hidrógeno y un sustituyente (R) que se denomina cadena lateral. Al ser los cuatro sustituyentes del  $C_\alpha$  diferentes, los aminoácidos son quirales. Todos los codificados genéticamente son exclusivamente L-aminoácidos. La glicina es una excepción ya que R es un átomo de hidrógeno. Las cadenas laterales pueden ser puramente alifáticas, incluir un anillo aromático (cadenas hidrófobas), estar sustituidas por grupos alcoholes, tioles o amidas (cadenas polares), o pueden estar sustituidas por ácidos carboxílicos, aminos o guanidino. Dado el carácter ácido base de estos últimos grupos, presentan carga eléctrica neta a pH fisiológico.

Las proteínas se forman por la polimerización lineal de aminoácidos. La unión de dos aminoácidos ocurre por formación de un enlace amida denominado peptídico. Formado este enlace, a los aminoácidos se les denomina residuos. Los aminoácidos polimerizan en cadenas de longitud variable. Las secuencias con menos de 50 residuos generalmente se denominan péptidos, mientras que los términos proteína o polipéptido se usan para secuencias más largas. Por convenio, el residuo con el grupo  $\alpha$ -amino libre es el extremo N-terminal y marca el inicio de la secuencia peptídica. El residuo el grupo  $\alpha$ -carboxilo libre se llama extremo C-terminal y marca el final de la secuencia (Berg *et al.*, 2002). Un elemento importante en péptidos y proteínas es el esqueleto o *backbone*, formado por los átomos N,  $C_\alpha$ , C y O de la cadena principal.

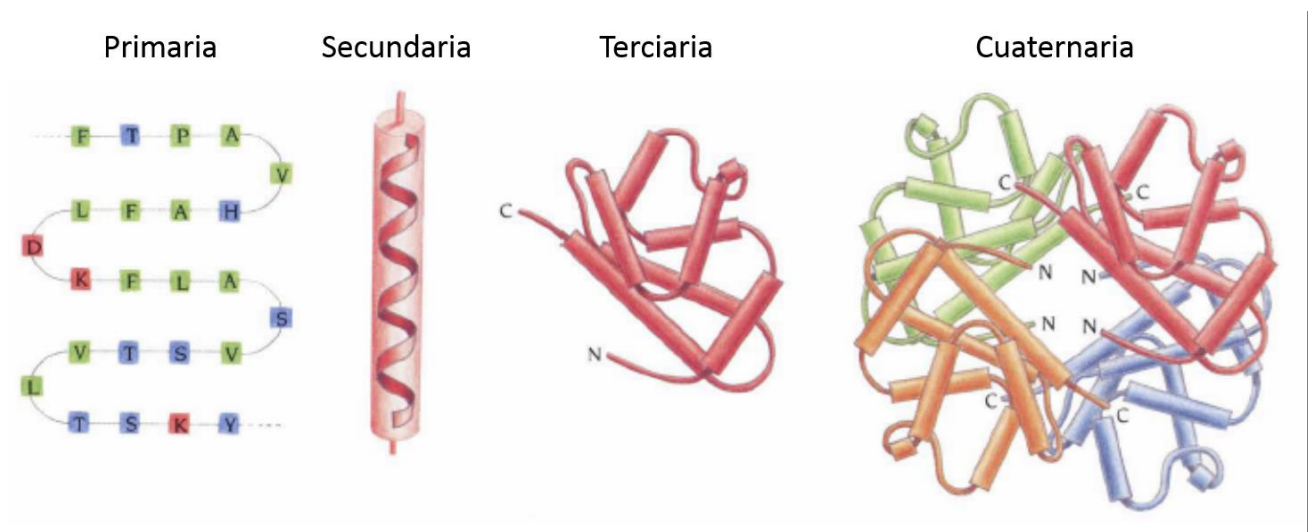
Las proteínas tienen cuatro niveles diferentes de estructura (Figura 1). La estructura primaria es la secuencia de aminoácidos unidos por enlace peptídico. La estructura secundaria es la conformación local que adquiere el backbone de la proteína (Fersht, 1999). Esta conformación se puede definir a partir de la torsión de los diedros  $\phi$  y  $\psi$  formados por los átomos del backbone ( $C_{i-1}$ ,  $N_i$ ,  $C\alpha_i$ ,  $C_i$ ) y ( $N_i$ ,  $C\alpha_i$ ,  $C_i$ ,  $N_{i+1}$ ) respectivamente. La capacidad de formación de enlaces de hidrógeno intramoleculares, la polaridad de las cadenas laterales, los potenciales de torsión de los diedros y factores estéricos favorecen determinadas conformaciones. Las más frecuentes son la hélice  $\alpha$ , lámina  $\beta$ , giros y *loops*. Las características principales de estos elementos son:

- **Hélice  $\alpha$ .** El backbone se pliega en forma de espiral levógira con una frecuencia de 3.6 residuos por giro. Para ello, los diedros  $\phi$  y  $\psi$  de residuos consecutivos tienen valores en torno a  $-67^\circ$  y  $-47^\circ$  respectivamente. Como consecuencia, el grupo carbonilo de un residuo  $i$  acepta un enlace de hidrógeno del grupo amida del residuo  $i+4$  en la secuencia.
- **Lámina  $\beta$ .** Están formadas por la asociación de dos o más secuencias peptídicas relativamente extendidas. Cada una de estas secuencias se denomina fibra  $\beta$  y no tienen por qué ser consecutivas. En las fibras, los diedros  $\phi$  adoptan valores en torno a  $(-119^\circ, -139^\circ)$  y  $\psi$  en torno a  $(113^\circ, 135^\circ)$ . Cada par de fibras se disponen paralelamente y forman enlaces de hidrógeno entre los grupos NH amida de una fibra con los carbonilos de la otra. Si la secuencia de las fibras aumenta en la misma dirección, se dice que éstas son paralelas. Si por el contrario una aumenta y otra decrece, son antiparalelas.

- **Giros.** Consisten en cambios de dirección de la secuencia peptídica en donde el residuo  $i$  que comienza el giro forma un enlace de hidrógeno con el residuo  $i+n$  que lo termina. Existen diferentes posibles giros según el valor de  $n$ , pero los más frecuentes son los giros  $\beta$ , en donde  $n$  es igual a 3.
- **Loops.** Los *loops* (bucles o rizos) son giros con un número grande de residuos entre el principio y el final. Esto los hace ser secciones muy flexibles y dinámicas.

Otras conformaciones menos frecuentes, por ser menos estables, son la hélice  $3_{10}$  y hélice  $\pi$ . La hélice  $3_{10}$ , es de especial interés en este trabajo. Podría considerarse una variación de la hélice  $\alpha$  con un giro más cerrado. En cada vuelta sólo hay 3 residuos y los enlaces de hidrógeno se forman entre el residuo  $i$  e  $i+3$ . Finalmente, a las formas menos regulares que no encajan con las anteriores clasificaciones se les suele designar como *random coil* (espiral aleatoria).

La estructura terciaria es la forma en que los elementos de estructura secundaria se agrupan o empaquetan debido al equilibrio entre las interacciones intramoleculares de las cadenas laterales (Rittle *et al.*, 2019), interacciones proteína-disolvente e interacciones disolvente-disolvente. Las interacciones intramoleculares son habitualmente no covalentes. Según la naturaleza de las cadenas laterales pueden ser interacciones de Coulomb (puentes salinos), enlaces de hidrógeno, interacciones dipolo-dipolo y fuerzas de dispersión. Las interacciones covalentes se dan por formación de enlaces disulfuro entre los grupos tiol de dos cisteínas próximas en el espacio (puente disulfuro). Existe una inmensa diversidad posible de disposiciones de elementos de estructura secundaria. Sin embargo, se observa que éstos tienden a formar patrones o motivos que se repiten entre proteínas diferentes (Branden & Tooze, 1999).



**Figura 1.** Jerarquía de plegamiento de proteínas. Tomado de Branden & Tooze, 1999

## 1.2. Relación entre la estructura y función de las proteínas

Las proteínas realizan multitud de funciones biológicas como estructurales, de transporte, enzimáticas, receptores de membranas, o canales o poros de membrana entre otras. Estas funciones están fuertemente relacionadas con su forma tridimensional. Por ejemplo, para interactuar con el sustrato adecuado (moléculas pequeñas u otras proteínas) en el caso de enzimas o en proteínas de cascadas de señales o para permitir el paso selectivo de ciertas especies a través de membranas celulares.

Los enfoques para el análisis de las relaciones estructura-función en proteínas se basan en similitudes globales o en similitudes locales (Lee *et al.*, 2007). Se ha demostrado que las similitudes de plegamiento se

asocian con la función y también se han utilizado para inferir patrones de secuencia específicos de la función (Pazos, 2004). Sin embargo, muchos pliegues, como el barril TIM y el pliegue de Rossmann, se encuentran en proteínas con diferentes funciones (Kinoshita *et al.*, 2003). Esto sugiere que mutaciones o cambios químicos en algunos residuos pueden causar cambios estructurales que, a su vez, alteren la función normal de las proteínas.

### **1.3. Plegamiento y desplegamiento de proteínas**

El plegamiento de proteínas es un proceso de autoensamblaje en el que una cadena polipeptídica desordenada colapsa para formar una estructura tridimensional compacta y bien definida. Por lo tanto, un plegamiento correcto es esencial para una función adecuada (Dobson, 2003). El plegamiento permite interacciones entre residuos que pueden estar distantes entre sí en la secuencia primaria de la proteína. Así, alteraciones en residuos que pueden estar distantes entre sí en la secuencia primaria pueden provocar cambios en el plegamiento. El plegamiento incorrecto puede conducir a la formación de agregados intra y extracelulares. Algunos agregados son solubles, pero otros insolubles tienen efectos dañinos para la célula, como ocurre en la amiloidosis y algunas enfermedades degenerativas como Alzheimer, Parkinson, Huntington o la esclerosis lateral amiotrófica; la diabetes tipo II y ciertas formas de enfermedad cardíaca (Valastyan & Lindquist, 2014; Hartl, 2017). El plegamiento incorrecto de proteínas también es responsable de muchos cánceres mediados por la proteína p53 supresora de tumores (Chaudhuri & Paul, 2006).

Una característica común de casi todas las enfermedades conformacionales de proteínas es la formación de un agregado causado por la desestabilización de la estructura  $\alpha$ -helicoidal y la formación simultánea de una hoja  $\beta$  (Dobson, 1999). El correcto o incorrecto plegamiento está acoplado a muchos procesos biológicos, incluido el tráfico de moléculas a ubicaciones celulares específicas, la regulación del crecimiento y diferenciación celular (Radford & Dobson, 1999).

#### **1.3.1. Interacciones hidrofóbicas en el plegamiento**

Un tipo de interacción que juega un papel importante en el correcto plegamiento de una proteína es la interacción hidrofóbica y es la principal fuerza impulsora del plegamiento de proteínas globulares solubles. Cuando un polipéptido se pliega en su forma correcta, los residuos con cadenas laterales no polares generalmente se agrupan en el núcleo de la proteína, manteniéndose alejados del agua. Por su parte, los residuos polares o cargados quedan en la superficie de la proteína, interaccionando con el disolvente. En el proceso, casi todas las moléculas de agua se excluyen del interior de una proteína (Lins & Brasseur, 1995).

La explicación aceptada de la formación de las interacciones hidrofóbicas está en la variación de entropía del disolvente. En el estado desplegado, las cadenas apolares forman interacciones dipolo-dipolo inducido con las moléculas de agua. Durante el plegamiento, estas interacciones se reemplazan por interacciones más débiles de tipo dipolo instantáneo-dipolo inducido, implicando un aumento entálpico. Además, el plegamiento reduce los grados de libertad de las cadenas que quedan en el interior de la proteína, por lo que la entropía de éstas es reduce. Por su parte, las moléculas de agua en torno a un soluto apolar se encuentran en cajas de solvatación más ordenadas y formando enlaces de hidrógeno más estables que en el seno del disolvente. El proceso de plegamiento reduce el área del soluto expuesta al disolvente, por lo que implica un coste entálpico para el disolvente ya que los enlaces de la caja de solvatación se reemplazan por otros más débiles. Sin embargo, al pasar al seno del disolvente el número de microestados accesibles de las moléculas



de agua incrementa mucho, causando un aumento de entropía. Normalmente, es este último término el que domina en la variación de energía libre total del proceso.

Las interacciones débiles de van der Waals en el interior de la proteína tienen efecto acumulativo y ayuda a dar a las proteínas su forma única (Olivares & García, 2004). Los enlaces de hidrógeno, dipolo-dipolo y las interacciones electrostáticas entre grupos polares o cargados también contribuyen a la estabilización de la estructura.

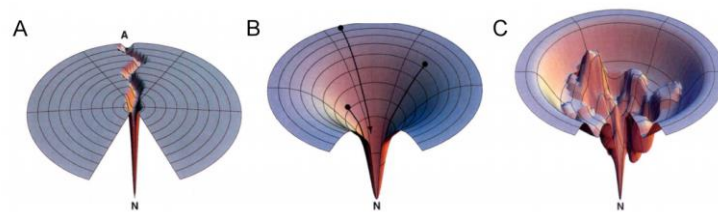
### 1.3.2. Mecanismos de plegamiento

A lo largo de los años han surgido diferentes modelos que pretenden explicar cómo se da el proceso de plegamiento de proteínas. El principio clásico del plegamiento de proteínas es que toda la información requerida para que una proteína adopte la conformación tridimensional correcta es proporcionada por su secuencia de aminoácidos. Esto fue inicialmente establecido por los experimentos de Christian Anfinsen (Anfinsen *et al.*, 1961) que demostraron que la ARNasa desnaturalizada puede replegarse espontáneamente *in vitro* hasta su conformación nativa, por lo que el plegamiento de proteínas parecía ser un proceso de autoensamblaje que no requería factores celulares adicionales. Sin embargo, estudios más recientes han demostrado que ésta no es una descripción adecuada dentro de la célula. El plegamiento dentro de las células está mediado por las actividades de otras proteínas. Un ejemplo de ello son las llamadas chaperonas moleculares, las cuales vigilan, ayudan y controlan el plegamiento (Cooper, 2000).

Unos años posteriores a la teoría de Anfinsen, Cyrus Levinthal (Levinthal, 1968), observó que, aunque las proteínas tienen extensos espacios conformacionales, pueden buscar y converger rápidamente a estados nativos. Lo que indica que el plegamiento no puede tener lugar únicamente por un proceso de prueba y error. Entonces, ¿Cómo las proteínas encuentran sus estados nativos tan rápidamente (típicamente  $\mu\text{s}$  -  $\text{ms}$ )?

Para responder a este tipo de cuestionamientos, otros investigadores han propuesto diversos modelos para ayudar en la comprensión del mecanismo del plegamiento. En el mecanismo de nucleación se sugiere que en las proteínas se forma uno o más núcleos estructurales partir de los cuales se propaga la estructura (Wetlaufer, 1973). En el modelo de difusión-colisión se considera que la proteína está compuesta de varias partes, las cuales se forman primero para luego difundirse, y posteriormente colisionar para formar estructuras de mayor tamaño (Karplus & Weaver, 1976). Posteriormente se propuso el modelo de armazón, en el que se asume que inicialmente se da la formación de algunos de los elementos de estructura secundaria, para que después interaccionen y den lugar a intermediarios que terminan plegándose (Kim & Baldwin, 1982). El modelo del colapso hidrofóbico, sin embargo, supone que la estructura proteica sufre un colapso hidrofóbico dirigido por la entropía del sistema, para después formar las consecuentes estructuras secundarias. Un modelo que combina algunos conceptos expuestos, es el de nucleación-condensación, en donde se forma un núcleo no estable que no presenta estructura secundaria, en el cual colapsa la proteína formándose la estructura secundaria y terciaria casi al mismo tiempo.

El modelo más reciente denominado "entropy funnel", surgió de simulaciones con estudios teóricos. De acuerdo a este modelo, el plegamiento se entiende como un movimiento en un paisaje energético. El concepto histórico del paisaje de energía libre fue propuesto en 1991 por Frauenfelder, Sligar y Wolynes (Frauenfelder *et al.*, 1991). El paisaje describe el conjunto de estados de proteínas como un mapeo de todas las conformaciones posibles por sus energías libres. Este concepto sugiere que el estado nativo de una proteína corresponde a su mínimo de energía libre bajo un conjunto específico de condiciones, dando al paisaje una forma de embudo con la proteína plegada de forma nativa en su parte inferior:



**Figura 2.** Esquema del modelo de plegamiento “entropy funnel”, posibles paisajes energéticos. En el eje vertical se representa la energía libre sin considerar la entropía conformacional. Un punto en la superficie representaría una conformación. Frente a un escenario en el que el paisaje energético es plano excepto para el estado nativo, si existiera un camino en la superficie energética preferente (A). Un paisaje en forma de embudo (B) considera conjuntos estadísticos de conformaciones energéticamente más favorables. En un paisaje más realista (C) hay que considerar que la superficie no es suave por la existencia de intermedios y otros mínimos locales metaestables. Tomado de Dill & Chan, 1997.

La descripción del embudo implica que el plegamiento de proteínas se debe al efecto hidrófobo, con la energía libre disminuida aún más por la formación subsiguiente de contactos energéticos favorables específicos. La profundidad del pozo representa la estabilización energética del estado nativo frente al desnaturalizado, y el ancho del pozo representa la entropía configuracional del sistema. En este modelo en lugar de proponer una única ruta, se producen diferentes que van convergiendo para formar diversas interacciones nativas.

Muchos estudios, experimentales y teóricos, se han dedicado a comprender los mecanismos que subyacen al proceso de plegamiento. Se ha ampliado el conocimiento de la cinética y la termodinámica, pero aún muchos de los detalles se desconocen debido a todas las complejidades intrínsecas que existen en el proceso de plegamiento.

#### 1.4. Modificaciones postraduccionales de proteínas

La secuencia de aminoácidos de una proteína está codificada en el ADN. La síntesis de proteínas se produce mediante en la traducción tras la transcripción. En la transcripción un segmento particular de ADN se copia en ARN; y en la traducción, la secuencia de ARN se utiliza como plantilla para guiar la síntesis de la proteína en cuestión (Nozawa *et al.*, 2017). Sin embargo, a menudo se producen modificaciones postraduccionales. Dichas modificaciones aumentan la diversidad funcional del proteoma mediante la adición covalente de grupos funcionales o proteínas, la escisión proteolítica de subunidades reguladoras o la degradación de proteínas completas. Las modificaciones postraduccionales cambian el tamaño y la estructura de la proteína, y juegan un papel fundamental en la regulación de su plegamiento, su orientación a compartimentos subcelulares específicos, su interacción con ligandos u otras proteínas, y su estado funcional, como la actividad catalítica en el caso de las enzimas o la función de señalización de proteínas involucradas en las vías de transducción de señales (Jensen, 2004).

Las modificaciones se producen en cadenas laterales o enlaces peptídicos, y en la mayoría de los casos están mediados por la actividad enzimática. Estas enzimas incluyen quinasas, fosfatasa, transferasas y ligasas, que agregan o eliminan grupos funcionales, proteínas, lípidos o azúcares hacia o desde las cadenas laterales; y proteasas, que escinden enlaces peptídicos para eliminar secuencias específicas o subunidades reguladoras (Bürkle, 2001). Las modificaciones postraduccionales no enzimáticas se producen estocásticamente, dependen de la estructura ternaria y pueden ser dañinas fisiológicamente (Šoškic *et al.*, 2008). Tales modificaciones incluyen oxidación, glicación, desamidación, racemización e isomerización.

(Nedić *et al.*, 2015). Por lo tanto, la identificación y comprensión de estas modificaciones, es fundamental en el estudio de la biología celular y el tratamiento y la prevención de enfermedades. (Walsh, 2006).

La modificación postraducciona puede ocurrir en cualquier etapa del de la vida útil de la proteína. Por ejemplo, muchas proteínas se modifican poco después de que se completa la traducción para mediar el plegamiento o la estabilidad adecuados de las proteínas o para dirigir la proteína naciente a distintos compartimentos celulares. Otras modificaciones se producen después de completar el plegamiento y la localización para activar o desactivar la actividad catalítica o para influir de otra manera en la actividad biológica de la proteína. (Omenn *et al.*, 2016).

Las modificaciones pueden ser reversibles dependiendo de su naturaleza. Por ejemplo, las quinasas y fosfatasas fosforilan y desfosforilan cadenas laterales específicas, comúnmente como señales de activación o inactivación en cascadas fisiológicas. Por otra parte, la escisión enlaces peptídicos elimina permanentemente las secuencias peptídicas o los dominios reguladores de la proteína original (Nedić *et al.*, 2017).

### **1.4.1. Glicación y formación de AGEs**

La glicación es una de las principales modificaciones post-traduccionales fisiopatológicas. En condiciones de hiperglucemia o estrés oxidativo, las proteínas y el ADN se modifican no enzimáticamente por la glicación oxidativa y dando lugar a productos finales de la glicación avanzada (AGEs). La formación de AGEs implica reacciones de deshidratación, condensación cíclica, enlaces cruzados intermoleculares y oxidación de radicales libres de oxígeno (Baynes *et al.*, 1989). El proceso puede ocurrir tanto extra como intracelularmente. Las proteínas glicadas son más propensas a formar enlaces cruzados con otras proteínas, lo que lleva a alteraciones estructurales y funcionales o incluso a la apoptosis (Seetharaman, 2016).

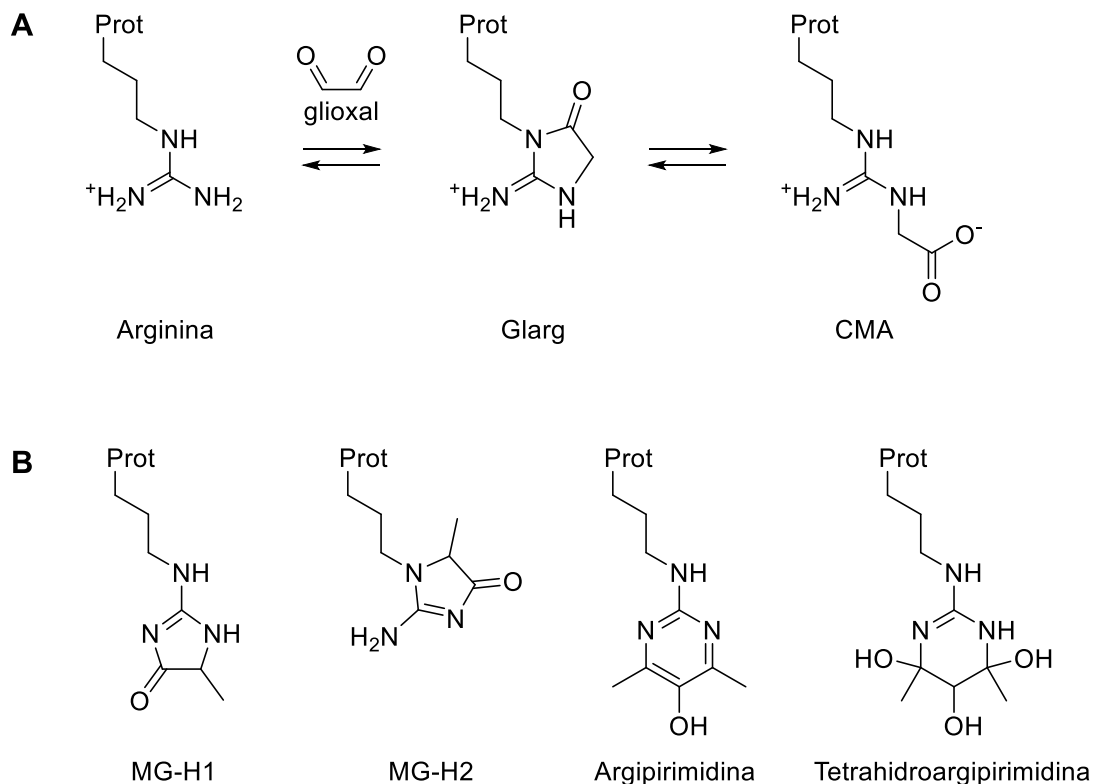
Los trastornos neurodegenerativos, como las enfermedades de Alzheimer y Parkinson o la esclerosis lateral amiotrófica son enfermedades debilitantes y hasta ahora incurables que requieren una investigación intensiva. En estas enfermedades, el plegamiento incorrecto, la agregación y la precipitación de proteínas parecen estar directamente relacionados con la neurotoxicidad. (Chiti & Dobson, 2006). La mayor prevalencia de estas enfermedades en personas diabéticas y la aparición de productos de glicación avanzada en los agregados tóxicos sugiere que la glicación es un factor importante en su desarrollo (Mukherjee, 2015; Salahuddin, 2014).

### **1.4.2. Glicación de Arginina**

Las cadenas laterales de lisina suelen ser más susceptibles a la glicación que las de arginina por su mayor reactividad nucleófila del grupo amina. La proteína de este estudio presenta un residuo lisina y otro arginina. Mientras que en el estado nativo, la arginina establece un puente salino con un residuo aspartato, la lisina sólo interacciona con el disolvente. Por tanto, la modificación del residuo arginina parece a priori potencialmente más dañina para el plegamiento que la modificación de la lisina. Así, se ha optado por estudiar un AGE de arginina.

Experimentalmente se han caracterizado varios AGEs de arginina formados por reacción con los agentes glicantes fisiológicamente más relevantes glioxal y metilglioxal (Figura 3) (Frolov *et al.*, 2014). La incubación con glioxal produce el AGE 1-(4-amino-4-carboxibutil)2-imino-5-oxo-imidazolidina (Glarg), que en condiciones fisiológicas se hidroliza lentamente al AGE carboximetilarginina (CMA) (Figura 3) (Schwarzenbolz *et al.*, 1997; Glomb *et al.*, 2001).

Al igual se ha identificado CMA en colágeno incubado en glucosa (Lijima *et al.*, 2000), y se ha demostrado que los niveles de CMA en proteínas séricas de personas diabéticas son significativamente más altos que los de controles de la misma edad (Odani *et al.*, 2001). Se ha observado que la incubación de colágeno con ribosa también aumenta los niveles de CMA y afecta negativamente al crecimiento de los fibroblastos dérmicos humanos (Shimasaki *et al.*, 2011). También se ha observado que la carboximetilarginina aislada puede actuar como inhibidor de la óxido nítrico sintasa, pudiendo contribuir a la patogénesis de complicaciones vasculares, como insuficiencia renal y aterosclerosis en pacientes con diabetes (Lai *et al.*, 2010).



**Figura 3.** A) Formación de Glarg y CMA por reacción de arginina con glioxal. B) AGEs de arginina con metilglioxal.

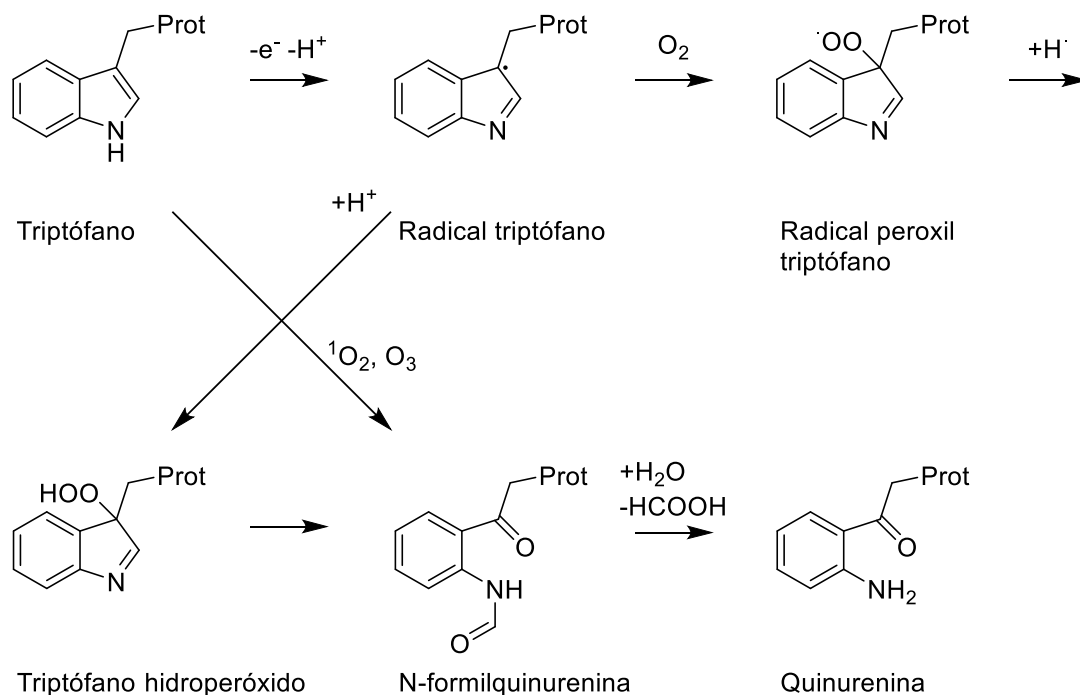
### 1.4.3. Degradación de algunos residuos por radicales

Un radical libre es una especie química que presenta al menos un electrón no apareado, son muy reactivos y pueden causar daños a nivel celular. En proteínas, los radicales libres pueden inducir fragmentación, lo que puede causar alteraciones en sus funciones. Diversas reacciones de oxidación pueden convertir algunos residuos de aminoácidos, como la prolina, la arginina y la lisina derivados de tipo carbonilo (González *et al.*, 2000)

Entre los 20 residuos de aminoácidos proteínogénicos, metionina, cisteína, tirosina, triptófano, histidina y fenilalanina son especialmente propensas a la oxidación (Berlett & Stadtman, 1997). Esta oxidación puede potencialmente causar cambios conformacionales, ligados a modificación en la función biológica (Houde *et al.*, 2010). Además, las alteraciones estructurales introducidas por oxidación pueden conducir a la agregación, o fragmentación, lo que aumenta la susceptibilidad proteolítica de las proteínas oxidadas (Davies, 2016). Por otra parte, los productos de oxidación pueden reaccionar con otras macromoléculas y afectar indiscriminadamente a otros procesos metabólicos (Piedrafita *et al.*, 2015).

### 1.4.4. Degradación del triptófano

Los residuos de triptófano son oxidados por especies reactivas del oxígeno, y pueden dar como resultado una mezcla de diversos productos (Figura 4). El radical triptófano puede reaccionar con oxígeno molecular para formar hidropérido de triptófano, que luego puede reorganizarse en N-formilquinurenina (NFK) y cinurenina (Stadtman & Levine, 2003; Davies, 2003). Las reacciones directas entre Trp y  $^1O_2$  u ozono también pueden conducir a la formación de los productos NFK y quinurenina (Figura 4, Ehrenshaft *et al.*, 2015).



**Figura 4.** Productos de la oxidación del triptófano. Adaptado de Ehrenshaft *et al.*, 2015.

Estos productos de oxidación podrían cambiar la estructura y función proteicas, pero generalmente son degradados por enzimas o el proteasoma (Louie *et al.*, 2002). Sin embargo, también pueden acumularse y agregarse en las células (Squier, 2001). Los efectos funcionales de estas oxidaciones al triptófano dependen del tipo de proteína y la localización de los residuos en la estructura. Por ejemplo, la ozonización de la lisozima de clara de huevo sólo convierte 2 de un total de 6 residuos Trp a NFK (Previero *et al.*, 1967). De éstos sólo uno es necesario para la conformación nativa y la actividad enzimática (Previero *et al.*, 1967).

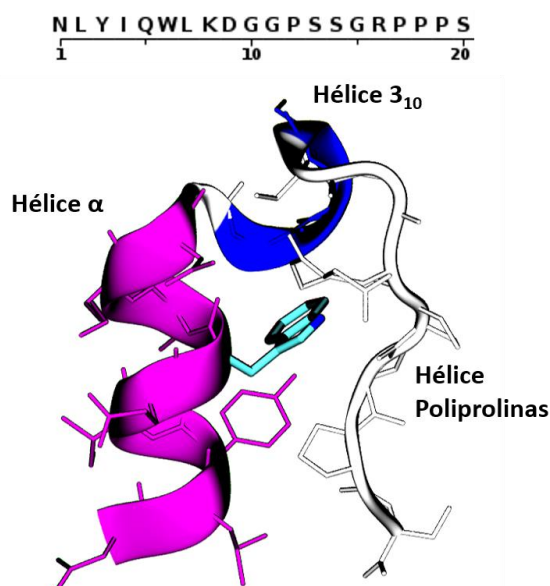
En general la modificación del triptófano se ha detectado en proteínas fisiológicamente importantes y se ha asociado con una serie de afecciones de enfermedades humanas. Se han vinculado niveles elevados de residuos de triptófano oxidados en las proteínas plasmáticas para la diabetes (D'Autréaux & Toledano, 2007) y otras enfermedades relacionadas con la edad. (Todorovski *et al.*, 2011). Adicionalmente la oxidación de los residuos de triptófano se correlaciona con el desarrollo de afecciones como el síndrome de dificultad respiratoria neonatal (Manzanares *et al.*, 2007), aterosclerosis (Batthyány *et al.*, 2000) y la esclerosis lateral amiotrófica (Taylor, *et al.*, 2007).

La modificación oxidativa NFK se ha detectado en una variedad de proteínas como: enzimas respiratorias mitocondriales (Rexroth *et al.* 2012); el fotosistema II (Dreaden *et al.* 2011; Dreaden Kasson *et al.* 2012); la apolipoproteína B-100 (Gießauf *et al.* 1996) y  $\alpha$ -crystallin (Guptasarma *et al.* 1992). A pesar de esto, bajo nuestro conocimiento actualmente no existen estudios que evalúen el efecto de tal producto de la oxidación del triptófano en el plegamiento y desplegamiento de las proteínas. Por tal razón la mini proteína trp-cage,

servirá como un excelente modelo para estudiar dicha modificación y su efecto en la termodinámica y cinética de plegamiento.

## 1.5. Trp-Cage

Trp-Cage es una miniproteína sintética de 20 residuos diseñada en 2001 a partir del péptido de 39 residuos exendina-4, de la saliva del monstruo de Gila (*Heloderma suspectum*) (Neidigh *et al.*, 2001). El rediseño a partir de los residuos 21 al 39 de esta proteína, ha demostrado ser el fragmento de plegamiento independiente más corto conocido, exhibiendo una transición de plegado cooperativa y significativamente más estable que cualquier otra miniproteína. Debido a su rápida cinética de plegamiento, alta estabilidad termodinámica y pequeño tamaño, Trp-Cage ha mostrado ser un modelo ideal, recibiendo una atención considerable en la comunidad computacional a lo largo de los últimos años (Qiu *et al.*, 2002; Ding *et al.*, 2005; Borgohain & Paul, 2018).



**Figura 5.** Características estructurales de Trp-Cage. El color lila representa la hélice  $\alpha$  conformada por los residuos 2 al 8, el color azul la hélice  $3_{10}$  con los residuos 11 al 14, y en color blanco la sección que incluye una hélice de poliprolinas.

La estructura de Trp-Cage revela un núcleo hidrófobo compacto donde tres residuos de prolina Pro12, Pro18, Pro19, una glicina Gly11, y una tirosina Tyr3 empaquetan la cadena lateral aromáticas Trp6 (Zhou, 2003). Los elementos de la estructura secundaria incluyen una hélice  $\alpha$  en los residuos 2 a 8, seguida de una hélice  $3_{10}$  en los residuos 11 a 14 y una hélice de poliprolina II C-terminal.

El mecanismo de plegado de Trp-Cage se ha investigado extensivamente tanto experimentalmente como teóricamente y se ha determinado que debe existir al menos un estado intermedio en la ruta de plegamiento/desplegamiento (Neuweiler *et al.*, 2005; Ahmed *et al.*, 2005; Zhou, 2013; Meuzelaar *et al.*, 2013).

Una combinación de técnicas experimentales y simulaciones de dinámica molecular, apuntaron a un proceso simple de despliegue en dos estados (Heyda *et al.*, 2011). Sin embargo, algunos investigadores han encontrado que la dinámica de relajación cercana a la temperatura de fusión, implica rápidas fluctuaciones en la región de la poliprolina II (Culik *et al.*, 2011; Byrne *et al.*, 2014). Estos autores han propuesto que la formación de una hélice  $\alpha$  bien definida precede a la formación de la estructura terciaria con el plegamiento

correcto de la hélice de poliprolina II para completar el núcleo hidrofóbico (Culik *et al.*, 2011; Byrne *et al.*, 2014), lo que implica al menos un intermedio (Meuzelaar *et al.*, 2013).

Adicionalmente, se han sugerido diferentes vías de plegamiento, posiblemente asociadas a la temperatura de trabajo empleada. En un estudio realizado a diferentes temperaturas se definieron dos vías. Para la vía I, se planteó que el colapso hidrófobo precede a la formación de la hélice  $\alpha$ , y la en la vía II, los eventos ocurren en orden inverso (Andryushchenko & Chekmarev, 2017). Estos autores encontraron que a medida que aumenta la temperatura, la vía I se transforma gradualmente en la vía II. La vía I se ha observado en las simulaciones a temperatura ambiente (Chowdhury *et al.*, 2004; Juraszek & Bolhuis 2006; Han & Schulten 2013; Pandini & Fornili, 2016), y la vía II en las simulaciones a temperaturas de fusión y mayores (Paschek *et al.*, 2008) (Deng *et al.*, 2013).

Otro aspecto en el cual se han observado resultados contradictorios en la literatura, es el relacionado con la influencia del puente salino entre Asp9-Arg16 en el plegamiento nativo. Algunas simulaciones en solvente implícito han enfatizado el papel del puente salino para la formación y estabilidad de la estructura plegada (Hu *et al.*, 2008; Snow *et al.*, 2002; Chowdhury *et al.*, 2003), pero en otros estudios con disolvente explícito se observó que la presencia de este puente salino no era indispensable para el plegamiento (Kannan & Zacharias 2014).

## 2. OBJETIVOS

Se realizarán simulaciones de dinámica molecular y posteriores análisis de las mismas para:

1. Identificar y caracterizar las estructuras más relevantes termodinámicamente (i.e. las más pobladas) de los estados plegado y desplegado de la miniproteína Trp-Cage a su temperatura de fusión.
2. Estimar la cinética de plegamiento y desplegamiento y describir el mecanismo asociado de la miniproteína Trp-Cage a su temperatura de fusión.
3. Describir el efecto de la degradación del triptófano Trp6 a NFK en la termodinámica y cinética de plegamiento/desplegamiento de Trp-Cage a su temperatura de fusión.
4. Describir el efecto de la formación del AGE CMA en Arg16, un residuo importante para la formación del núcleo hidrofóbico de Trp-Cage, en la termodinámica y cinética de plegamiento/desplegamiento a su temperatura de fusión.



### 3. MATERIALES Y MÉTODOS

#### 3.1. Dinámica molecular

La dinámica molecular (DM) es una técnica que permite la simulación de la evolución temporal de un sistema de partículas interaccionantes. El resultado de una simulación se denomina trayectoria y es el conjunto de posiciones y velocidades de las partículas a lo largo de un cierto período de tiempo. Dadas las coordenadas y velocidades iniciales de las partículas, la integración de la segunda ley de Newton del movimiento permite calcular las nuevas posiciones y velocidades a tiempos posteriores:

$$-\frac{dV}{d\mathbf{r}_i} = m_i \frac{d^2\mathbf{r}_i}{dt^2}, \quad i = 1, \dots, N \quad [1]$$

Donde  $V$  es la energía potencial,  $m_i$  y  $\mathbf{r}_i$  son la masa y posición de la partícula  $i$ . La resolución analítica de las ecuaciones de movimiento [1] no es posible para más de dos partículas, por lo que se recurre a métodos de diferencias finitas. Conocidas las posiciones a tiempo  $t$ , según una expansión de Taylor, las posiciones tras un pequeño incremento de tiempo  $\Delta t$  son:

$$\mathbf{r}_{t+\Delta t} = \mathbf{r}_t + \frac{d\mathbf{r}}{dt} \Delta t + \frac{1}{2} \frac{d^2\mathbf{r}}{dt^2} \Delta t^2 + \frac{1}{6} \frac{d^3\mathbf{r}}{dt^3} \Delta t^3 + \dots \quad [2]$$

Las posiciones a un tiempo  $\Delta t$  anterior son:

$$\mathbf{r}_{t-\Delta t} = \mathbf{r}_t - \frac{d\mathbf{r}}{dt} \Delta t + \frac{1}{2} \frac{d^2\mathbf{r}}{dt^2} \Delta t^2 - \frac{1}{6} \frac{d^3\mathbf{r}}{dt^3} \Delta t^3 + \dots \quad [3]$$

Según el método de Verlet (Verlet, 1967), considerando los tres primeros términos (posiciones, velocidades y aceleraciones) y sumando [2] y [3] se pueden calcular las posiciones a tiempo  $t+\Delta t$ :

$$\mathbf{r}_{t+\Delta t} = 2\mathbf{r}_t - \mathbf{r}_{t-\Delta t} + \frac{d^2\mathbf{r}}{dt^2} \Delta t^2 \quad [4]$$

Las velocidades a tiempo  $t$  se obtienen restando [2] y [3]:

$$v_t = (\mathbf{r}_{t+\Delta t} - \mathbf{r}_{t-\Delta t}) / 2\Delta t \quad [5]$$

El análisis posterior de las trayectorias permite obtener propiedades macroscópicas del sistema como densidad, viscosidad o constante dieléctrica. La concordancia de éstas con medidas experimentales es un indicador de la representatividad del sistema real. Quizá más interesante, es que las simulaciones otorgan un acceso inmediato al cálculo de propiedades microscópicas cuya medida experimental debe ser indirecta, o es difícil o imposible. En estos casos, las simulaciones son un complemento ideal a las técnicas experimentales para comprender con detalle el sistema de estudio.

En química es habitual encontrar tres niveles de resolución espacial en las simulaciones de DM. En el primer nivel, las simulaciones *ab initio* describen los electrones y núcleos de los átomos y sus interacciones de naturaleza cuántica. Estas simulaciones son las más costosas computacionalmente y permiten tratar sistemas de decenas a pocos cientos de átomos en un espacio temporal de hasta decenas de picosegundos. En el segundo nivel, las dinámicas *all atom* no tratan electrones y núcleos, pero aún hay resolución atómica. Los átomos se tratan como esferas blandas con un potencial de tipo Lennard-Jones y una carga atómica parcial puntual. Si ciertos átomos forman una molécula, los enlaces químicos y ángulos de enlace se describen con potenciales armónicos y los potenciales de torsión de enlace con combinación de funciones trigonométricas. El conjunto de funciones que modelan las interacciones entre los átomos se denomina campo de fuerza o *force field*. Esta aproximación no permite describir la formación y ruptura de enlaces químicos, pero es computacionalmente mucho menos costosa y da acceso al estudio de sistemas de cientos de miles de átomos en escalas temporales de pocos microsegundos. En el tercer nivel de resolución se agrupan átomos o incluso diferentes moléculas para formar superpartículas. Esta aproximación, denominada *coarse grained*, permite estudiar millones de partículas en escalas temporales de milisegundos o incluso segundos a costa de perder la resolución atómica.

Las proteínas son moléculas formadas por cientos o miles de átomos y su plegamiento/desplegamiento tiene lugar en escalas de microsegundos a minutos. Por tanto, si las proteínas son pequeñas o se pliegan/despliegan rápidamente, su estudio es posible con simulaciones *all atom*. Por otra parte, la simulación de plegamiento de proteínas lentas requiere métodos *coarse grained* (Caflish & Paci 2005). Diferentes propiedades estructurales y dinámicas de la proteína se pueden calcular a partir de las trayectorias resultantes para obtener una comprensión cinética y termodinámica del proceso (Berhanu & Masunov, 2014).

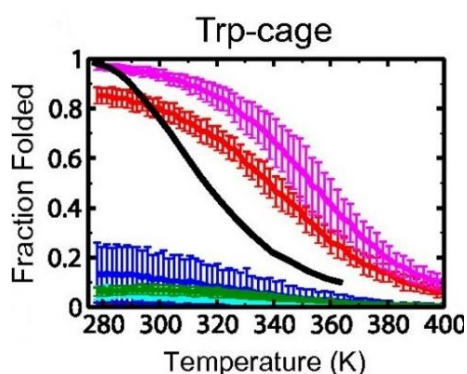
### 3.2. Detalles computacionales

Para este estudio, las simulaciones se realizaron con el programa Gromacs-2016.4 (Berendsen *et al.*, 1995; Van Der Spoel *et al.*, 2005; Abraham *et al.*, 2015), utilizando el campo de fuerza Amber a99SB\*-ILDN (Best & Hummer, 2009; Lindorff-Larsen 2010) y el modelo de agua TIP3P (Jorgensen, 1981). Esta combinación fue seleccionada porque según estudios realizados anteriormente, a99SB\*-ILDN se desempeñó bien para proteínas plegadas, péptidos pequeños desordenados y proteínas de plegamiento rápido (Robustelli 2018). Específicamente para Trp-Cage, el campo de fuerza a99SB\*-ILDN proporciona una curva de fusión en concordancia con la experimental (Figura 6). La temperatura de fusión obtenida con este campo de fuerza es mayor que la experimental (i.e. 350 K vs 315 K). Sin embargo, de todos los campos de fuerza evaluados, es el único que muestra un 100% de plegamiento a bajas temperaturas y casi un 0% de plegamiento a altas temperaturas (Figura 6).

La estructura inicial de Trp-Cage para las simulaciones se tomó del ensemble en disolución obtenido con experimentos de RMN por Neidigh *et al.* (Código PDB: 1L2Y) (Neidigh, 2002). La proteína fue situada en una caja de simulación con forma de dodecaedro truncado cuyos límites estaban al menos a 2.4 nm de la proteína. La caja de simulación se rellenó con moléculas de agua para solvatar la proteína. Las proteínas Trp-Cage y NFK presentan carga total +1 debido a la protonación de sus cadenas laterales. En estos dos sistemas se añadió un anión cloruro para neutralizar la carga eléctrica total del sistema.

Una vez preparado el sistema, se realizó una optimización geométrica para asegurar una geometría adecuada sin impedimentos estéricos al inicio de las simulaciones. Las estructuras resultantes fueron sometidas al procedimiento de equilibración. En todas las simulaciones se empleó un *cutoff* de 1.0 nm para

el cálculo de interacciones electrostáticas y de van der Waals. La primera fase de la equilibración consistió en una simulación de 1 ns bajo condiciones de número de partículas, volumen y temperatura constante (NVT) hasta alcanzar la temperatura deseada de 352 K. El termostato empleado para equilibrar la temperatura fue el CSVR (*canonical sampling with velocity rescaling*) (Bussi, 2007). En la segunda fase de equilibración, se realizó una simulación de 1 ns bajo condiciones de número de partículas, presión y temperatura constante (NPT) a la temperatura 352 K y a una presión de 1 bar. Para conseguir que la presión se equilibrara al valor deseado, se empleó el baróstato Parrinello-Rahman (Parrinello, 1981). Tras estas dos etapas, el sistema se encuentra estable en condiciones de T y P equivalentes a las que se podría realizar un experimento de plegamiento/desplegamiento de la proteína. Debe indicarse que se estableció 352 K como la temperatura de trabajo por corresponder a una temperatura cercana a la temperatura de fusión para el campo de fuerza empleado (Figura 6). Esta elección tiene el propósito de facilitar el muestreo tanto del estado plegado como del desplegado.



**Figura 6.** Comparación de las curvas de fusión obtenidas con diferentes campos de fuerza con la curva de fusión experimental (negro). En rosa se muestra la curva obtenida con el campo de fuerza a99SB\*-ILDN. *Figura tomada de (Robustelli, 2018).*

Durante la equilibración, se restringió el movimiento de los átomos de la proteína mediante potenciales armónicos con constantes de fuerza de  $10^3 \text{ kJ}\cdot\text{nm}^{-2}\cdot\text{mol}^{-1}$  con el objetivo de que el proceso no causara cambios artificiales de la conformación inicial de la proteína obtenida de RMN en disolución. Para permitir la relajación de la proteína a las condiciones de trabajo, se eliminaron las restricciones y se descartaron los primeros 10 ns de todas las simulaciones de producción. Es diagrama de flujo desde la preparación del sistema hasta el análisis de resultados se muestra en el Esquema 1.



**Esquema 1.** Pasos generales en una simulación de DM.

En las primeras simulaciones de producción se observaron varios eventos de desplegamiento de la proteína Trp-Cage. Se observó que la caja de simulación elegida no era lo suficientemente grande como para evitar contactos entre la proteína y sus imágenes periódicas en el estado de máxima extensión. Por este motivo se extrajeron las coordenadas de la proteína en su máxima extensión y sobre esa conformación se generó un nuevo sistema en el que los límites de la caja dodecaédrica de simulación estaban a una distancia mínima de 7.5 nm de la proteína. De esta manera, la distancia mínima con las imágenes periódicas será de

15 nm, mayor que el *cutoff* elegido de 1.0 nm. Los parámetros de esta última caja de simulación se emplearon para repetir la creación del sistema con las coordenadas de la proteína plegada. Posteriormente, los sistemas con la proteína plegada y desplegada fueron sometidos a un nuevo proceso de equilibración como el descrito anteriormente.

Los sistemas descritos para la Trp-Cage se usaron como punto de partida para construir las modificaciones postraduccionales NFK y CMA sobre la cadena lateral del residuo triptófano Trp6 o de la arginina Arg16 respectivamente. Las modificaciones se realizaron con el programa de modelización Pymol (DeLano, 2002).

Con el objetivo de tener un mejor muestreo de los estados plegado y desplegado y de los procesos de plegamiento y desplegamiento, se obtuvieron 3 sistemas (i.e. proteína Trp-Cage, NFK y CMA), con la proteína en dos conformaciones posibles: plegada (RMN) y totalmente extendida. Para cada sistema se realizaron 15 simulaciones de producción con velocidades iniciales aleatorias empezando con cada conformación (i.e. 30 simulaciones por sistema). En el caso de la CMA, se realizaron 17 simulaciones comenzando desde la conformación plegada. Cada una de las simulaciones de producción se extendió como mínimo 200 ns y las estructuras de todo el sistema se guardaron a intervalos de 1 ns. Sumando todas las simulaciones, se muestrearon 11.6  $\mu$ s, 11.4  $\mu$ s y 10.5  $\mu$ s respectivamente para las proteínas Trp-Cage, NFK y CMA.

Para el análisis de las trayectorias y representación de los resultados se emplearon los siguientes programas: Gromacs-2016.4 (Berendsen *et al.*, 1995; Van Der Spoel *et al.*, 2005; Abraham *et al.*, 2015); VMD (Humphrey, 1996); Gnuplot (Williams & Kelly, 2011); Plumed 2.4.0 (Tribello *et al.*, 2014); Pyemma 2.5.6 (Senne *et al.*, 2012). También se escribieron scripts de BASH y programas con el lenguaje AWK para calcular promedios, desviaciones estándar, medias móviles y RMSDs de los datos obtenidos con los programas indicados.

### 3.3. Parametrización de NFK y CMA

Los campos de fuerza para la simulación de proteínas están parametrizados para los residuos estándar. Como las modificaciones postraduccionales alteran la estructura química de los residuos, los residuos modificados deben parametrizarse de forma coherente al campo de fuerza utilizado. La parametrización de una nueva molécula dentro de un campo de fuerza es un proceso delicado puesto que de la calidad de la parametrización dependerá la calidad de las futuras simulaciones. Un aspecto crítico en los campos de fuerza de proteínas es la parametrización de los potenciales de torsión de los ángulos  $\phi$  y  $\psi$  de cada residuo, ya que éstos determinan la propensión a formar los diferentes tipos de estructura secundaria o estructura desordenada.

Por suerte, la formación de N-formilquínurena sobre triptófano y de carboximetilarginina sobre arginina sólo afecta a las cadenas laterales y lejos del carbono alfa. Esto simplifica mucho la parametrización de los residuos NFK y CMA ya que se puede considerar como una aproximación bastante buena que los potenciales de torsión de los diedros  $\phi$  y  $\psi$  son iguales a los de los residuos originales. Así, únicamente es necesario parametrizar las cadenas laterales. Todos los tipos de átomos necesarios para representar la N-formilquínurena y la carboximetilarginina se encuentran en el campo de fuerza a99SB\*-ILDN. Casi todos los parámetros que describen enlaces, ángulos y torsiones también se encuentran en el campo de fuerza a99SB\*-ILDN. En los pocos casos en que no es así, se emplearon los valores correspondientes del campo de fuerza general de amber GAFF (Wang *et al.*, 2004).

Las cargas parciales de los átomos en los residuos modificados se calcularon siguiendo el mismo procedimiento por el que se parametrizaron los residuos estándar en a99SB\*-ILDN. En primer lugar, se modelizó el aminoácido modificado NFK o CMA con sus extremos N-terminal acetilado y C-terminal N-

metilado. De cada aminoácido se generaron estructuras aleatorias con un algoritmo de búsqueda conformacional del programa OpenBabel (O'Boyle *et al.*, 2011). Las 30 conformaciones de menor energía se optimizaron con cálculos según la Teoría del Funcional de la Densidad (DFT) con el funcional B3LYP (Lee *et al.*, 1998), las bases 6-31+G(d,p) y el modelo de disolvente SMD (Marenich *et al.*, 2009) con el programa Gaussian 09 (). Todas las estructuras optimizadas fueron sometidas a un cálculo vibracional para asegurar que fueran mínimos energéticos por la ausencia de frecuencias imaginarias. Tras la optimización geométrica se seleccionaron los 8 confórmeros de menor energía. De cada uno de ellos se calculó el potencial electrostático generado por la distribución de núcleos y electrones con el método de Hartree-Fock con las bases 6-31G(d). El método RESP multiconfiguracional se utilizó para obtener las cargas parciales de los átomos en NFK y CMA a partir de los potenciales calculados. Este método sitúa cargas puntuales sobre las posiciones de los átomos y varía su valor hasta conseguir reproducir el potencial electrostático calculado cuánticamente. Durante ese proceso, las cargas parciales de los átomos del backbone, C $\beta$  y sus hidrógenos se mantuvieron fijas a los valores de los residuos originales Trp y Arg.

### 3.4. Parámetros de orden

En física o química de materiales un parámetro de orden es una medida de los cambios en una transición de fase. Por extensión, en estudios de proteínas se usa el término parámetro de orden como cualquier medida que permite cuantificar el plegamiento o desplegamiento. Sería, por tanto, una aproximación más o menos simple a la coordenada multidimensional real de plegamiento o desplegamiento.

#### 3.4.1. RMSD

La raíz de la desviación cuadrática media (RMSD) se usa para cuantificar la diferencia de una conformación respecto a otra de referencia. Para calcular el RMSD, las estructuras comparadas deben alinearse con movimientos de rotación y/o traslación. Entonces se determina la diferencia de posiciones de los átomos considerados en una conformación respecto a las posiciones de los mismos átomos en la conformación de referencia. Se calcula el cuadrado de dichas referencias para que las desviaciones sólo sumen, y se calcula su promedio. El RMSD es la raíz cuadrada de dicho promedio (ec. 6). Un valor bajo de RMSD significa similitud estructural de las conformaciones comparadas y valores altos indican diferencia estructural.

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n (\mathbf{r}_i - \mathbf{r}_{ref})^2} \quad [6]$$

#### 3.4.2. Radio de giro

El radio de giro es definido como el promedio cuadrático medio de la distancia ponderada en masa de los átomos al centro de masas de la molécula (ec. 7). Para dos proteínas que tienen la misma masa molecular, la que tiene el mayor radio de giro es la más extendida o menos esférica. Por tal razón, el radio de giro se usa como un indicador de compacidad de la estructura de la proteína.

$$Rg = \sqrt{\frac{\sum_{i=1}^n |\mathbf{r}_i|^2 m_i}{\sum_{i=1}^n m_i}} \quad [7]$$

### 3.4.3. Superficie accesible al disolvente (SASA)

La superficie accesible al disolvente (*solvent accessible surface area*, SASA) es el área de la superficie barrida por el centro de una esfera sonda que rueda sobre la superficie de una molécula. (Lesk & Chothia, 1980). Cuando el agua es el disolvente, la molécula de agua se representa por una esfera de radio 1.4 Å. Las moléculas de soluto están representadas por un conjunto de esferas interconectadas de radios de van der Waals apropiados para cada átomo. Este parámetro será útil en el estudio de la afectación del núcleo hidrofóbico de Trp-Cage por la modificación en los residuos 6 y 16.

### 3.4.4. Medida de la estructura secundaria (DSSP)

El algoritmo DSSP (Define Secondary Structure of Proteins) (Kabsch & Sander, 1983) es el método estándar para la asignación de la estructura secundaria de los aminoácidos de una proteína, dadas sus coordenadas de resolución atómica (Touw *et al.*, 2014). El algoritmo y programa del mismo nombre calcula la energía de enlace de hidrógeno entre todos los átomos del backbone dadas las posiciones de los átomos en una proteína. Los dos mejores enlaces de hidrógeno para cada átomo se usan para determinar la estructura secundaria más probable de cada residuo.

## 3.5. Superficie de energía libre

La superficie de energía libre (*free energy Surface*, FES) es la colección de valores de energía libre correspondiente a cada posible configuración de las N partículas que forman el sistema. Es decir, es una hipersuperficie 3N dimensional. Así, para analizar la variación de energía libre de un proceso determinado es necesario proyectar la hipersuperficie en una o dos coordenadas adecuadas que permitan distinguir todos los estados relevantes (e.g. estado plegado y desplegado, reactivos y productos, etc.). Las coordenadas sobre las que se proyecta la energía libre son parámetros de orden o variables colectivas que intentan aproximar la coordenada de reacción multidimensional real.

La FES de un sistema proporciona información muy importante ya que permite caracterizar la termodinámica de un sistema. Por ejemplo, permite identificar las configuraciones más representativas por ser las de menor energía y más pobladas. A partir de las cuales es posible calcular observables macroscópicos como promedios ponderados por la distribución de poblaciones. Igualmente, es posible determinar la contribución de cada configuración (i.e. estado microscópico) en las propiedades macroscópicas. La FES también permite calcular las barreras energéticas entre los estados más poblados. En combinación con la Teoría del Estado de Transición, las barreras de energía permiten estimar la cinética de los procesos observados.

A pesar de la relevancia que tiene la FES, las técnicas experimentales sólo permiten una descripción parcial y normalmente macroscópica. Al contrario, la reconstrucción de la FES es relativamente sencilla a partir de simulaciones computacionales. Esta es una de las características sobresalientes de las simulaciones

de DM (Tuckerman, 2007). A partir de las configuraciones muestreadas en una simulación, la energía libre  $F(s)$  en función de una variable colectiva  $s$  se puede calcular como

$$F(s_i) = -RT \ln P(s_i) + C \quad [8]$$

Donde  $R$  es la constante de los gases,  $T$  la temperatura absoluta y  $P(s)$  es la población de la configuración  $s_i$ . El término  $C$  es una constante cuyo valor no es relevante para el cálculo de energías relativas. En sistemas condensados, la variación de volumen por fluctuaciones de la presión es tan pequeña, que se puede asumir como válida la ecuación 8 para el cálculo de la energía libre de Gibbs  $G(s)$ . En la práctica, una vez realizada la simulación, debe elegirse un parámetro de orden  $s$  apropiado, discretizar el espacio de muestreo de  $s$  y contar las veces que el sistema está en cada configuración posible  $s_i$  para obtener la población  $P(s_i)$ .

La función biológica de las proteínas depende tanto de su estructura como de su dinámica, lo que a su vez está conectado a la FES. Para calcular las FES de Trp-Cage y de sus variantes modificadas, se probaron diversas variables colectivas con el objetivo de determinar: 1) las que muestran mejor separación de los estados plegado y desplegados; y 2) las que mejor describieran el proceso de plegamiento/desplegamiento. Se usaron como variables colectivas: 1) RMSD de los carbonos  $\alpha$  de toda la proteína; 2) el RMSD de los átomos del backbone de los residuos 2 a 8 que forman la hélice  $\alpha$  en la Trp-Cage en su plegamiento nativo; 3) el RMSD de los átomos del backbone de los residuos 11 a 14 que forman la hélice  $3_{10}$  en la Trp-Cage en su plegamiento nativo. La FES se representó frente a diferentes combinaciones 2 a 2 de estas variables colectivas.

### 3.6. Modelos de estados de Markov para la estimación de cinéticas de plegamiento y desplegamiento

Los modelos de estados de Markov permiten analizar la dinámica de sistemas. En este caso, trayectorias de DM. La aplicación original a DM se puede encontrar en (Prinz *et al.*, 2011) y una revisión exhaustiva en (Bowman, Pande, Noé, 2014). Recientemente, estos métodos se han aplicado con gran éxito a la descripción de dinámica de proteínas (Chodera & Noé, 2014; Plattner & Noé, 2015) o interacción entre proteínas y fármacos (Shukla *et al.*, 2014; Plattner & Noé, 2015) entre otros (Husic & Pande, 2018).

El método requiere definir estados entre los que calcular el número de transiciones y los tiempos de las mismas. Esto se denomina discretización espacial. Habitualmente la discretización se realiza con métodos de clúster automáticos para clasificar conformaciones. Uno de los más usados es el método de clúster de  $k$ -medias porque permite asignar todas las conformaciones de las trayectorias a un número de estados definido por el usuario. Así se puede variar de forma sistemática el número de estados y controlar su influencia en la descripción dinámica del sistema. Sin embargo, no es una condición necesaria usar métodos de clúster. Además, no todos los estados deben corresponder a mínimos en la hipersuperficie de energía libre. De hecho, la partición del espacio tiene que ser completa de forma que no haya solapamiento entre estados para que todas las conformaciones pertenezcan únicamente a un estado, incluyendo las de más alta energía.

En este punto debe indicarse que comportamiento Markoviano indica que la probabilidad de transición entre estados no depende de los estados visitados previamente. Para que esta condición se cumpla debe permitirse que el sistema se relaje tras cada transición. Esto requiere definir un tiempo de espera o "lag time"  $\tau$ . Definidos el número de estados  $n$  y el lag time, se analizan las trayectorias contando cuántas transiciones y entre qué estados se dan tras cada intervalo de tiempo  $\tau$ .

La cuenta de las transiciones se usa para construir una matriz  $n \times n$  de probabilidad de transición entre todos los estados. El primer vector propio de esta matriz proporciona la población de cada estado en el equilibrio. Los posteriores vectores propios indican entre qué estados ocurren los principales modos de relajación del sistema, ordenados del más lento a más rápido según sus valores propios. De hecho, el tiempo de relajación  $t_i$  de cada modo de relajación se determina a partir del valor propio correspondiente  $\lambda_i$ :

$$t_i = -\frac{\tau}{\ln \lambda_i} \quad [9]$$

Tanto el lag time como el número de estados elegidos introducen cierto error en la descripción dinámica del sistema. El lag time determina la resolución temporal con la que se describe la dinámica, de forma que no se pueden describir eventos que ocurren en tiempos inferiores a  $\tau$  por el simple hecho de que sólo se analiza el estado del sistema a intervalos de tiempo igual a  $\tau$ . Por tanto, un lag time elevado implicará comportamiento Markoviano, pero la descripción de la dinámica tendrá baja resolución temporal. Además, cuanto mayor sea el lag time, menor número de transiciones se observarán en una trayectoria de longitud finita. Por otra parte, si se elige un número pequeño de estados, éstos corresponderán a conformaciones significativamente diferentes, con tiempos de transición grandes, y típicamente se observarán muy pocas transiciones en las trayectorias. Al contrario, si se elige un número muy grande de estados, éstos se encontrarán cinéticamente próximos, pero cada estado estará formado por pocas conformaciones. Así, el número de transiciones observado también será bajo porque cada conformación se visita pocas veces. En definitiva, es necesario elegir el número de estados y el lag time de forma que se observe un comportamiento Markoviano con la mayor resolución temporal posible y con el mayor número de transiciones posibles entre estados.

Finalmente, la ventaja que ofrece este método es que permite combinar múltiples trayectorias relativamente cortas para describir eventos que ocurren a escalas temporales mayores. Esto permite comenzar las trayectorias desde conformaciones diferentes para facilitar el muestreo del espacio configuracional. Esta aproximación suele ser más eficiente y proporciona menor error estadístico que simular una única trayectoria muy larga y esperar que los eventos lentos se repitan un número mínimo de veces. Típicamente se acepta que un modelo de estados Markoviano puede describir eventos que ocurren en una escala temporal igual al tiempo total de las múltiples trayectorias empleadas.



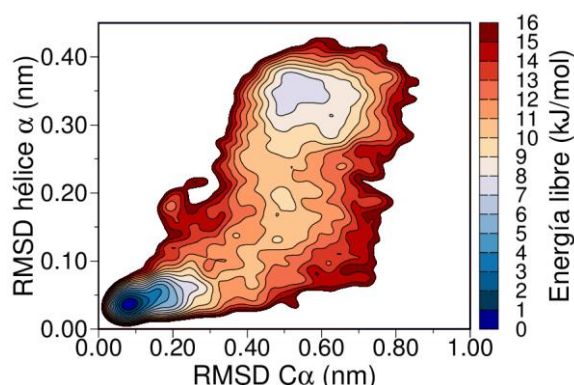
## 4. RESULTADOS Y DISCUSIÓN

### 4.1. Variables Colectivas y muestreo de las trayectorias

Para describir el proceso de plegamiento/desplegamiento, debemos poder cuantificarlo. Esto requiere elegir una aproximación a la coordenada de reacción real. Para diferenciar la coordenada de reacción real de la coordenada de reacción aproximada, llamaremos a esta última variable colectiva (VC) o variables colectivas (VCs) si usamos una combinación de varias. Una buena VC debe permitir distinguir el estado plegado, desplegado, los posibles intermedios y estados de transición. Por eso, también permitirá decidir si el muestreo de las trayectorias es satisfactorio.

Como se indica en la sección metodológica, se consideraron tres posibles VCs. En primer lugar, el RMSD de todos los  $C\alpha$  medido respecto al plegamiento nativo de Trp-Cage obtenido por NMR (PDB 1L2Y). Valores bajos de esta VC indican la formación de estructura secundaria y terciaria (i.e. plegamiento nativo), y valores altos indican desplegamiento total. El problema es que los valores intermedios pueden deberse a pérdida parcial de la estructura terciaria manteniendo los elementos de estructura secundaria, pérdida parcial de los elementos de estructura secundaria manteniendo una forma cercana a la estructura terciaria plegada, o cualquier situación intermedia. En segundo y tercer lugar, el RMSD de los átomos del backbone de los residuos 2 a 8 que forman la hélice  $\alpha$  y el RMSD de los átomos del backbone de los residuos 11 a 14 que forman la hélice  $3_{10}$  medidos respecto al plegamiento nativo indican el grado de formación de estos elementos de estructura secundaria, pero no dan información acerca del grado de formación de estructura terciaria. Por eso, se probaron diferentes combinaciones de estas tres VCs para describir el proceso de plegamiento lo más detalladamente posible.

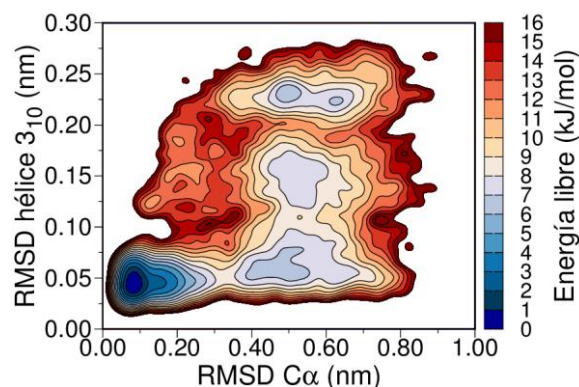
La combinación RMSD de la hélice  $\alpha$  vs RMSD de  $C\alpha$  se ha reportado ampliamente en la literatura como coordenadas de reacción para el plegamiento de Trp-Cage (Juraszek2006, Meuzelaar *et al.*, 2013; Kannan, & Zacharias, 2009; Kim *et al.*, 2016). A continuación, se muestra la superficie de energía libre (FES) obtenida de las trayectorias de Trp-Cage (Figura 7).



**Figura 7.** Superficie de energía libre del plegamiento de Trp-Cage representada en función del RMSD de la hélice  $\alpha$  y del RMSD de los  $C\alpha$ .

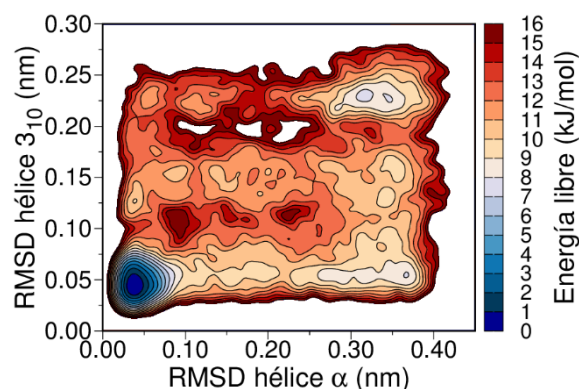
Esta combinación permite distinguir claramente el estado plegado, mostrado como el mínimo absoluto (en azul intenso  $\text{RMSD}(C\alpha) < 0.20 \text{ nm}$ ,  $\text{RMSD}(\text{hélice } \alpha) < 0.10 \text{ nm}$ , Figura 7), del estado desplegado, mostrado como el mínimo local (en azul claro  $\text{RMSD}(C\alpha) > 0.40 \text{ nm}$ ,  $\text{RMSD}(\text{hélice } \alpha) > 0.30 \text{ nm}$ , Figura 7). Sin embargo, no se observan otros mínimos locales (intermedios de reacción) y únicamente hay un canal de reacción entre el estado plegado y desplegado.

Otra posibilidad es usar el RMSD de los residuos que forman la hélice  $3_{10}$  en lugar de los que forman la hélice  $\alpha$ . En ese caso, se observa que entre el estado plegado (azul intenso, valores bajos de los dos RMSDs, Figura 8) y el desplegado (azul claro, RMSD (hélice  $3_{10}$ ) > 0.20 nm, Figura 8), aparecen dos intermedios como mínimos locales de la FES.



**Figura 8.** Superficie de energía libre del plegamiento de Trp-Cage representada en función del RMSD de la hélice  $3_{10}$  y del RMSD de los  $C\alpha$ .

La combinación de RMSD de la hélice  $\alpha$  con el RMSD de la hélice  $3_{10}$  proporciona una FES aún más detallada (Figura 9). Además del estado plegado (azul intenso, valores bajos de los dos RMSDs, Figura 9), y del estado desplegado (azul claro, RMSD (hélice  $\alpha$ ) > 0.30 nm, RMSD (hélice  $3_{10}$ ) > 0.20 nm, Figura 9), se pueden distinguir 4 zonas con mínimos locales de energía más o menos definidos. Se consideró que algunos mínimos locales próximos, que presentaban conformaciones similares, podían agruparse en un mismo macroestado si además eran próximos en energía (Figura 9).

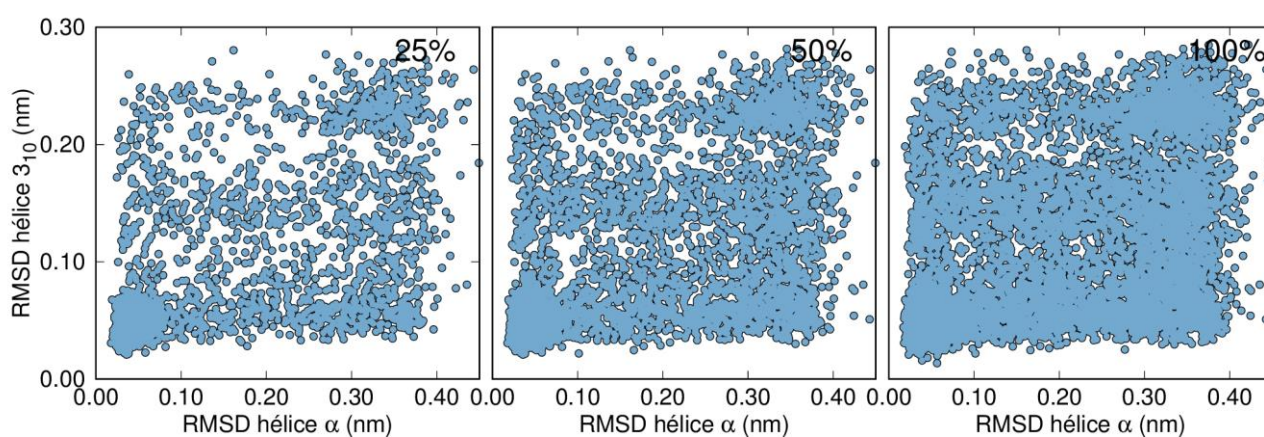


**Figura 9.** Superficie de energía libre del plegamiento de Trp-Cage representada en función del RMSD de la hélice  $3_{10}$  y del RMSD de los  $C\alpha$ .

También se probó usar el RMSD de los residuos que forman la secuencia de prolina en la región C-terminal de Trp-Cage, pero se observó que estaba muy correlacionada con el RMSD( $C\alpha$ ) y no proporcionaba información adicional.

Un factor adicional a considerar es que el RMSD( $C\alpha$ ) y el RMSD (hélice  $\alpha$ ) están bastante correlacionados. Como se puede apreciar en la Figura 7, el aumento o disminución de una variable implica el aumento o disminución de la otra. Esto implica que el RMSD (hélice  $\alpha$ ) informa indirectamente del plegamiento total de Trp-Cage. Por tanto, la combinación de VCs que mejor captura el plegamiento total con la descripción más detallada de los intermedios es RMSD (hélice  $3_{10}$ ) y RMSD (hélice  $\alpha$ ).

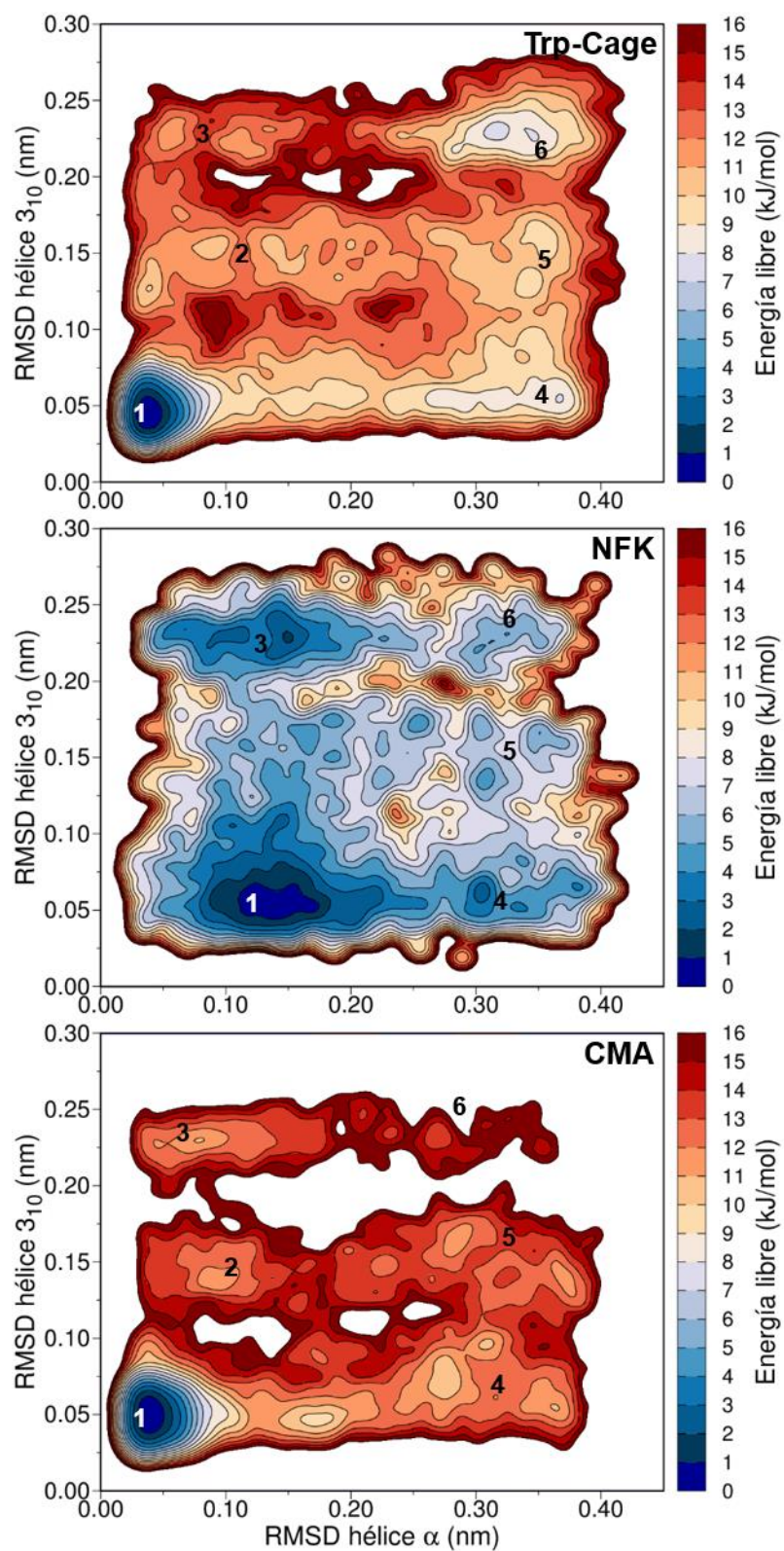
El muestreo de las trayectorias en el espacio de las VCs escogidas es satisfactorio para describir la FES de Trp-Cage y sus variantes CMA y NFK. En la Figura 10, cada punto de cada representación corresponde a una conformación de una trayectoria guardada en disco para su posterior análisis. En cada representación se muestra un porcentaje diferente de conformaciones de todas las trayectorias de Trp-Cage. Como se aprecia, incluso con sólo el 25% de las conformaciones, el muestreo de los mínimos locales observados en la Figura 9 es casi completo. Análogamente, el muestreo del espacio de las VCs de CMA y NFK también es satisfactorio (Figura A1 y A2).



**Figura 10.** Muestreo considerando el 25% (izquierda), 50% (centro) y 100% (derecha) de las conformaciones de cada trayectoria, para todas las trayectorias simuladas de Trp-Cage.

En el caso de NFK, las trayectorias iniciadas desde el plegamiento nativo de NMR mostraron desplegamiento de la proteína hasta un estado mal plegado en pocos ns. Además, las trayectorias iniciadas desde el estado desplegado adquieren el mal plegamiento también en pocos ns. Sin embargo, Trp-Cage y CMA requirieron mucho más tiempo en promedio. Para intentar observar el mayor número de eventos completos de plegamiento o desplegamiento, algunas de las trayectorias se extendieron hasta 1  $\mu$ s. La baja frecuencia de eventos completos de plegamiento y desplegamiento demuestra que la observación directa no es la mejor estrategia para estudiar la cinética y mecanismos de plegamiento si se quiere hacer con significancia estadística. A pesar de todo, se observaron respectivamente 29, 23 y 22 plegamientos o desplegamientos completos en las trayectorias de Trp-Cage, NFK y CMA.

## 4.2. Termodinámica de plegamiento



**Figura 11.** Superficies de energía libre del plegamiento representada en función del RMSD de la hélice 3<sub>10</sub> y del RMSD de los C $\alpha$ .



En todas las superficies se observa un mínimo de energía absoluto diferenciado del resto de mínimos de energía. En Trp-Cage y CMA el mínimo absoluto se encuentra en valores relativamente bajos de RMSD y está formado por aquellas conformaciones similares a la estructura nativa obtenida por NMR. Es decir, corresponde al estado plegado correcto. Así como la modificación de Arg16 por CMA no impide el plegamiento, en NFK el mínimo está desplazado a mayores valores de RMSD hélice  $\alpha$ , indicando menor similitud con el plegamiento nativo. Por otra parte, en Trp-Cage y CMA el mínimo está muy localizado y la energía aumenta rápidamente al alejarse de él, indicando que el plegamiento nativo será la conformación predominante de este estado energético. Por otra parte, en NFK la energía varía poco en los alrededores del mínimo absoluto, indicando que a la misma temperatura la proteína podrá adquirir mayor diversidad conformacional. Más adelante se observará que el estado plegado de NFK debe describirse como “mal plegado”, del inglés *misfolded*.

Además del mínimo absoluto, para todos los casos se observan otras 4 o 5 regiones con mínimos locales o que agrupan mínimos locales cercanos. Todos estos mínimos diferentes del estado 1 corresponden a conformaciones en las que la hélice  $3_{10}$  y/o hélice  $\alpha$  están más o menos formadas. Por tanto, en general los estados 2 a 6 corresponden a conformaciones desplegadas. En NFK no se observa claramente el estado 2, por lo que se ha omitido y se ha seguido con misma la numeración de estados de Trp-Cage y CMA.

En Trp-Cage, el estado 6, que corresponde al de menor contenido en estructura secundaria (i.e. más desplegado), es el más estable entre los desplegados. Este estado es  $\sim 7$  kJ/mol más energético que el estado plegado. Por otra parte, los estados desplegados menos estables son  $\sim 11$  kJ/mol más energéticos que el estado plegado (estado 3, Figura 11). Por el contrario, en CMA el estado 6 es el menos estable a  $\sim 13$  kJ/mol del estado plegado, y el estado desplegado más estable es el 4 ( $\sim 10$  kJ/mol, Figura 11). En NFK, el estado desplegado más estable es el 3, a sólo  $\sim 1$  kJ/mol del estado mal plegado. Además, el resto de estados desplegados también presentan bajas energías a sólo 2-5 kJ/mol del estado plegado. Ya que a la temperatura de simulación la energía térmica es de  $\sim 2.9$  kJ/mol, en el equilibrio NFK mostrará una población significativa de conformaciones en diferentes estados desplegados.

Para Trp-Cage, la barrera de energía mínima para escapar del estado 1 plegado es de  $\sim 10$  kJ/mol. En CMA es  $\sim 12$  kJ/mol. Estas dos barreras son significativamente superiores a la energía térmica. En ambos casos, las barreras de energía desde los estados desplegados hacia el plegado son bastante inferiores. Todo esto indica que el estado plegado de Trp-Cage y CMA es cinéticamente estable. Sin embargo, en NFK la barrera de energía para escapar del estado mal plegado es sólo  $\sim 5$  kJ/mol, y la barrera de energía para entrar en el estado mal plegado es  $\sim 4$  kJ/mol. Además de mostrar mayor diversidad conformacional por presentar un estado mal plegado más amplio y con estados desplegados más próximos en energía, la cinética de plegamiento y desplegamiento será mucho más rápida que en Trp-Cage y CMA.

Estos resultados muestran que diferentes modificaciones pueden tener consecuencias muy diferentes, favorecer o desfavorecer el plegamiento nativo y acelerar o ralentizar los equilibrios de plegamiento, desplegamiento o entre estados desplegados. Para entender mejor las bases moleculares de estos resultados se describirán las estructuras de los diferentes estados.

### 4.3. Estado plegado

Las estructuras del estado plegado de la Trp-Cage y CMA, se extrajeron de los mínimos energéticos absolutos de las FES respectivas. Para NFK se emplearon las superficies de energía libre de los RMSD de la hélice  $\alpha$  y de la hélice  $3_{10}$ .

### 4.3.1. Características estructurales

Los elementos de la estructura secundaria de Trp-Cage en su plegamiento nativo incluyen una hélice  $\alpha$  en los residuos 2 a 8, seguida de una hélice  $3_{10}$  en los residuos 11 a 14, y una hélice de poliprolina C-terminal. Empleando el programa DSSP se asignó el contenido de estructura secundaria para cada conformación plegada (estado 1, Figura 11) de todas las trayectorias de cada proteína (Tabla 1).

**Tabla 1.** Contenido en estructura secundaria del estado plegado de Trp-Cage, NFK y CMA.

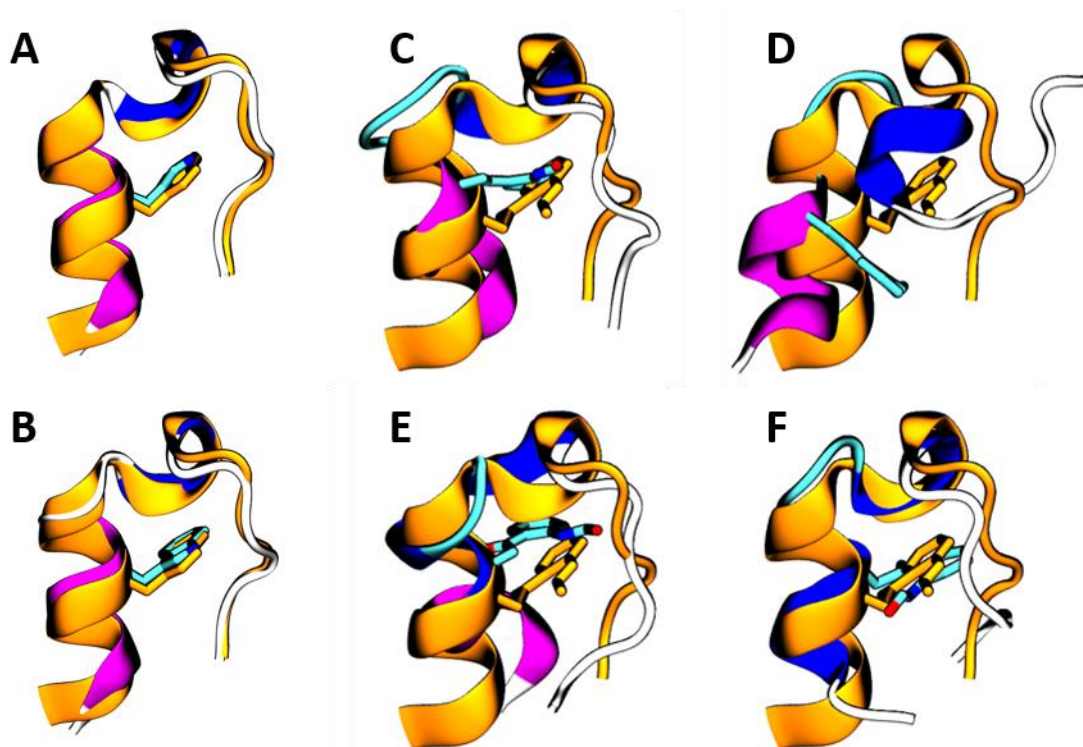
Proteína	# Residuos Hélice $\alpha$	# Residuos Hélice $3_{10}$
<i>Trp-Cage</i>	$6.3 \pm 0.9$	$3.1 \pm 0.5$
<i>NFK</i>	$2.5 \pm 1.8$	$1.6 \pm 1.2$
<i>CMA</i>	$6.2 \pm 0.8$	$3.1 \pm 0.9$

Existe una diferencia significativa entre el número de residuos que adoptan la conformación de hélice  $\alpha$  y hélice  $3_{10}$  de Trp-Cage y la proteína modificada en el residuo Arg16 (CMA) con respecto a la modificada en el Trp6 (NFK). Este es un primer indicio de que la modificación en el triptófano altera considerablemente la estructura secundaria de la proteína. El estado plegado de Trp-Cage y CMA contiene en su estructura un 89 % y 88 % de hélice  $\alpha$  respectivamente en los residuos del 2 al 8 (Tabla 1). Valores que se ven contrastados por un 34 % correspondiente a NFK. De igual manera para la estructura secundaria hélice  $3_{10}$ , el NFK presenta un número de residuos inferior que adoptan dicha conformación (38 %), con respecto a los observados en la Trp-Cage (78 %) y CMA (77 %).

Las diferencias en contenido de estructura secundaria también se observan en los valores de RMSD(C $\alpha$ ), que además indica sobre la formación de estructura terciaria. Los valores respectivos de Trp-Cage no modificada, NFK y CMA en sus estados plegados son  $0.10 \pm 0.04$  nm,  $0.46 \pm 0.16$  nm y  $0.11 \pm 0.04$  nm. Los valores de Trp-Cage y CMA son equivalentes al RMSD de las estructuras del ensemble experimental (i.e.  $0.17 \pm 0.03$  nm para PDB 1L2Y), lo que permite afirmar que el estado plegado obtenido en las simulaciones de estas dos especies es el nativo correcto. Por el contrario, NFK muestra valores de RMSD(C $\alpha$ ) bastante mayores, lo que indica que su estado plegado se diferencia significativamente del nativo. Por este motivo, se le denomina estado mal plegado. Además, la desviación estándar elevada indica mayor diversidad conformacional en el estado plegado de NFK.

Para verificar dicha suposición, y determinar la homogeneidad o heterogeneidad de los estados plegados de cada proteína, se realizaron análisis de clúster con un *cutoff* de 0,2 nm. Para Trp-Cage y CMA, más del 99.9% de las estructuras corresponden al plegamiento nativo, indicando así, la uniformidad de sus estados plegados. Sin embargo, para NFK se obtuvieron 325 clústeres. Si consideramos que cada clúster obtenido es una conformación significativa del estado, el hecho de que sea necesario considerar 267 conformaciones para abarcar el 90% de la población evidencia la amplia diversidad de estructuras del estado plegado de NFK.

La Figura 12 muestra las conformaciones (clústeres) más importantes de Trp-Cage, CMA y NFK. Mientras que en Trp-Cage y CMA los elementos de estructura secundaria se aprecian muy bien definidos, en NFK se observan hélices  $\alpha$  y  $3_{10}$  parcialmente formadas. Esto se relaciona con los resultados obtenidos con el programa DSSP, que mostró que el porcentaje de residuos que se encuentran dispuestos como estructura secundaria en NFK es mucho menor con respecto a las otras dos proteínas.



**Figura 12.** Estructuras representativas del estado plegado de (A) Trp-Cage, (B) CMA y (C-F) NFK, superpuestas con la primera estructura del ensemble de RMN PDB 1L2Y (en naranja). Se muestran los 4 primeros clústeres de NFK, que representan el 15% de la población del estado mal plegado.

Otro parámetro de orden que cuantifica la movilidad de los residuos es el RMSF. Los RMSFs observados para Trp-Cage y CMA son bajos ( $\sim 0.1$  nm, Figura A3), consistentes con fluctuaciones normales de una única conformación plegada dominante. Por el contrario, el RMSF de los residuos de NFK es bastante mayor ( $\sim 0.4$  nm, Figura A3), consistente con grandes fluctuaciones entre múltiples conformaciones.

### 4.3.2. Núcleo Hidrofóbico

En general, las trayectorias mostraron una tendencia ya constatada por otros autores (Meuzelaar *et al.*, 2013, Andryushchenko & Chekmarev, 2017), en donde inicialmente se forman los elementos de estructura secundaria, seguido por las diferentes interacciones hidrofóbicas en torno al triptófano. El núcleo hidrofóbico de Trp-Cage está compuesto por las interacciones que se dan entre Trp6 y las cadenas laterales de los residuos Pro12, Pro18, Pro19, Leu7 y Tyr3. Adicionalmente a las interacciones hidrofóbicas, se tuvieron en cuenta otras interacciones que ayudan a la formación de la caja de Trp6. En concreto el enlace de hidrógeno entre el oxígeno carbonílico de Arg16/CMA16 y el nitrógeno indólico de Trp6/NFK6 y el puente salino de Arg16/CMA16 con Asp9. La Tabla 2 muestra las distancias entre Trp6 y los residuos clave mencionados.

Según los resultados presentados, no hay una diferencia significativa en la disposición del núcleo hidrofóbico de Trp-Cage y el de la proteína modificada en el residuo Arg16 (i.e. CMA) (Tabla 2). Caso contrario al observado para la proteína NFK, en donde se evidenció un aumento considerable en las distancias, especialmente de las prolinas hacia el triptófano modificado NFK6.

En la Figura 13, se puede observar claramente la distorsión de la caja hidrofóbica generada por la oxidación de Trp6. Se evidencia la modificación de las interacciones hidrofóbicas importantes. Por ejemplo, las

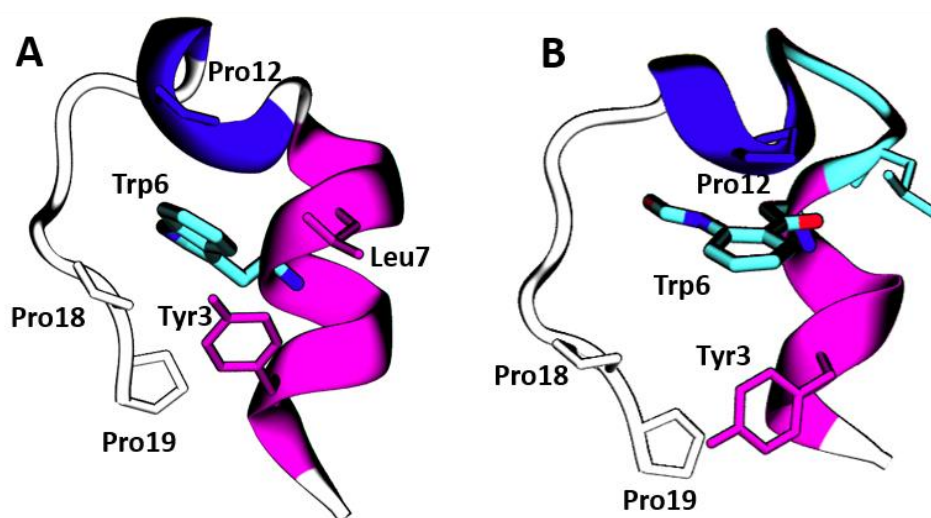
posiciones de Pro19 y Tyr3 se encuentran totalmente desalineadas del NFK6, generando un espacio en el que fácilmente pueden acceder moléculas de agua y de esta manera desestabilizar el núcleo hidrofóbico.

Otra interacción importante que se vio afectada de gran manera por la modificación del residuo 6, es la de Arg16 con el residuo NFK6. Las distancias promedio obtenidas tanto para Trp-Cage como para CMA estuvieron en el orden de 0.31 nm, en cambio para NFK el valor de esta distancia fue 3 veces mayor, indicando así la ausencia de alguna interacción.

El puente salino que se forma entre Asp9 y Arg16, puede ayudar a proteger el núcleo hidrofóbico. A pesar de que el promedio de las distancias entre dichos residuos para Trp-Cage y CMA, fueron mucho menores que las obtenidas para NFK, las desviaciones estándares para los tres casos fueron muy altas, evidenciando fluctuaciones significativas en esa interacción. Esta observación sobre el puente salino también ya se había informado en un estudio anterior (Kannan & Zacharias, 2014).

**Tabla 2.** Interacciones relevantes del núcleo hidrofóbico de Trp-Cage y sus variantes. Las distancias que involucran a Trp6 están medidas desde los centros geométricos de las cadenas laterales de los residuos indicados.

Interacción	Distancias (nm)		
	Trp-Cage	NFK	CMA
<i>Trp6 – Tyr3</i>	0.59 ± 0.06	0.66 ± 0.19	0.59 ± 0.06
<i>Trp6 – Leu7</i>	0.59 ± 0.60	0.74 ± 0.14	0.59 ± 0.60
<i>Trp6 – Pro12</i>	0.54 ± 0.04	1.07 ± 0.51	0.54 ± 0.04
<i>Trp6 – Pro18</i>	0.42 ± 0.04	1.18 ± 0.54	0.43 ± 0.04
<i>Trp6 – Pro19</i>	0.60 ± 0.10	1.28 ± 0.56	0.60 ± 0.10
<i>Trp6 – Arg16</i>	0.31 ± 0.05	1.03 ± 0.48	0.31 ± 0.04
<i>Asp9 - Arg16</i>	0.52 ± 0.23	0.95 ± 0.45	0.54 ± 0.25



**Figura 13.** Caja hidrofóbica de A) Trp-Cage B) NFK.

### 4.3.3. Superficie accesible al disolvente

Para ampliar la comprensión de cómo se modifican las interacciones hidrofóbicas con la modificación de residuos, se calculó el área de la superficie accesible al disolvente (SASA) de la cadena lateral del residuo 6 (Trp/NFK), de la cadena lateral del residuo 16 (Arg/CMA), y de la superficie total de la proteína.



Con el objetivo de realizar una comparación adecuada entre las cadenas laterales con y sin modificaciones, se reportan las SASA relativas a la SASA de la cadena lateral del aminoácido aislado en disolución acuosa. Los aminoácidos y sus variantes modificadas, todos acetilados en el N-terminal y N-metilados en el C-terminal, se simularon en disolución acuosa bajo las mismas condiciones que las proteínas durante 20 ns y posteriormente se calculó la SASA promedio de la cadena lateral correspondiente. Las SASAs de los aminoácidos aislados se usaron para normalizar las SASAs de los residuos (Tabla 3).

**Tabla 3.** Superficies accesibles al disolvente total y de los residuos 6 y 16 para cada proteína.

Proteína	Superficie accesible al disolvente (%)		
	Total	Trp6/NFK6	Arg6/CMA6
<i>Trp-Cage</i>	67 ± 3	10 ± 5	54 ± 22
<i>NFK</i>	75 ± 6	44 ± 18	64 ± 28
<i>CMA</i>	66 ± 3	9 ± 5	56 ± 20

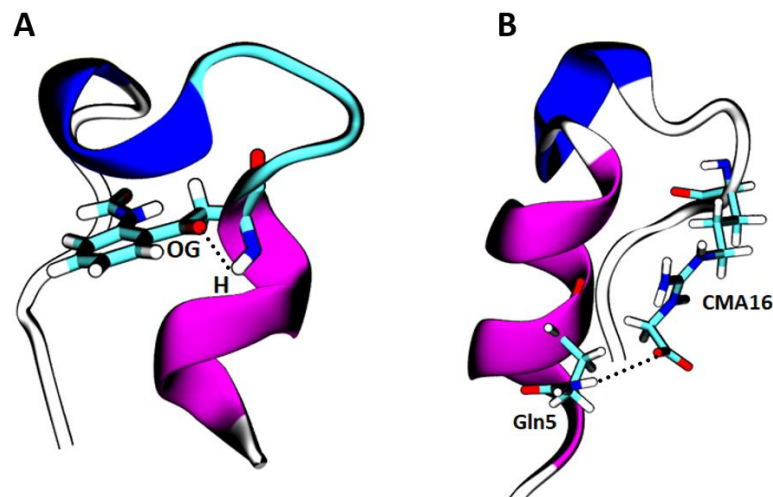
El porcentaje de la superficie total expuesta al disolvente de la proteína NFK es mayor que la de Trp-Cage y la de CMA. Esto se relaciona con la ausencia parcial de la jaula hidrofóbica para la proteína con el residuo 6 modificado, lo que facilita el acceso del disolvente hacia los alrededores del residuo, y de esta manera aumenta la superficie total. El aumento de SASA total de la proteína está correlacionado con cierto desplegamiento o aumento de la extensión de la proteína, como se aprecia en el incremento del 14% en el radio de giro de NFK respecto a Trp-Cage y CMA (Tabla A1).

Hay una diferencia notoria de SASAs de NFK6 (44 %), con respecto a los obtenidos para Trp6 en Trp-Cage (10 %) y CMA (9 %), por lo que es posible indicar que la modificación de este residuo influye de manera importante en el valor de superficie total obtenido para NFK. Aún más teniendo en cuenta la poca diferencia de los valores de la SASA para el residuo Arg16/CMA16 entre las tres proteínas.

Existe una leve diferencia entre la SASA de la cadena lateral del Trp6 de Trp-Cage (10 %), con respecto a la de CMA (9 %). El menor valor en CMA puede ser explicado porque en esa proteína, el residuo 16 modificado tiene un tamaño mayor que el no modificado y, adicionalmente, se encuentra en una posición adecuada para proteger al triptófano de las moléculas de agua. Además, se observaron interacciones propias de CMA entre el oxígeno del carboxilato de CMA16 y el grupo NH de la Gln5 de la región N terminal (Figura 14 B).

En NFK las distancias entre Arg16 y NFK6 son superiores a las de CMA y Trp-Cage (Tabla 2), por lo que la SASA de Arg16 es mayor. Además, los altos valores de desviación estándar para dicha superficie reflejan la elevada movilidad del aminoácido en la cadena proteica.

Una interacción adicional importante para destacar de la proteína NFK, es la que se da en el residuo 6 modificado con su propio grupo amida del backbone mediante el oxígeno carbonílico OG producto de la oxidación. El promedio de la distancia entre ambos átomos de las estructuras en el estado plegado es de 0,31 ± 0,08 nm. Esta nueva interacción, que es inexistente en Trp-Cage y CMA, influye en la distorsión de la caja hidrofóbica. La oxidación del Trp6 a NFK6 rompe el anillo de pirrol e introduce 2 grupos carbonilos, favoreciendo la aparición de enlaces de hidrógeno con otros grupos de la proteína y con moléculas de agua, pero disminuyendo la estabilidad de las interacciones hidrofóbicas.



**Figura 14.** A) Interacción del oxígeno OG de la cadena lateral de NFK6 con su H amídico del backbone.  
 B) Interacción entre el carboxilato de CMA16 con el grupo amida de la cadena lateral de Gln5.

#### 4.4. Estados desplegados

Se realizó un análisis de clúster de las estructuras que conforman cada estado termodinámico (Figura 11), los cuales se definieron con base en los valores de RMSD de la hélice  $3_{10}$  y RMSD de la hélice  $\alpha$  de cada proteína. En la Figura 15 se presentan las estructuras centrales obtenidas de los clústeres más poblados. Las estructuras mostradas son las representativas de la estructura mayoritaria. Sin embargo, el porcentaje máximo que representa cada estructura central del clúster más poblado es del 19 % para Trp-Cage, 13 % para NFK y 18% para CMA. (Para más detalles ver Tabla A2).

Para las tres proteínas en el estado 3 se aprecia claramente la ausencia de la hélice  $3_{10}$ , lo que es coherente con los resultados de la asignación de estructura secundaria (Tabla 1), donde se obtuvieron menos de 0.3 residuos en hélice  $3_{10}$  para Trp-Cage, NFK y CMA. Estas estructuras también son consistentes con los valores altos de RMSD de la hélice  $3_{10}$  en los estados 3 de las FES de cada proteína (Figura 11). Además, cualitativamente se puede observar que la hélice  $\alpha$  en la estructura central del clúster más poblado de CMA, está más definida que la de Trp-Cage y NFK; observación que es nuevamente respaldada por la asignación de estructura secundaria, donde se obtuvo un 45 % de residuos en hélice  $\alpha$  para Trp-Cage, un 36 % para NFK y un 48 % para CMA.

A pesar de que en el estado 4 hay muy pocos residuos dispuestos en la conformación de hélice  $\alpha$  para las tres proteínas, CMA es la que presenta un mayor número de residuos (0.5), seguido de Trp-Cage (0.2) y finalmente NFK para el cual se reportaron 0 residuos con dicha estructura secundaria. Sin embargo, cuando se analizan la cantidad de residuos dispuestos en hélice  $3_{10}$  en el mismo estado, se nota que para CMA hay menos cantidad de residuos (0.6), que en Trp-Cage (1.3) y que en NFK (1.5).

Es evidente la ausencia de elementos de estructura secundaria en los estados 5 y 6 en las tres proteínas. Para el estado 6 la cantidad de residuos dispuestos en hélice  $3_{10}$  es menor que los observados en el estado 5. (Tabla 4).

En los estados 1, 2 y 3 se observan un número mayor de residuos en hélice  $\alpha$  en las tres proteínas, lo que tiene concordancia con la ubicación de dichos estados en las FES, que se encuentran en la parte izquierda de las superficies a valores bajos de RMSD de hélice  $\alpha$ .

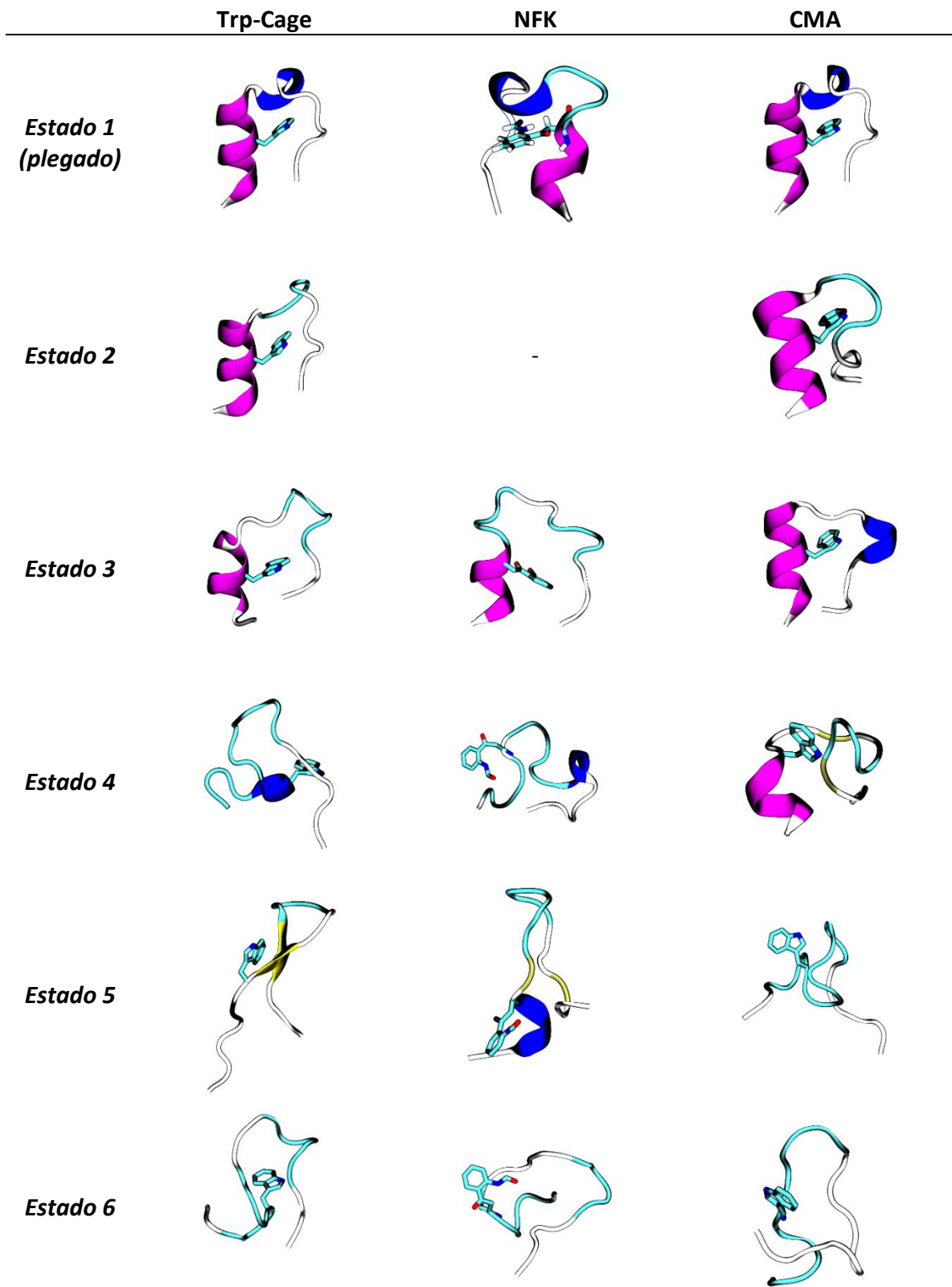


Figura 15. Estructuras representativas de los estados desplegados de Trp-Cage, CMA y NFK.

**Tabla 4.** Contenido en estructura secundaria para los estados de Trp-Cage, NFK y CMA.

	DSSP Hélice $\alpha$ (# residuos)			DSSP Hélice $3_{10}$ (# residuos)		
	Trp-Cage	NFK	CMA	Trp-Cage	NFK	CMA
<i>Estado 1</i>	6.3 $\pm$ 0.8	2.5 $\pm$ 1.8	6.2 $\pm$ 0.8	3.1 $\pm$ 0.5	1.6 $\pm$ 1.2	3.1 $\pm$ 0.9
<i>Estado 2</i>	3.4 $\pm$ 2.6	-	3.1 $\pm$ 2.6	0.3 $\pm$ 0.8	-	0.1 $\pm$ 0.5
<i>Estado 3</i>	3.2 $\pm$ 2.4	2.5 $\pm$ 2.1	3.4 $\pm$ 2.5	0.2 $\pm$ 0.6	0.1 $\pm$ 0.3	0.3 $\pm$ 0.7
<i>Estado 4</i>	0.2 $\pm$ 0.8	0	0.5 $\pm$ 1.3	1.3 $\pm$ 1.5	1.5 $\pm$ 1.5	0.6 $\pm$ 1.2
<i>Estado 5</i>	0.1 $\pm$ 0.6	0.5 $\pm$ 1.3	0.03 $\pm$ 0.4	0.4 $\pm$ 1.0	0.5 $\pm$ 0.5	0.6 $\pm$ 1.2
<i>Estado 6</i>	0.04 $\pm$ 0.4	0.2 $\pm$ 0.8	0.1 $\pm$ 0.6	0.2 $\pm$ 0.6	0.1 $\pm$ 0.4	0.2 $\pm$ 0.6

En general los resultados de asignación estructural para Trp-Cage y para CMA son similares, pero para NFK hay diferencias significativas en ambos elementos de estructura secundaria con respecto a Trp-Cage, reflejando así el efecto producido por la oxidación del triptófano.

En la Tabla 5 se muestran las superficies accesibles al disolvente del residuo 6 para las tres proteínas en los diferentes estados:

**Tabla 5.** Superficies accesibles al disolvente del residuo 6 de cada proteína en todos los estados.

	% SASA Residuo 6		
	Trp-Cage	NFK	CMA
<i>Estado 1</i>	10 $\pm$ 5	44 $\pm$ 18	9 $\pm$ 5
<i>Estado 2</i>	40 $\pm$ 23	-	40 $\pm$ 18
<i>Estado 3</i>	37 $\pm$ 20	48 $\pm$ 18	38 $\pm$ 18
<i>Estado 4</i>	55 $\pm$ 21	66 $\pm$ 20	58 $\pm$ 22
<i>Estado 5</i>	60 $\pm$ 24	55 $\pm$ 19	57 $\pm$ 20
<i>Estado 6</i>	51 $\pm$ 20	56 $\pm$ 24	58 $\pm$ 22

Para las proteínas Trp-Cage y CMA, los valores de % de SASA del Trp6 para los estados desplegados son hasta 5 veces mayores comparados con los del estado plegado (estado 1). Los estados desplegados en NFK también presentan porcentajes más altos de SASA que su estado 1, pero la diferencia no es tan grande como la mencionada para Trp-Cage y CMA. Esto se debe a que el núcleo hidrofóbico de NFK en el estado plegado ya se encuentra distorsionado por la modificación realizada en Trp6.

El estado 2 en Trp-Cage y CMA, y el estado 3 en las tres proteínas, son los estados que presentan los valores de % de SASA más bajos de los estados desplegados (Tabla 5). Esto es debido a que esos estados se encuentran ubicados a bajos valores de RMSD de la hélice  $\alpha$  en las FES de cada proteína. Al contrario, los estados 4, 5 y 6 presentan altos valores de RMSD de la hélice  $\alpha$ , lo que conlleva la exposición del triptófano al disolvente, pues este residuo se encuentra dentro de la hélice  $\alpha$ .

Esta información va de la mano con los resultados obtenidos de radio de giro para cada estado (Tabla A1), en donde se observa que a medida que se va perdiendo la estructura secundaria, la compacidad de la estructura también disminuye, y por lo tanto las estructuras pertenecientes a los estados desplegados se parecen menos a la estructura plegada de referencia.

Finalmente se determinaron las distancias de enlace de hidrógeno que establece NFK6 con su backbone, encontrando que la interacción se mantiene en todos los estados con valores de 0.3 a 0.4 nm. (Tabla A4). También se calcularon las distancias del puente salino para cada estado de las tres proteínas, obteniendo valores por encima de 1 nm en todos los estados desplegados, lo que indica que este tipo de interacción se pierde fuera del estado plegado para CMA y Trp-Cage.

**Tabla 6.** Distancias entre los residuos 16 y 9 de cada proteína en todos los estados.

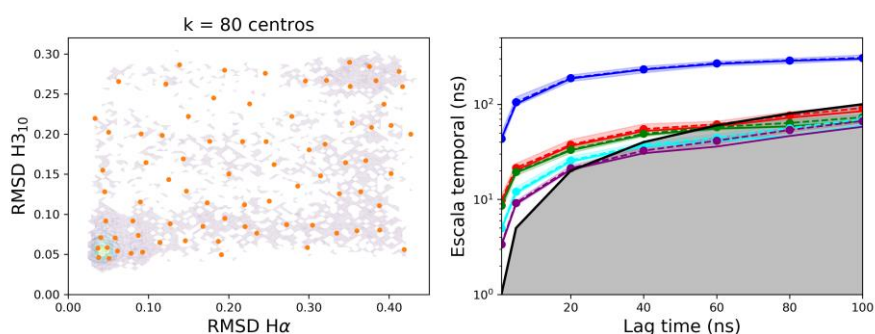
	Distancia (nm) Arg16/CMA16-Asp9		
	Trp-Cage	NFK	CMA
<i>Estado 1</i>	0.5 ± 0.2	1.0 ± 0.5	0.5 ± 0.2
<i>Estado 2</i>	0.9 ± 0.5	-	1.2 ± 0.5
<i>Estado 3</i>	1.1 ± 0.6	1.1 ± 0.6	1.4 ± 0.5
<i>Estado 4</i>	1.0 ± 0.5	1.1 ± 0.5	1.0 ± 0.4
<i>Estado 5</i>	1.0 ± 0.4	1.0 ± 0.5	1.2 ± 0.4
<i>Estado 6</i>	1.1 ± 0.6	1.3 ± 0.6	1.4 ± 0.6

## 4.5. Estudios Cinéticos

Tanto para Trp-Cage y las dos modificadas se utilizaron todas las trayectorias, las que empiezan con la proteína plegada y las que empiezan con la proteína desplegada, para construir un modelo de estados Markoviano con el que extraer información cinética sobre el plegamiento/desplegamiento.

### 4.5.1. Trp-Cage

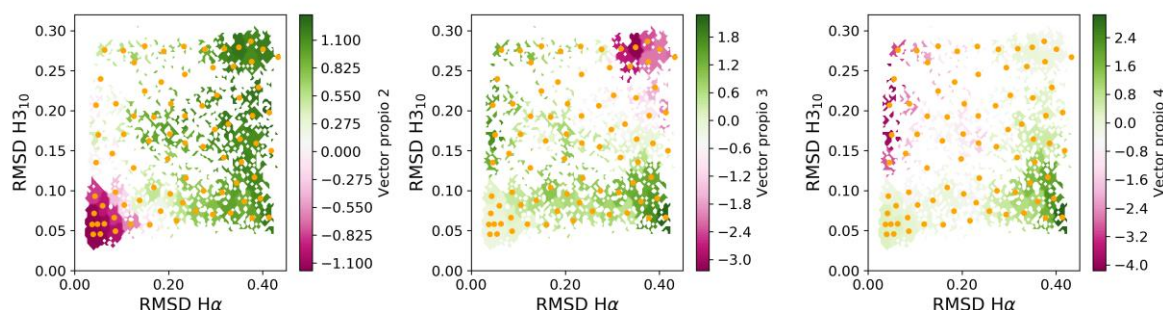
La discretización con 20 clústeres proporciona buena resolución temporal del proceso dominante (i.e. el más lento), pero es necesario incrementar el número de clústeres a 80 para poder resolver el segundo y tercer procesos (Figura A4). Para todas las discretizaciones, 40 ns de lag time es un buen compromiso entre la convergencia de la escala temporal del proceso más lento y la resolución temporal. Así, con una discretización de 80 clústeres y un lag time de 40 ns, la escala temporal del segundo y tercer proceso está un poco por encima del lag time (Figura 16).



**Figura 16.** Izquierda: Centros de los clústeres sobre la densidad de población en función de los RMSD de la hélice  $3_{10}$  y la hélice  $\alpha$ . Se ha representado la densidad de población con transparencia para una mejor visualización. Derecha: Escalas temporales para los 5 procesos más lentos (colores y en orden temporal ascendente: azul, rojo, verde, cyan, morado). Las líneas discontinuas muestran la escala temporal medida de las trayectorias. Las líneas continuas muestran la escala temporal predicha por el MSM al lag time correspondiente. La zona sombreada de color muestra el intervalo de un 95% de confianza. El lag time se muestra con la línea negra y la zona en gris.

Con la discretización de 80 clústeres y un lag time de 40 ns se pueden resolver 3 procesos. Siguiendo la enumeración de los estados termodinámicos, el segundo vector propio de la matriz de transiciones indica que el proceso más lento corresponde a una transición entre el estado 1 (plegado) y todos los estados desplegados (Figura 17). La escala temporal asociada a este vector propio, indica que el plegamiento/desplegamiento tiene requiere  $238 \pm 9$  ns (Tabla 7). El segundo proceso más lento corresponde a la transición entre los estados 5, 6 y todos los demás (Figura 17) y tiene lugar en una escala temporal de  $54 \pm 2$  ns. Como se observa en la FES de Trp-Cage (Figura 11), el estado 6 es la segunda conformación más

estable después del estado plegado. Consecuentemente, la transición hacia o desde este estado presenta la segunda cinética más lenta. El siguiente proceso corresponde a transiciones entre los estados 1, 4, 5 y 6 y los estados más energéticos 2 y 3 (Figura 17). Este proceso es un poco más rápido que el anterior, con una escala temporal de  $48 \pm 2$  ns. Los estados que más participan en este proceso son el 2, 3 y 4. Así esta transición implica intercambio entre hélice  $3_{10}$  (estado 4) y hélice alfa (estados 2 y 3).

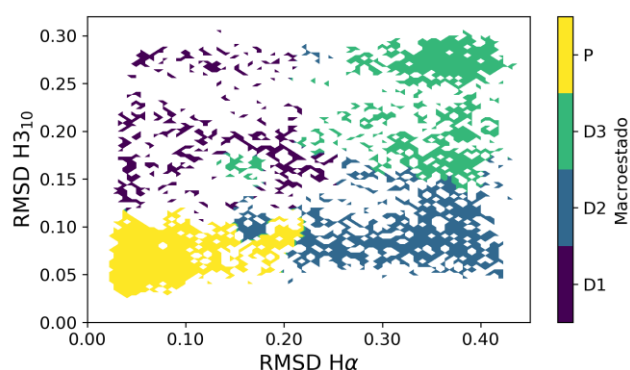


**Figura 17.** Vectores propios 2º a 4º de la matriz de transiciones del modelo Markoviano para la proteína Trp-Cage.

**Tabla 7.** Escalas temporales y desviación estándar de los 3 procesos más lentos de la dinámica de Trp-Cage.

	<b>Vector propio 2</b>	<b>Vector propio 3</b>	<b>Vector propio 4</b>
Escala temporal (ns)	$238 \pm 9$	$54 \pm 2$	$48 \pm 2$

Como para la discretización y lag time elegidos se pueden resolver 3 procesos (Figura 17), es razonable pensar que corresponden a transiciones entre 4 estados cinéticos. Estos 4 estados se obtienen de agrupar los 80 microestados cinéticos (un microestado cinético corresponde a un clúster de conformaciones). El proceso resultante es satisfactorio puesto que se obtienen 4 estados metaestables o macroestados bien separados. Es decir, cada microestado tiene una probabilidad muy alta de pertenecer únicamente a uno de los macroestados (Figura A5). Debe indicarse que un análisis de Chapman-Kolmogorov indicó que las dinámicas del modelo simple de 4 macroestados y el modelo con los 80 microestados no presentan diferencias significativas con un intervalo de confianza del 95% (Figura A6).



**Figura 18.** Macroestados cinéticos para Trp-Cage.

Los macroestados cinéticos resultantes tienen correspondencia con los mínimos de la superficie de energía libre. Así, el estado P corresponde al mínimo termodinámico 1 y a la conformación plegada (Figura 18). Los estados D1, D2 y D3 corresponden a conformaciones desplegadas. Las etiquetas D1, D2 y D3 se asignan en orden creciente de población. D1 incluye los mínimos termodinámicos 2 y 3. El estado D2 se corresponde con el mínimo 4, y el estado D3 con los estados 5 y 6. Esta agrupación indica que dentro de cada macroestado cinético las transiciones entre conformaciones son más rápidas que la resolución temporal elegida. En este caso los 40 ns de lag time.

Según este modelo de 4 estados, el estado plegado P acumula una población del 49% y la suma de los estados desplegados el 51% restante (Tabla 8). Este resultado es indicativo de que las simulaciones se han realizado aproximadamente a la temperatura de fusión de la proteína. Entre los estados desplegados, D2 y D3 tienen aproximadamente la misma población (19-20%) y la población menor se da en D1 (12%), de acuerdo con orden de energías observado en la FES de Trp-Cage (Figura 11).

A partir de las probabilidades de transición entre estados y los tiempos de transición observados en las trayectorias, es posible determinar el tiempo medio que tarda la proteína en escapar por primera vez de un estado para llegar a otro diferente. Es decir, el tiempo de transición entre macroestados conformacionales. Las transiciones desde los estados desplegados individuales hacia el estado plegado son más rápidas, con tiempos entre 311 y 406 ns, que al revés, con tiempos entre 617 y 765 ns (Tabla 8). Esto explica la mayor población del estado plegado, sin embargo, si se consideran todos los estados desplegados como un único macroestado, los tiempos de plegamiento y desplegamiento son estadísticamente equivalentes (Tabla 9), lo que vuelve a indicar que la temperatura de simulación corresponde a la de fusión de la proteína.

**Tabla 8.** Poblaciones de los macroestados cinéticos. Primeros tiempos medios de transición entre estados (ns).  
Transiciones desde (fila) hasta (columna).

Estado	Población		D1	D2	D3	P
D1	0.12	D1	-	358	437	311
D2	0.19	D2	444	-	358	360
D3	0.20	D3	431	271	-	406
P	0.49	P	647	617	765	-

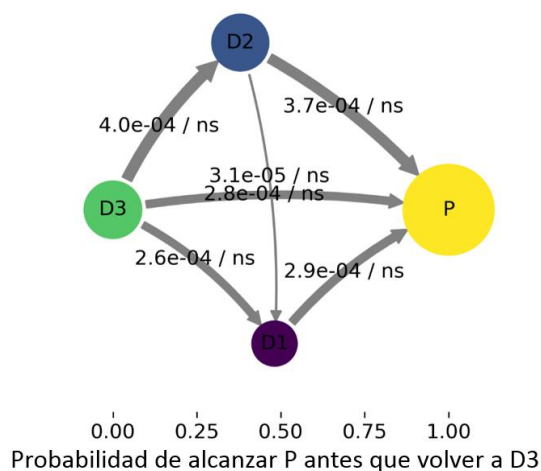
**Tabla 9.** Primeros tiempos medios de transición  $\pm$  desviación estándar e intervalos de confianza para un modelo de dos estados. Todas las magnitudes se dan en ns.

	D	P	Intervalo de confianza 95%
D	-	370 $\pm$ 15	346, 407
P	400 $\pm$ 19	-	360, 438

A partir de la población entre estados microscópicos y de la probabilidad de transición entre ellos se puede calcular el flujo entre dos estados. Desde el estado completamente desplegado (D3) hasta el plegado (P), la ruta dominante con un 40% del flujo total implica primero la formación de la hélice  $3_{10}$  (D3  $\rightarrow$  D2) o de una conformación tipo horquilla, y luego la formación de la hélice alfa (D2  $\rightarrow$  P, Tabla 10, Figura 19). La segunda y tercera rutas tienen un peso similar. La segunda implica la transición directa desde D3 hasta P, mientras que la tercera implica primero la formación de la hélice alfa y luego la formación de la hélice  $3_{10}$  (D3  $\rightarrow$  D1  $\rightarrow$  P, Tabla 10, Figura 19).

**Tabla 10.** Flujos de las rutas de plegamiento de D3 a P.

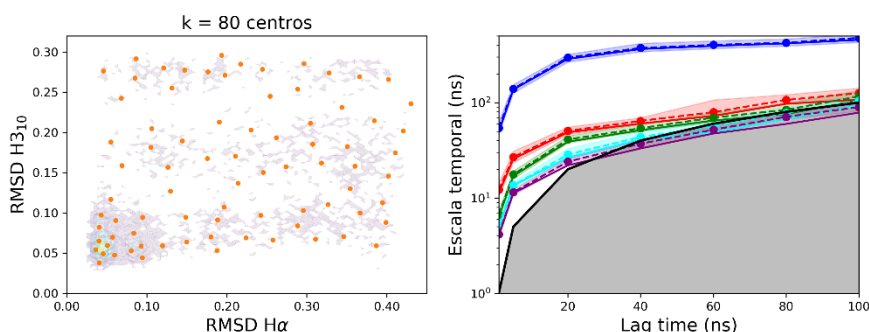
Ruta	Flujo (%)
D3 $\rightarrow$ D2 $\rightarrow$ P	39
D3 $\rightarrow$ P	29
D3 $\rightarrow$ D1 $\rightarrow$ P	28
D3 $\rightarrow$ D2 $\rightarrow$ D1 $\rightarrow$ P	3



**Figura 19.** Flujos de plegamiento para Trp-Cage medido como número de transiciones por ns. El eje horizontal indica la probabilidad de alcanzar el estado P antes que a D3 partiendo desde cada estado. El ancho de las flechas es proporcional al flujo. El área de cada estado es proporcional a la población del mismo.

#### 4.5.2. Trp-Cage CMA

De manera similar a Trp-Cage, la discretización con 80 clústeres permite resolver el primer, segundo y tercer proceso para CMA por encima de un lag time de 40 ns (Figura 20). Mientras que la escala temporal del primer proceso converge correctamente de forma asintótica, la convergencia de las escalas temporales del segundo y tercer proceso no se puede observar al estar las escalas muy próximas al lag time y perder resolución temporal.

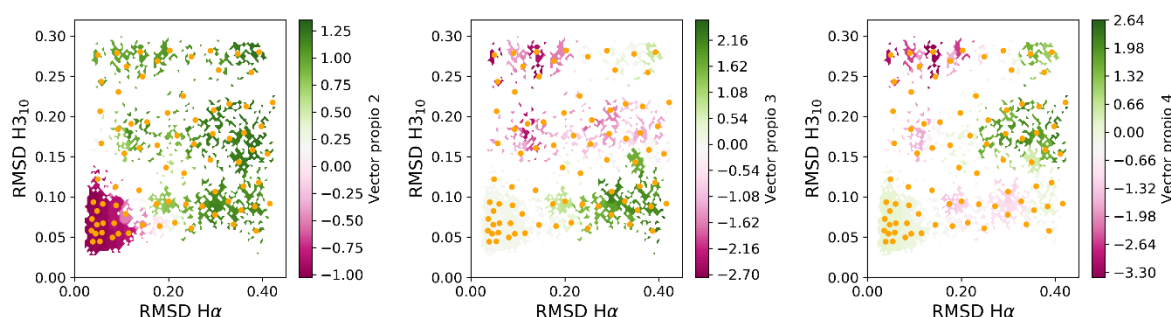


**Figura 20.** Izquierda: Centros de los clústeres de CMA sobre la densidad de población en función de los RMSD de la hélice  $3_{10}$  y la hélice  $\alpha$ . Se ha representado la densidad de población con transparencia para una mejor visualización. Derecha: Escalas temporales para los 5 procesos más lentos (colores y en orden temporal ascendente: azul, rojo, verde, cyan, morado). Las líneas discontinuas muestran la escala temporal medida de las trayectorias. Las líneas continuas muestran la escala temporal predicha por el MSM al lag time correspondiente. La zona sombreada de color muestra el intervalo de un 95% de confianza. El lag time se muestra con la línea negra y la zona en gris.

Según la Figura 21, el vector propio del proceso más lento corresponde a la transición entre el estado termodinámico 1 y todos los estados desplegados (2 a 6) considerados en la FES de CMA (Figura 11). Este proceso se da en una escala temporal de 370 ns (Tabla 11). Para CMA, a diferencia de Trp-Cage el segundo proceso más lento con una escala temporal de 64 ns, representa las transiciones entre los estados 1, 4 y 6 y los estados 2, 3, 5, aunque la mayor variación de población se da entre el estado 4 y los estados 2 y 3. Esta diferencia tiene una total concordancia con lo observado en las superficies de energía libre de ambas proteínas (Figura 11), en donde se evidencia que para CMA, el estado 6 aumenta considerablemente de energía, y no es la segunda conformación más estable. El estado 4 se convierte en la segunda conformación



más estable, por lo que está implicado en la segunda transición más lenta (Figura 21). El tercer proceso tiene lugar en una escala temporal de 55 ns. De manera similar a Trp-Cage, este proceso corresponde a las transiciones entre los estados 1, 4, 5 y 6 y los estados 2 y 3.

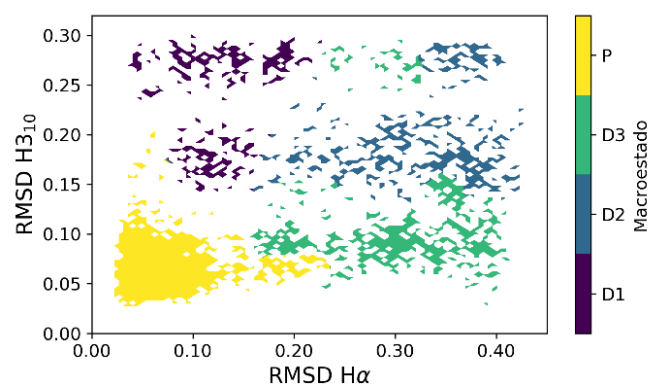


**Figura 21.** Vectores propios 2<sup>o</sup> a 4<sup>o</sup> de la matriz de transiciones del modelo Markoviano para la proteína Trp-Cage CMA.

**Tabla 11.** Escalas temporales y desviación estándar de los 3 procesos más lentos de la dinámica de Trp-Cage CMA.

	Vector propio 2	Vector propio 3	Vector propio 4
Escala temporal (ns)	370 ± 19	64 ± 3	55 ± 3

Al poder resolverse 3 procesos por encima del lag time, se puede pensar que corresponden a transiciones entre 4 estados cinéticos. La Figura A8 corrobora que los 4 macroestados se encuentran bien definidos. También se llevó a cabo el test de Chapman-Kolmogorov, el cual validó el comportamiento Markoviano del sistema (Figura A9). Para CMA, el estado plegado P corresponde al estado 1 termodinámico (Figura 22). El estado D1 corresponde a los mínimos termodinámicos 2 y 3. El estado D2 se relaciona con los estados 5 y 6, y el estado D3 con el mínimo 4.



**Figura 22.** Macroestados cinéticos para Trp-Cage CMA.

La población del macroestado cinético P es del 56% (Tabla 12), valor que se encuentra un poco por encima del obtenido en Trp-Cage (49%). Esto indica que la glicación del residuo Arg16 produce que el estado plegado P sea más estable. Los estados desplegados representan en total una población del 44%. A diferencia de Trp-Cage, en donde el estado completamente desplegado era el más poblado de los desplegados (i.e. D3, Tabla 7, Figura 18), en CMA este estado es el segundo desplegado más poblado (D2). Esto era de esperar según valores relativos de energía libre de los respectivos estados termodinámicos (Figura 11).

De nuevo se observa que las transiciones desde los estados desplegados hacia el estado plegado son más rápidas, con tiempos entre 502 y 554 ns, que al revés, con tiempos entre 981 y 1414 ns (Tabla 12). Esta diferencia se constata al agrupar todos los estados desplegados, resultando un tiempo de desplegamiento significativamente mayor al de plegamiento (Tabla 13). Este resultado concuerda con una estabilización del

estado plegado (Figura 11). Por otra parte, como tanto los tiempos de plegamiento como los de desplegamiento son mayores para CMA, se deduce que la barrera de energía aumenta en los dos sentidos. En la FES (Figura 11) se aprecia que para CMA la barrera de energía es mayor en el sentido de desplegamiento, pero no se observan grandes diferencias entre CMA y Trp-Cage para las barreras de plegamiento. Sí se aprecia que en la FES de CMA hay más zonas de alta energía inexploradas (zonas en blanco) y que los canales de reacción entre los estados son más estrechos, indicativo de que el aumento de los tiempos de plegamiento puede ser debido a causas entrópicas. Sin embargo, no deben descartarse errores en los valores de la FES por muestreo limitado de las zonas de alta energía.

**Tabla 12.** Poblaciones de los macroestados cinéticos y primeros tiempos medios de transición entre estados. Transiciones desde (fila) hasta (columna).

Estado	Población		D1	D2	D3	P
D1	0.08	D1	-	427	390	502
D2	0.17	D2	753	-	325	554
D3	0.19	D3	831	433	-	521
P	0.56	P	1414	1138	981	-

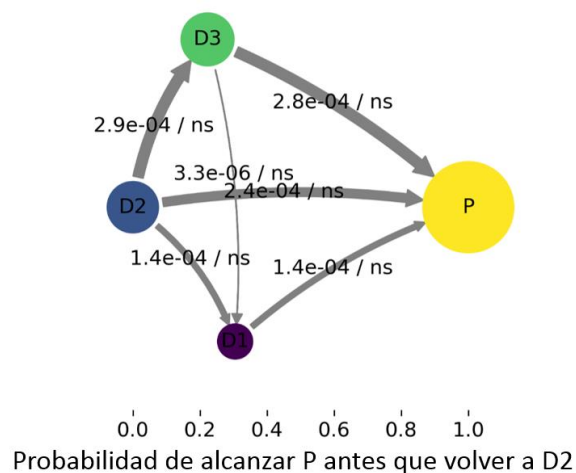
**Tabla 13.** Primeros tiempos medios de transición  $\pm$  desviación estándar e intervalos de confianza para un modelo de dos estados. Todas las magnitudes se dan en ns.

	D	P	Intervalo de confianza 95%
D	-	549 $\pm$ 42	465, 629
P	789 $\pm$ 60	-	691, 930

El análisis de flujo desde el estado completamente desplegado (D2 en CMA) hasta el plegado (P) indica que los mecanismos de plegamiento son esencialmente iguales, y con porcentajes similares a los de Trp-Cage (Tabla 14, Figura 22). La ruta dominante con un 43% implica primero la formación de la hélice 3<sub>10</sub> o de una conformación tipo horquilla, y luego la formación de la hélice alfa (Tabla 14, Figura 22). La transición directa en donde la proteína desplegada colapsa hacia el estado plegado tiene una contribución del 36%, y la formación de la hélice alfa (D2  $\rightarrow$  D1) y después de la hélice 3<sub>10</sub> (D1  $\rightarrow$  P) un 21% (Tabla 14, Figura 22).

**Tabla 14.** Flujos de las rutas de plegamiento de D2 a P.

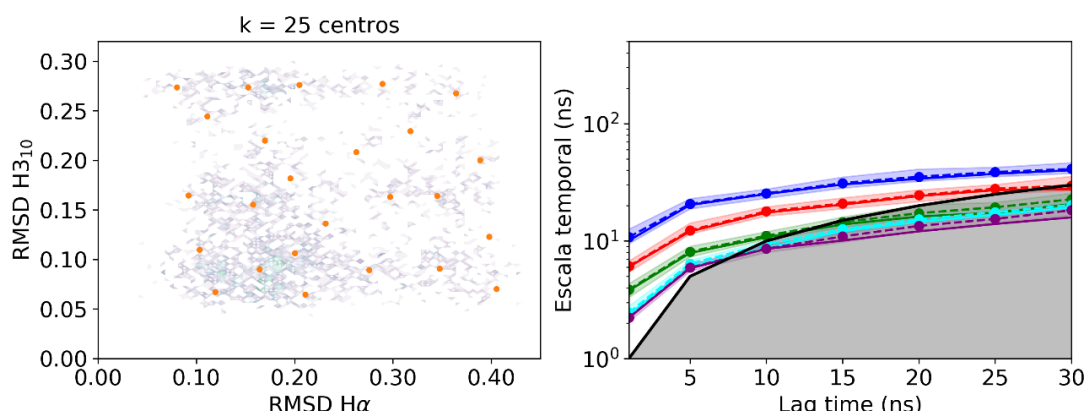
Ruta	Flujo (%)
D2 $\rightarrow$ D3 $\rightarrow$ P	43
D2 $\rightarrow$ P	36
D2 $\rightarrow$ D1 $\rightarrow$ P	21



**Figura 23.** Flujos de plegamiento para Trp-Cage CMA medido como número de transiciones por ns. El eje horizontal indica la probabilidad de alcanzar el estado P antes que a D2 partiendo desde cada estado. El ancho de las flechas es proporcional al flujo. El área de cada estado es proporcional a la población del mismo.

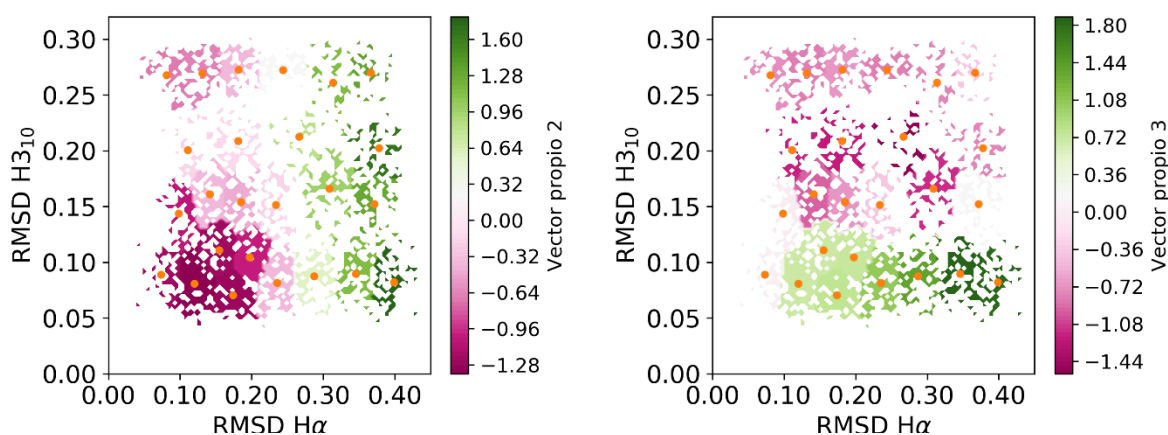
### 4.5.3. Trp-Cage NFK

Para NFK, la discretización con 25 clústeres permite resolver el segundo y tercer proceso por encima de un lag time de 20 ns, un tiempo menor que en CMA y en Trp-Cage porque en NFK las transiciones se dan mucho más rápido como se mostrara más adelante.



**Figura 24.** Izquierda: Centros de los clústeres de NFK sobre la densidad de población en función de los RMSD de la hélice  $H_{310}$  y la hélice  $H_{\alpha}$ . Se ha representado la densidad de población con transparencia para una mejor visualización. Derecha: Escalas temporales para los 5 procesos más lentos (colores y en orden temporal ascendente: azul, rojo, verde, cian, morado). Las líneas discontinuas muestran la escala temporal medida de las trayectorias. Las líneas continuas muestran la escala temporal predicha por el MSM al lag time correspondiente. La zona sombreada de color muestra el intervalo de un 95% de confianza. El lag time se muestra con la línea negra y la zona en gris.

El vector propio del proceso más lento corresponde a la transición entre los estados 1 y 3, y los estados 4, 5 y 6 (Figura 25). Esto está de acuerdo con la FES de NFK (Figura 11), que muestra una zona de baja energía libre por debajo de RMSD de hélice alfa 0.20 nm, y una zona de más alta energía para valores superiores a 0.20 nm. Esta transición se da en una escala temporal de 39 ns (Tabla 15), 1 orden de magnitud menor que la escala temporal asociada al proceso más lento de CMA (370 ns) y Trp-Cage (238 ns). De manera similar que en CMA, el segundo proceso más lento corresponde a las transiciones entre 1 y 4 y los demás estados. Este proceso tiene lugar en una escala temporal de 27 ns. De nuevo este bajo valor indica que las transiciones se dan más rápido en NFK que en las transiciones equivalentes de Trp-Cage y CMA.

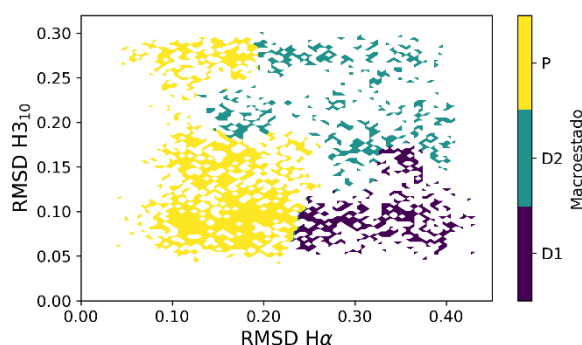


**Figura 25.** Vectores propios 2º y 3º de la matriz de transiciones del modelo Markoviano para la proteína Trp-Cage NFK.

**Tabla 15.** Escalas temporales de los 3 procesos más lentos de la dinámica de Trp-Cage.

	<b>Vector propio 2</b>	<b>Vector propio 3</b>
Escala temporal (ns)	$39 \pm 1.8$	$28 \pm 1.4$

Dados dos procesos, se considera agrupar los microestados en 3 macroestados. En las distribuciones metaestables se observa cierto solapamiento entre uno de los estados desplegados y el estado plegado (Figura A11). Esto es debido a que las barreras de energía libre entre estados son bajas para la temperatura de simulación (Figura 11). Sin embargo, el modelo de 3 macroestados cinéticos superó satisfactoriamente el test de Markovianidad de Chapman-Kolmogorov (Figura A12).



**Figura 26.** Macroestados cinéticos para Trp-Cage NFK.

El macroestado P (plegado) corresponde a los estados 1 y 3 termodinámicos. En la FES de NFK (Figura 11) se aprecia que la barrera de energía entre 1 y 3 es de 5 kJ/mol mientras que la barrera entre 1 y 4 es de 4 kJ/mol. Así, podría parecer que los estados 1 y 4 deberían formar parte del mismo macroestado en lugar de 1 y 3. Pero la diferencia de energía entre las dos barreras está dentro del error estadístico del muestreo. La población del estado P es del 52%, pero no se debe interpretar como una estabilización del estado plegado ya que las conformaciones que lo forman (estados 1 y 3) son muy diversas y no contienen una cantidad importante de plegamiento nativo. Por eso al estado P se le puede denominar como “mal plegado” del inglés *misfolded*. El estado D1 corresponde al estado 4, y el estado D2 a los estados 5 y 6 (Figura 26). Los dos estados desplegados representan en total una población del 48%, donde el mayoritario es el estado D2 con un 28%, seguido del estado D1 con un 20% (Tabla 16).

Para NFK, las transiciones desde los estados desplegados hacia el estado mal plegado también son más rápidas, con tiempos entre 70 y 76 ns, que al revés, con tiempos entre 117 y 176 ns. Al agrupar los dos

estados desplegados en uno solo, se aprecia que los tiempos de paso entre el estado mal plegado (P) y los desplegados (D) son estadísticamente iguales (Tabla 17). En cualquiera de los dos casos, la modificación del triptófano en NFK incrementa significativamente la dinámica de la proteína, de acuerdo con las bajas barreras entre estados observadas en la FES correspondiente (Figura 11).

**Tabla 16.** Poblaciones de los macroestados cinéticos y primeros tiempos medios de transición entre estados. Transiciones desde (fila) hasta (columna).

Estado	Población		D1	D2	P
D1	0.20	D1	-	105	76
D2	0.28	D2	156	-	70
P	0.52	P	176	117	-

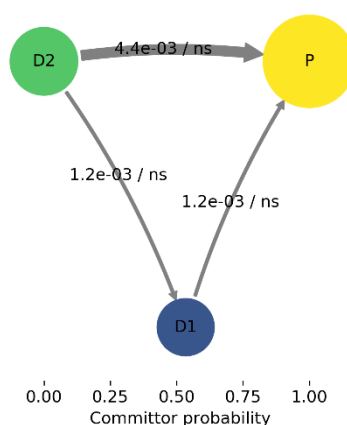
**Tabla 17.** Primeros tiempos medios de transición  $\pm$  desviación estándar e intervalos de confianza para un modelo de dos estados. Todas las magnitudes se dan en ns.

	D	P	Intervalo de confianza 95%
D	-	72 $\pm$ 4	67, 79
P	77 $\pm$ 4	-	70, 84

El análisis de flujo de NFK desde el estado completamente desplegado (D2) hasta el “mal plegado” (P) indica que la ruta predominante (79%, Tabla 18) es el paso directo, lo que implica la formación del segmento de hélice  $\alpha$  que interacciona con la cadena NFK. El segundo mecanismo, con un peso del 21%, implica primero la formación de la hélice  $3_{10}$  o de una conformación tipo horquilla (D2  $\rightarrow$  D1), y luego la formación del segmento de hélice alfa observado para NFK (D1  $\rightarrow$  P, Tabla 18, Figura 27).

**Tabla 18.** Flujos de las rutas de plegamiento de D2 a P.

Ruta	Flujo (%)
D2 $\rightarrow$ P	79
D2 $\rightarrow$ D1 $\rightarrow$ P	21



**Figura 27.** Flujos de plegamiento para Trp-Cage CMA medido como número de transiciones por ns. El eje horizontal indica la probabilidad de alcanzar el estado P antes que a D2 partiendo desde cada estado. El ancho de las flechas es proporcional al flujo. El área de cada estado es proporcional a la población del mismo.

## 4.6. Comparación bibliográfica

A continuación, se presenta una tabla comparativa de algunos resultados obtenidos por otros investigadores y los datos obtenidos en este trabajo para la proteína Trp-Cage:

**Tabla 19.** Parámetros asociados a Trp-Cage obtenidos de diferentes fuentes.

<b>Parámetro</b>	<b>Resultado</b>	<b>Temperatura (K)</b>	<b>Técnica</b>	<b>Referencia</b>
<i>Tiempos de Plegamiento (<math>\mu</math>s)</i>	0.4	352	DM - campo de fuerza a99SB*-ILDN	<b>Este estudio</b>
	3.6	300	DM - campo de fuerza PARAM19	Andryushchenko & Chekmarev, 2015
	5.5	290	DM - campo de fuerza CHARMM22* y Modelo Markov	Deng <i>et al.</i> , 2013
	4	300	Espectroscopia de fluorescencia	Qiu <i>et al.</i> , 2002
<i>Intermedios</i>	Sí	352	DM- campo de fuerza a99SB*-ILDN	<b>Este estudio</b>
	Sí	300	DM con intercambio de réplicas- campo de fuerza OPLSAA	Zhou, 2013
	Sí	322	Espectroscopia IR	Meuzelaar <i>et al.</i> , 2013
	Sí	300	Espectroscopía de correlación de fluorescencia	Neuweiler <i>et al.</i> , 2005
	Sí	277-343	Espectroscopia Raman de Resonancia UV	Ahmed <i>et al.</i> , 2005
	No	300	Espectroscopia de salto	Qiu <i>et al.</i> , 2002
<i>RMSD estructura de energía libre más baja (Å)</i>	1	352	DM - campo de fuerza a99SB*-ILDN	<b>Este estudio</b>
	1.5	300	DM - campos de fuerza aff03 y aff99SB	Kannan & Zacharias 2014
	1	300	DM con intercambio de réplicas- campo de fuerza RSFF2	Zhou, 2013
	1.5	300	DM con intercambio de réplicas- campo de fuerza OPLSAA	Zheng <i>et al.</i> , 2011

Es muy importante tener en cuenta que otras investigaciones que se han realizado de Trp-Cage no se han hecho bajo las mismas condiciones que las empleadas en este estudio. Por lo tanto, los datos aquí presentados tienen como propósito mostrar un panorama más amplio en cuanto a tendencias y comportamientos.

Considerando los tiempos de plegamiento que se han obtenido computacionalmente, se encuentran diferencias significativas probablemente asociadas a las diferentes temperaturas y campos de fuerza empleados. Además, en los estudios que emplearon modelos de disolvente implícito, es habitual observar cinéticas más rápidas debido a la ausencia de viscosidad del solvente (Chowdhury *et al.*, 2003).

Experimentalmente Qiu y sus colaboradores (Qiu *et al.*, 2002), determinaron que el tiempo de plegamiento para Trp-cage es de 4  $\mu$ s a 300 K. Posteriormente, otros investigadores emplearon un modelo de estado de Markov construido a partir de una simulación de DM con el campo de fuerza CHARMM22\*, determinando un tiempo de plegamiento de 5.5  $\mu$ s a 290 K (Deng *et al.*, 2013). Adicionalmente, empleando la técnica de DM evaluaron una temperatura por debajo y otra por encima de la temperatura de fusión, encontrando que a la menor temperatura había mayor porcentaje de población plegada, lo que está en concordancia con lo observado experimentalmente (Qiu *et al.*, 2002). Andryushchenko & Chekmarev determinaron un tiempo de plegamiento de 3.6  $\mu$ s a 330 K con el campo de fuerza PARAM19 (Andryushchenko & Chekmarev, 2015).

El tiempo de plegamiento obtenido en nuestro estudio es de 0.4  $\mu$ s a 352 K. Esta temperatura de trabajo está elegida para reproducir la temperatura de fusión de la proteína (i.e. 315 K) con el campo de fuerza empleado. Por lo que corresponde a una temperatura efectiva mayor que las mencionadas en las anteriores investigaciones. Así, el menor valor de tiempo de plegamiento obtenido está en buena concordancia con el valor experimental a la temperatura de fusión (i.e. 2.0-2.5  $\mu$ s, Qiu *et al.*, 2002) considerando las limitaciones de los campos de fuerza y los errores estadísticos en el muestreo.

Diversos estudios experimentales (Ahmed *et al.*, 2005; Neuweiler *et al.*, 2005) y computacionales (Halabis *et al.*, 2012; Meuzelaar *et al.*, 2013; Zhou, 2013) han comprobado que la cinética de plegamiento de Trp-Cage no se puede definir como un modelo simple de dos estados. Por ejemplo, Meuzelaar y colaboradores (Meuzelaar *et al.*, 2013) observaron una cinética bimodal con tiempos de relajación 100 ns y 770 ns a 322 K, lo que sugirió que el plegamiento implica un estado intermedio. Estos tiempos de relajación son similares a los tiempos de plegamiento obtenidos en este trabajo (Tabla 8). Otros experimentos también sugirieron la existencia de un intermedio muy similar a la estructura nativa con una hélice  $\alpha$  completamente formada (Chowdhury *et al.*, 2004). Esta conclusión es nuevamente comprobada por nuestros estudios, los cuales respaldan la idea de que en el proceso de plegamiento/desplegamiento existe al menos un intermedio (Tabla 8, Figura 19).

Los estudios que emplearon espectroscopia Raman (Ahmed *et al.*, 2005), destacaron la presencia de estructuras  $\alpha$ -helicoidales incluso en las estructuras de Trp-cage desplegadas, lo que es consistente con los resultados de simulación observados en este trabajo, en donde se ha proporcionado evidencia de la existencia de la hélice  $\alpha$  en diferentes estados intermedios (Tabla 4, Figura 15).

Otro tema ampliamente discutido en la literatura es el asociado a las posibles vías de plegamiento. Las vías más características son: 1) el colapso del núcleo hidrofóbico precede a la formación de la hélice  $\alpha$ ; y 2) la formación del  $\alpha$ -hélice precede al colapso del núcleo hidrofóbico. Existen estudios en donde mencionan la existencia de contactos hidrofobos incluso en el estado desnaturalizado con urea (Kannan & Zacharias 2014), indicando que estos contactos pueden darse de forma previa a la formación de los elementos de estructura secundaria. Sin embargo, Juraszek y colaboradores (Juraszek & Bolhuis, 2006) mediante DM con un campo de fuerza OPLSAA FF a 300 K, encontraron que el 20 % de los caminos primero forman la hélice, y el 80 % restante primero forman los contactos terciarios. Eso indica que Trp-Cage se podría plegar tanto por un mecanismo de difusión-colisión como por nucleación-condensación. Posteriormente, en otro estudio computacional con DM y un campo de fuerza PARAM19 a 285, 295 y 325 K (Andryushchenko & Chekmarev, 2017), se encontró que a medida que se aumenta la temperatura es más probable que la trayectoria de plegado comience con la formación de los elementos de estructura secundaria. Nuestros resultados de simulación concuerdan cualitativamente con este estudio considerando que nuestra temperatura de simulación es más alta y se emplea un campo de fuerza diferente (Figura 19).

En cuanto a las superficies de energía libre observadas en otras investigaciones, es importante tener en cuenta que los mínimos energéticos no siempre corresponden a una única conformación. Dependiendo de los parámetros de orden elegidos para la representación, puede ocurrir una superposición entre estados y ocultar un estado dentro de otro. Además, dependiendo de la temperatura, pueden encontrarse diversas conformaciones en un único mínimo energético si no hay barreras de energía entre ellas. Por ejemplo, en un estudio termodinámico de Trp-Cage a diferentes temperaturas, se observaron diferencias relevantes entre las superficies de energía libre (Zhou, 2003). En otro estudio con DM y un campo de fuerza a99SB (Meuzelaar *et al.*, 2013) también se observó que, a temperaturas bajas hay menos fluctuaciones y la FES es estrecha. Caso contrario a cuando se evalúa una temperatura por encima de la temperatura de fusión, donde un gran número de vías de plegamiento diversas con flujos comparables inundan el paisaje energético. En nuestro

estudio se empleó una temperatura aproximada al punto de fusión que, en conjunto con las variables colectivas elegidas, permitieron observar diferentes estados termodinámicos (Figura 11).

Las estructuras proteicas que corresponden a los mínimos energéticos de una investigación realizada con el campo de fuerza RSFF2 a 300 K (Zhou *et al.*, 2015), tienen mucha similitud con los expuestos en este trabajo. Ellos emplearon variables colectivas relacionadas con las nuestras que les permitieron elucidar diversos estados. Entre ellos, el más poblado corresponde al estado plegado. También describieron otros estados con características similares a los obtenidos en este estudio, como conformaciones con la hélice alfa y hélice 3<sub>10</sub> parcialmente formadas (Figura 15).

En esa misma investigación, se estudiaron cinco variantes de Trp-Cage. La mutación Lys8Ala en un residuo de la hélice alfa causa una estabilización del estado plegado y aumenta su barrera de energía libre (Zhou *et al.*, 2015). Resultado similar al obtenido para CMA en este trabajo, en donde la modificación del residuo Arg16 trajo consigo una estabilización termodinámica (Figura 11) y cinética (Tabla 8, Tabla 11) de la conformación plegada.

Por el contrario, la mutación de Tyr3 (Tyr3Ala), que interacciona directamente con Trp6, desestabiliza la hélice  $\alpha$ , a pesar de la mayor propensión a la  $\alpha$ -helicidad de Ala (Zhou, 2015). Esto es debido a la eliminación de una de las interacciones clave en la formación de la caja hidrofóbica. Este resultado se relaciona con lo observado en NFK, en donde la modificación de Trp6 también produce una desestabilización termodinámica y cinética, producto de las distorsiones de las interacciones hidrofóbicas que son claves para el buen empaquetamiento de Trp-Cage.

#### 4.7. Implicaciones Biológicas

La glicosilación y otras modificaciones postraduccionales pueden cambiar la función y la estructura de las biomoléculas. Diversos estudios han demostrado que el efecto de las modificaciones en proteínas depende de donde se encuentre el residuo modificado y de las interacciones en dicha zona (Previero *et al.*, 1967; Zhou *et al.*, 2015; Ehrenshaft *et al.*, 2017). Una de las interacciones de mayor influencia en la estabilidad de la estructura de las proteínas son las hidrofóbicas. En la mayoría de las proteínas, los aminoácidos de cadena lateral no polar se ubican en el interior de la proteína evitando las interacciones con moléculas de agua e interaccionando con las cadenas laterales de otros aminoácidos no polares; este tipo de fuerzas tienen un efecto directo en el plegamiento correcto de la proteína. Por tales razones, al presentarse dichas asociaciones en el corazón hidrofóbico de nuestra proteína en estudio, los resultados obtenidos para Trp-Cage se pueden escalar de hacia otras proteínas de mayor tamaño en donde las interacciones hidrofóbicas tengan un peso importante.

Por ejemplo, si se pretende estudiar la oxidación de un residuo de triptófano que se encuentre en un núcleo hidrofóbico de una proteína de mayor tamaño, es probable que se observen resultados similares a los evidenciados en este estudio. Así como se mostró anteriormente, también es posible que dentro de una sola proteína, la alteración oxidativa de múltiples residuos de aminoácidos pueda tener resultados funcionales diferentes para la proteína.

Ehrenshaft y *col.* (Ehrenshaft *et al.*, 2013), investigaron el efecto de la fotosensibilización de la  $\alpha$ -cristalina, una de las principales proteínas del cristalino que también actúa como una chaperona para garantizar su propia integridad. Dicha fotosensibilización produjo una acumulación significativa de NFK. Observaron que cualquier modificación de la  $\alpha$ -cristalina que afecte su función de chaperona crea un entorno en el que puede producirse un mayor daño proteico. En relación con nuestros resultados, la modificación de Trp a NFK en Trp-Cage destruye el núcleo hidrofóbico y causa un mal plegamiento. Con base a esto, podemos plantear



la hipótesis de que la oxidación de Trp a NFK en  $\alpha$ -cristalina también afecte a su plegamiento, a su función como chaperona y, como consecuencia, se promueve la formación de cataratas.

Mediante DM y empleando un campo de fuerza CHARMM19 a 300 K se investigaron los efectos de la formación de CMA en las proximidades de los sitios de escisión de colagenasa (Stultz & Edelman, 2003). Se sugiere que el colágeno puede existir en un estado conformacional nativo y en otro estado vulnerable no nativo. En la conformación vulnerable, las regiones parcialmente desplegadas cerca de los sitios de escisión permiten que las colagenasas degraden eficientemente el colágeno. Las simulaciones mostraron que la cadena lateral de CMA establece un enlace de hidrógeno mediado por agua con el backbone de la proteína en el estado nativo y dos enlaces de hidrógeno en el estado vulnerable. Como consecuencia, el estado vulnerable pasa a ser el más estable (Stultz & Edelman, 2003). En este trabajo también se observa que CMA establece enlaces de hidrógeno con el grupo amino N-terminal y con los grupos amida de la región N-terminal. Sin embargo, en Trp-Cage, estas nuevas interacciones no desestabilizan el estado nativo.

Un estudio reciente de nuestro grupo de investigación muestra resultados experimentales y computacionales coherentes con los observados en este trabajo para CMA y Trp-Cage. A pH ácido, el péptido Ala(Glu)<sub>6</sub>Lys(Glu)<sub>6</sub>Ala muestra estructura de hélice  $\alpha$ . La formación de carboximetilisina (AGE análogo a la carboximetilarginina, pero formado sobre lisina) estabiliza la estructura de hélice  $\alpha$  porque se establecen enlaces de hidrógeno adicionales entre los grupos carboxílicos protonados de los residuos Glu y de la carboximetilisina (Mariño *et al.*, 2019).

Se observó que la formación de AGEs en la lisozima de clara de huevo de gallina (HEWL) favorece la agregación proteica sin causar desplegamiento (Mariño *et al.*, 2015). El motivo es que la glicación ocurre en las lisinas de la superficie de la proteína, disminuyendo la carga eléctrica y aumentando la hidrofobicidad, lo que favorece la asociación entre proteínas glicadas. Los resultados de este trabajo para CMA muestran un efecto análogo para la glicación de Arg16 en la superficie de la proteína: reducción de la carga eléctrica, aumento de la hidrofobicidad y mantenimiento del estado nativo. Por lo que incluso los AGEs que no causan modificaciones estructurales importantes pueden tener efectos negativos si se consideran las interacciones de la proteína glicada con otras biomoléculas del entorno fisiológico.

Estos resultados indican que la modificación de las interacciones por formación de los AGEs es similar en diferentes proteínas, pero la modificación de la estructura y la función depende del entorno químico particular de cada proteína.

## 5. CONCLUSIONES

1. Múltiples simulaciones de dinámica molecular de Trp-Cage y sus variantes modificadas NFK y CMA, empezando desde las conformaciones plegadas y completamente desplegadas, permiten caracterizar termodinámica y cinéticamente el plegamiento y desplegamiento de las mismas.
2. Los RMSD de la hélice  $3_{10}$  y de la hélice  $\alpha$  son la mejor combinación de VCs, ya que permitieron obtener una descripción detallada de las superficies de energía libre de Trp-Cage y sus modificaciones.
3. Los valores de RMSD bajos del estado plegado de la proteína modificada en Arg16 por CMA respecto a Trp-Cage nativa, el contenido de estructura secundaria y la superficie accesible al disolvente de Trp6 y de la proteína completa indican la modificación no impide el plegamiento correcto. La interacción del carboxilato de CMA con los grupos amina y amida de la región N-terminal es responsable de esta estabilización.
4. La modificación de Trp6 en NFK incrementa los valores de RMSD de la conformación plegada respecto a Trp-Cage nativa, disminuye el contenido en estructura secundaria de la hélice  $\alpha$  y aumenta la superficie accesible al disolvente de NFK6. Todo esto indica un mal plegamiento de la proteína modificada. La interacción entre un oxígeno carbonílico de la cadena lateral y la amida del backbone de NFK6 estabiliza los estados mal plegados.
5. Tomando como referencia Trp-Cage no modificada, la formación de NFK6 desestabiliza el estado plegado respecto al desplegado. También reduce las barreras de energía entre los mínimos energéticos observados. Por el contrario, la formación de CMA16 estabiliza el estado plegado respecto a los desplegados y aumenta las barreras de energía entre ellos.
6. Los modelos de Markov apoyan las tendencias observadas en las superficies de energía libre. La formación de CMA aumenta la población del estado plegado y aumenta los tiempos de plegamiento y desplegamiento. La formación de NFK aumenta la población de un estado plegado diferente al nativo (mal plegado) y disminuye en un orden de magnitud los tiempos de plegamiento y desplegamiento.
7. Estos resultados demuestran que diferentes variaciones pueden tener efectos muy diversos. Éstos dependen del residuo modificado y del entorno proteico y las interacciones formadas en dicha zona. La modificación del núcleo hidrofóbico de la proteína impide el plegamiento nativo en favor de estados mal plegados. Modificación de residuos de la superficie proteica que disminuyen la carga eléctrica y aumentan la hidrofobicidad, estabilizan estados plegados.

## 6. BIBLIOGRAFIA

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2, 19–25.
- Ahmed, Z., Beta, I. A., Mikhonin, A. V., & Asher, S. A. (2005). UV-Resonance Raman Thermal Unfolding Study of Trp-Cage Shows That It Is Not a Simple Two-State Miniprotein. *Journal of the American Chemical Society*, 127(31), 10943–10950.
- Andryushchenko, V. A., & Chekmarev, S. F. (2015). A hydrodynamic view of the first-passage folding of Trp-cage miniprotein. *European Biophysics Journal*, 45(3), 229–243.
- Andryushchenko, V. A., & Chekmarev, S. F. (2017). Temperature evolution of Trp-cage folding pathways: An analysis by dividing the probability flux field into stream tubes. *Journal of Biological Physics*, 43(4), 565–583.
- Anfinsen, C. B., Haber, E., Sela, M., & White Jr, F. H. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, 47(9), 1309
- Banach, M., Konieczny, L., & Roterman, I. (2019). Secondary and Supersecondary Structure of Proteins in Light of the Structure of Hydrophobic Cores. *Protein Supersecondary Structures*, 347–378
- Banerjee, S. (2018). Glyoxal-induced modification enhances stability of hemoglobin and lowers iron-mediated oxidation reactions of the heme protein: An in vitro study. *International Journal of Biological Macromolecules*, 107, 494–501.
- Batthyány, C., Santos, C. X., Botti, H., Cerveñansky, C., Radi, R., Augusto, O., & Rubbo, H. (2000). Direct evidence for apo B-100-mediated copper reduction: studies with purified apo B-100 and detection of tryptophanyl radicals. *Archives of biochemistry and biophysics*, 384(2), 335-340.
- Baynes, J., Watkins, N., Fisher, C., Hull, C., Patrick, J., Ahmed, M., Dunn, J., & Thorpe, S. (1989). The Amadori product on protein: structure and reactions. *Prog Clin Biol Res* 304: 43–67.
- Berendsen, H.J., Spoel, D.V., & Drunen, R.V. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.*, 1995, 91, 43–56.
- Berg, J., Tymoczko J., & Stryer, L. (2002). *Estructura y función de la proteína*. Estados Unidos, Nueva York, WH Freeman.
- Berhanu, W. M., & Masunov, A. E. (2014). Chapter 6 - The Atomic Level Interaction of Polyphenols with the A $\beta$  Oligomer Aggregate, a Molecular Dynamic Guidance for Rational Drug Design. *Polyphenols in Human Health and Disease*, 59–70.
- Berlett, B. S., & Stadtman, E. R. (1997). Protein oxidation in aging, disease, and oxidative stress. *Journal of Biological Chemistry*, 272(33), 20313-20316.
- Best, R. B., & Hummer, G. (2009). Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *The Journal of Physical Chemistry B*, 113(26), 9004–9015.

- Borghain, G., & Paul, S. (2018). Atomistic level understanding of the stabilization of protein Trp cage in denaturing and mixed osmolyte solutions. *Computational and Theoretical Chemistry*, 1131, 78–89.
- Bowman, G. R., Pande, V. S., Noé, F. *Eds.* (2014). *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*. *Advances in Experimental Medicine and Biology*. 797. Springer.
- Branden, C. I., Tooze, J. (1999). *Introduction to Protein Structure* 2<sup>nd</sup> Ed. Garland Science.
- Bürkle, A. (2001). Posttranslational Modification. *Encyclopedia of Genetics*, 1533.
- Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1), 014101.
- Byrne, A., Williams, D. V., Barua, B., Hagen, S. J., Kier, B. L., & Andersen, N. H. (2014). Folding Dynamics and Pathways of the Trp-Cage Miniproteins. *Biochemistry*, 53(38), 6011–6021.
- Cafilisch, A., & Paci, E. (2005). Molecular dynamics simulations to study protein folding and unfolding. *Protein Folding Handbook*, 1143-1169.
- Chaudhuri, T. K., & Paul, S. (2006). Protein-misfolding diseases and chaperone-based therapeutic approaches. *FEBS Journal*, 273(7), 1331–1349.
- Chiti, F., & Dobson, C. (2006). Protein misfolding, functional amyloid and human disease. *Annu. Rev. Biochem.* 75, 333–366.
- Chodera, J. D., Noé, F. (2014). Markov state models of biomolecular conformational dynamics. *Curr. Op. Struct. Biol.* 25:135-144.
- Chowdhury, S., Lee, M. C., & Duan, Y. (2004). Characterizing the rate-limiting step of Trp-cage folding by all-atom molecular dynamics simulations. *The Journal of Physical Chemistry B*, 108(36), 13855-13865.
- Chowdhury, S., Lee, M. C., Xiong, G., & Duan, Y. (2003). Ab initio Folding Simulation of the Trp-cage Mini-protein Approaches NMR Resolution. *Journal of Molecular Biology*, 327(3), 711–717.
- Cooper G. (2000). *Plegado y procesamiento de proteínas. La célula: un enfoque molecular*. 2<sup>a</sup> edición. Sunderland (MA): Sinauer Associates
- Cotham, W. E., Metz, T. O., Ferguson, P. L., Brock, J. W. C., Hinton, D. J. S., Thorpe, S. R., ... Ames, J. M. (2004). Proteomic Analysis of Arginine Adducts on Glyoxal-modified Ribonuclease. *Molecular & Cellular Proteomics*, 3(12), 1145–1153.
- Culik, R. M., Serrano, A. L., Bunagan, M. R., & Gai, F. (2011). Achieving Secondary Structural Resolution in Kinetic Measurements of Protein Folding: A Case Study of the Folding Mechanism of Trp-cage. *Angewandte Chemie*, 123(46), 11076–11079.
- D'Autréaux, B., & Toledano, M. B. (2007). ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis. *Nature reviews Molecular cell biology*, 8(10), 813.
- Davies, M. J. (2003). Singlet oxygen-mediated damage to proteins and its consequences. *Biochemical and biophysical research communications*, 305(3), 761-770.

- Davies, M. J. (2016). Protein oxidation and peroxidation. *Biochemical Journal*, 473(7), 805–825.
- DeLano, W. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, 40, 82-92.
- Deng, N. J., Dai, W., & Levy, R. M. (2013). How kinetics within the unfolded state affects protein folding: An analysis based on Markov state models and an ultra-long MD trajectory. *The Journal of Physical Chemistry B*, 117(42), 12787-12799.
- Dill, K. A., & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Structural & Molecular Biology*, 4(1), 10–19.
- Ding, F., Buldyrev, S. V., & Dokholyan, N. V. (2005). Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model. *Biophysical journal*, 88(1), 147–155.
- Dobson, C. M. (2003). Protein folding and misfolding. *Nature*, 426(6968), 884–890.
- Dobson, CM (1999) Protein misfolding, evolution and disease. *Trends Biochem Sci* 24, 329– 332.
- Dreaden, T. M., Chen, J., Rexroth, S., & Barry, B. A. (2011). N-formylkynurenine as a marker of high light stress in photosynthesis. *Journal of Biological Chemistry*, 286(25), 22632-22641.
- Ehrenschaft, M., de Oliveira Silva, S., Perdivara, I., Bilski, P., Sik, R. H., Chignell, C. F., ... Mason, R. P. (2009). Immunological detection of N-formylkynurenine in oxidized proteins. *Free Radical Biology and Medicine*, 46(9), 1260–1266.
- Ehrenschaft, M., Deterding, L. J., & Mason, R. P. (2015). Tripping up Trp: Modification of protein tryptophan residues by reactive oxygen species, modes of detection, and biological consequences. *Free Radical Biology and Medicine*, 89, 220–228.
- Ehrenschaft, M., Roberts, J. E., & Mason, R. P. (2013). Hypericin-mediated photooxidative damage of  $\alpha$ -crystallin in human lens epithelial cells. *Free Radical Biology and Medicine*, 60, 347–354.
- Feher, J. (2017). Protein Structure. *Quantitative Human Physiology*, 130–141.
- Fersht, A. (1999). *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. Macmillan.
- Frauenfelder, H., Sligar, S., & Wolynes, P. (1991). The energy landscapes and motions of proteins. *Science*, 254(5038), 1598–1603.
- Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels,

- O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, & D. J. Fox. (2013). Revision D.01, Gaussian, Inc., Wallingford CT.
- Frolov, A., Schmidt, R., Spiller, S., Greifenhagen, U., Hoffmann, R. (2014). Journal of Agricultural and Food Chemistry, 62, 3626-3635.
- Gießauf, A., van Wickern, B., Simat, T., Steinhart, H., & Esterbauer, H. (1996). Formation of N-formylkynurenine suggests the involvement of apolipoprotein B-100 centered tryptophan radicals in the initiation of LDL lipid peroxidation. *FEBS letters*, 389(2), 136-140.
- Glomb, M. A., Lang, G. (2001) Isolation and characterization of glyoxalarginine modifications. *J. Agric. Food Chem.*, 49, 1493–1501.
- Godbey, W. T. (2014). *Proteins. An Introduction to Biotechnology*, 9–33.
- González-Torres, M., & Betancourt-Rule, M., & Ortiz-Muñiz, R. (2000). Daño Oxidativo y Antioxidantes. *Bioquímica*, 25 (1), 3-9.
- Gromiha, M. M., Ahmad, S., & Suwa, M. (2005). TMBETA-NET: discrimination and prediction of membrane spanning -strands in outer membrane proteins. *Nucleic Acids Research*, 33(Web Server), W164–W167.
- Guptasarma, P., Balasubramanian, D., Matsugo, S., & Saito, I. (1992). Hydroxyl radical mediated damage to proteins, with special reference to the crystallins. *Biochemistry*, 31(17), 4296-4303.
- Halliwell, B. (2009). The wanderings of a free radical. *Free Radical Biology and Medicine*; 46(5):531–542. 19111608
- Han, W., & Schulten, K. (2013). Characterization of folding mechanisms of Trp-cage and WW-domain by network analysis of simulations with a hybrid-resolution model. *The Journal of Physical Chemistry B*, 117(42), 13367-13377.
- Hartl, F. U. (2017). Protein Misfolding Diseases. *Annual Review of Biochemistry*, 86(1), 21–26.
- Heyda, J., Kožíšek, M., Bednárova, L., Thompson, G., Konvalinka, J., Vondrášek, J., & Jungwirth, P. (2011). Urea and Guanidinium Induced Denaturation of a Trp-Cage Mini-protein. *The Journal of Physical Chemistry B*, 115(28), 8910–8924.
- Houde, D., Peng, Y., Berkowitz, S. A., & Engen, J. R. (2010). Post-translational Modifications Differentially Affect IgG1 Conformation and Receptor Binding. *Molecular & Cellular Proteomics*, 9(8), 1716–1728.
- Hu, Z., Tang, Y., Wang, H., Zhang, X., & Lei, M. (2008). Dynamics and cooperativity of Trp-cage folding. *Archives of Biochemistry and Biophysics*, 475(2), 140–147.
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33–38.
- Husic, B. E., Pande, V. S., (2018) Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.*, 140(7):2386-2396.

- Hvidsten, T. R., Laegreid, A., Kryshtafovych, A., Andersson, G., Fidelis, K., & Komorowski, J. (2009). A comprehensive analysis of the structure-function relationship in proteins based on local structure similarity, *4*(7), e6266.
- Ijima, K., Murata, M., Takahara, H., Irie, S. And Fujimoto, D. (2000 ) Identification of N $\omega$ -carboxymethylarginine as a novel acid-labile advanced glycation end product in collagen. *Biochem. J.* 347, 23– 27.
- Jensen, O. (2004). Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Current Opinion in Chemical Biology*, 8(1), 33–41.
- Jorgensen, W. L. (1981). Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *Journal of the American Chemical Society*, 103(2), 335–340.
- Juraszek, J., & Bolhuis, P. G. (2006). Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proceedings of the National Academy of Sciences*, 103(43), 15859-15864.
- Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12), 2577–2637.
- Kannan, S., & Zacharias, M. (2009). Folding simulations of Trp-cage mini protein in explicit solvent using biasing potential replica-exchange molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*, 76(2), 448–460.
- Kannan, S., & Zacharias, M. (2014). Role of Tryptophan Side Chain Dynamics on the Trp-Cage Mini-Protein Folding Studied by Molecular Dynamics Simulations. *PLoS ONE*, 9(2), e88383.
- Karplus, M., & Weaver, D. L. (1994). Protein folding dynamics: the diffusion-collision model and experimental data. *Protein science: a publication of the Protein Society*, 3(4), 650–668.
- Kasson, T. M. D., & Barry, B. A. (2012). Reactive oxygen and oxidative stress: N-formyl kynurenine in photosystem II and non-photosynthetic proteins. *Photosynthesis Research*, 114(2), 97–110.
- Kasson, T. M. D., Rexroth, S., & Barry, B. A. (2012). Light-induced oxidative stress, N-formylkynurenine, and oxygenic photosynthesis. *PLoS One*, 7(7), e42220.
- Kim, P. S., & Baldwin, R. L. (1982). Specific Intermediates in the Folding Reactions of Small Proteins and the Mechanism of Protein Folding. *Annual Review of Biochemistry*, 51(1), 459–489.
- Kim, S. B., Palmer, J. C., & Debenedetti, P. G. (2016). Computational investigation of cold denaturation in the Trp-cage miniprotein. *Proceedings of the National Academy of Sciences*, 113(32), 8991–8996.
- Kinoshita, K., & Nakamura, H. (2003). Protein informatics towards function identification. *Current Opinion in Structural Biology*, 13(3), 396–400.
- Kundu, S. 2005. Amino acids network within protein. *Physica A*. 346:104–109.
- Lai, Y.-L., Aoyama, S., Nagai, R., Miyoshi, N., & Ohshima, H. (2010). Inhibition of L-Arginine Metabolizing Enzymes by L-Arginine-Derived Advanced Glycation End Products. *Journal of Clinical Biochemistry and Nutrition*, 46(2), 177–185.

- Lee, C., Yang, W., & Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2), 785–789.
- Lee, D., Redfern, O., & Orengo, C. (2007). Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol*; 8:995–1005.
- Lesk, A. M., & Chothia, C. (1980). Solvent accessibility, protein surfaces, and protein folding. *Biophysical Journal*, 32(1), 35–47.
- Levinthal, C. (1968). Are there pathways for protein folding?. *Journal de chimie physique*, 65, 44-45.
- Lindorff-Larsen, K., Piana, S., Dror, R. O., & Shaw, D. E. (2011). How Fast-Folding Proteins Fold. *Science*, 334(6055), 517–520.
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., & Shaw, D. E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics*, NA–NA.
- Lins, L., & Brasseur, R. (1995). The hydrophobic effect in protein folding. *The FASEB Journal*, 9(7), 535–540.
- Louie, J. L., Kappahn, R. J., & Ferrington, D. A. (2002). Proteasome function and protein oxidation in the aged retina. *Experimental eye research*, 75(3), 271-284.
- Manzanares, D., Rodriguez-Capote, K., Liu, S., Haines, T., Ramos, Y., Zhao, L., & Possmayer, F. (2007). Modification of tryptophan and methionine residues is implicated in the oxidative inactivation of surfactant protein B. *Biochemistry*, 46(18), 5604-5615.
- Marenich, A. V., Cramer, C. J., & Truhlar, D. G. (2009). Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *The Journal of Physical Chemistry B*, 113(18), 6378–6396.
- Mariño, L., Casasnovas, R., Ramis, R., Vilanova, B., Ortega-Castro, J., Frau, J., Adrover, M. (2019). Does glycation really distort the peptide  $\alpha$ -helicity? *International Journal of Biological Macromolecules*, 129, 254-266.
- Mariño, L., Pauwels, K., Casasnovas, R., Sanchis, P., Vilanova, Muñoz, F., Donoso, J., Adrover, M. (2015) Ortho-methylated 3-hydroxypyridines hinder hen egg-white lysozyme fibrillogenesis. 5, 12052.
- Meuzelaar, H., Marino, K. A., Huerta-Viga, A., Panman, M. R., Smeenk, L. E. J., Kettelarij, A. J., ... Woutersen, S. (2013). Folding Dynamics of the Trp-Cage Miniprotein: Evidence for a Native-Like Intermediate from Combined Time-Resolved Vibrational Spectroscopy and Molecular Dynamics Simulations. *The Journal of Physical Chemistry B*, 117(39), 11490–11501.
- Mukherjee, A., Morales-Scheihing, D., Butler, P.C., Soto, C. (2015). Type 2 diabetes as a protein misfolding disease, *Trends Mol. Med.* 21, 439–449.
- Nedić, O., Rattan, S. I. S., Grune, T., & Trougakos, I. P. (2013). Molecular effects of advanced glycation end products on cell signalling pathways, ageing and pathophysiology. *Free Radical Research*, 47(sup1), 28–38.



- Nedić, O., Rogowska-Wrzesinska, A., & Rattan, S. (2015). Standardization and quality control in quantifying non-enzymatic oxidative protein modifications in relation to ageing and disease: Why is it important and why is it hard? *Redox biology*, 5, 91–100.
- Neidigh, J. W., R. M. Fesinmeyer, K. S. Prickett, & N. H. Andersen. (2001). Exendin-4 and glucagon-like-peptide-1: NMR structural comparisons in the solution and micelle-associated states. *Biochemistry*, 40:13188–13200.
- Neidigh, J., & Fesinmeyer, R., & Niels, A. (2002). Designing a 20-residue protein. *Nature structural biology*, 9, 425–3.
- Neuweiler, H., Doose, S., & Sauer, M. (2005). A microscopic view of miniprotein folding: Enhanced folding efficiency through formation of an intermediate. *Proceedings of the National Academy of Sciences*, 102(46), 16650–16655.
- Nowotny, K., Jung, T., Grun,e T., & Höhn, A. (2014). Accumulation of modified proteins and aggregate formation in aging. *Experimental Gerontology*; 57:122–131. 24877899.
- Nozawa, K., Schneider, T. R., & Cramer, P. (2017). Core Mediator structure at 3.4 Å extends model of transcription initiation complex. *Nature*, 545(7653), 248–251.
- O’Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(1), 33.
- Odani, H., Iijima, K., Nakata, M., Miyata, S., Kusunoki, H., Yasuda, Y., Hiki, Y., Irie, S., Maeda, K., & Fujimoto, D. (2001). Identification of N $\omega$ -carboxymethylarginine, a new advanced glycation endproduct in serum proteins of diabetic patients: Possibility of a new marker of aging and diabetes. *Biochem. Biophys. Res. Commun.* 285, 1232– 1236.
- Olivares-Quiroz, L., & García-Colín, L. (2004). Plegamiento de las proteínas: Un problema interdisciplinario. *Revista de la Sociedad Química de México*, 48(1), 95-105.
- Omenn G. S., Lane L., Lundberg E. K., Beavis R. C., Overall C. M., Deutsch E. W. Progress on identifying and characterizing the human proteome, including post-translational modifications. *Journal of Proteome Research*. 2016;15(11):3951–3960.
- Pandini, A., & Fornili, A. (2016). Using local states to drive the sampling of global conformations in proteins. *Journal of chemical theory and computation*, 12(3), 1368-1379.
- Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12), 7182–7190.
- Paschek, D., Hempel, S., & García, A. E. (2008). Computing the stability diagram of the Trp-cage miniprotein. *Proceedings of the National Academy of Sciences*, 105(46), 17754-17759.
- Pazos, F., & Sternberg, M. J. E. (2004). Automated prediction of protein function and detection of functional sites from structure. *Proceedings of the National Academy of Sciences*, 101(41), 14754–14759.
- Pelley, J. (2007). *Protein Structure and Function*. Elsevier's Integrated Biochemistry. Mosby. 19-28.

- Piedrafita, G., Keller, M., & Ralser, M. (2015). The Impact of Non-Enzymatic Reactions and Enzyme Promiscuity on Cellular Metabolism during (Oxidative) Stress Conditions. *Biomolecules*, 5(3), 2101–2122.
- Plattner, N., Noé, F. (2015). Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* 6:7653.
- Previero, A., Coletti-Previero, M. A., & Jolles, P. (1967). Localization of non-essential tryptophan residues for the biological activity of lysozymet. *Journal of molecular biology*, 24(2), 261-268.
- Prinz, J. H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., Chodera, J. D., Schütte, C., Noé, F. (2011). Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* 134(17):174105.
- Qiu, L., Pabit, S. A., Roitberg, A. E., & Hagen, S. J. (2002). Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4  $\mu$ s. *Journal of the American Chemical Society*, 124(44), 12952–12953.
- Radford, S. E. & Dobson, C. M. (1999). From computer simulations to human disease: emerging themes in protein folding. *Cell* 97, 291–298.
- Rehman, I., & Botelho, S. (2018). *Estructura de la proteína secundaria*. Bioquímica.Treasure Island (FL), StatPearls.
- Reigle, K. L., Di Lullo, G., Turner, K. R., Last, J. A., Chervoneva, I., Birk, D. E., ... San Antonio, J. D. (2008). Non-enzymatic glycation of type I collagen diminishes collagen-proteoglycan binding and weakens cell adhesion. *Journal of Cellular Biochemistry*, 104(5), 1684–1698.
- Rittle, J., Field, M. J., Green, M. T., & Tezcan, F. A. (2019). An efficient, step-economical strategy for the design of functional metalloproteins. *Nature Chemistry*.
- Robustelli, P., Piana, S., & Shaw, D. E. (2018). Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences*, 115(21), E4758–E4766.
- Salahuddin, P., Rabbani, G., Khan, R. H. (2014) The role of advanced glycation end products in various types of neurodegenerative disease: a therapeutic approach, *Cell. Mol. Biol. Lett.* 19, 407–437.
- Schwarzenbolz, U.; Henle, T.; Haeßner, R.; Klostermeyer, H. (1997). On the reaction of glyoxal with proteins. *Z. Lebensm. Unters. Forsch. A*, 205, 121–124
- Seetharaman, S. (2016). The Influences of Dietary Sugar and Related Metabolic Disorders on Cognitive Aging and Dementia. *Molecular Basis of Nutrition and Aging*, 331–344.
- Senne, M., Trendelkamp-Schroer, B., Mey, A. S. J. S., Schütte, C., & Noé, F. (2012). EMMA: A Software Package for Markov Model Building and Analysis. *Journal of Chemical Theory and Computation*, 8(7), 2223–2238.
- Shimasaki, S., Kubota, M., Yoshitomi, M., Takagi, K., Suda, K., Mera, K., Nagai, R. (2011). N $\omega$ -(carboxymethyl)arginine Accumulates in Glycated Collagen and Klotho-deficient Mouse Skin. *ANTI-AGING MEDICINE*, 8(6), 82–87.
- Shukla, D., Meng, Y., Roux, B., Pande, V. S. (2014) Activation Pathway of Src Kinase Reveals Intermediate States as Targets for Drug Design. *Nat. Commun.*, 5, 3397.

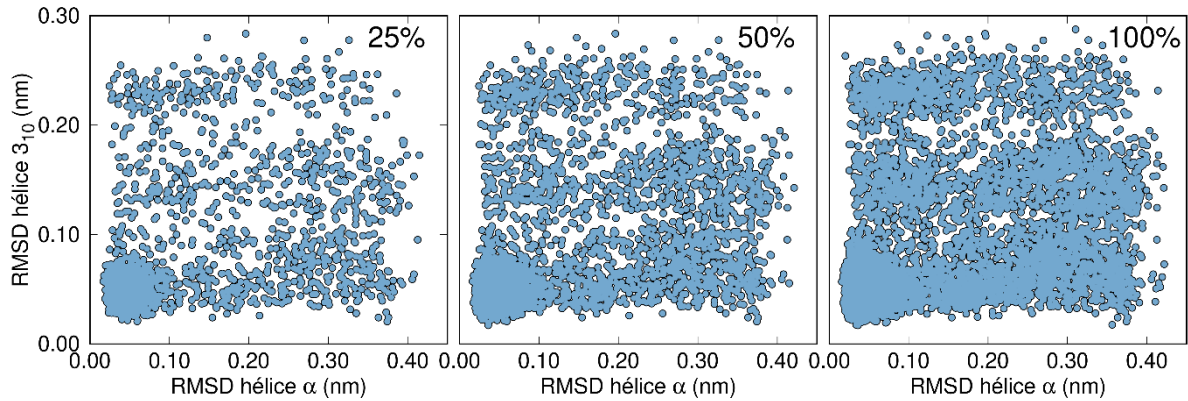
- Snow, C. D., Zagrovic, B., & Pande, V. S. (2002). The Trp Cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations. *Journal of the American Chemical Society*, 124(49), 14548–14549.
- Šoškić, V., Groebe, K., & Schratzenholz, A. (2008). Nonenzymatic posttranslational protein modifications in ageing. *Experimental Gerontology*, 43(4), 247–257.
- Squier, T. C. (2001). Oxidative stress and protein aggregation during biological aging. *Experimental gerontology*, 36(9), 1539-1550.
- Stadtman, E. R., & Levine, R. L. (2003). Free radical-mediated oxidation of free amino acids and amino acid residues in proteins. *Amino acids*, 25(3-4), 207-218.
- Stultz, C. M., & Edelman, E. R. (2003). A Structural Model that Explains the Effects of Hyperglycemia on Collagenolysis. *Biophysical Journal*, 85(4), 2198–2204.
- Taylor, D. M., Gibbs, B. F., Kabashi, E., Minotti, S., Durham, H. D., & Agar, J. N. (2007). Tryptophan 32 potentiates aggregation and cytotoxicity of a copper/zinc superoxide dismutase mutant associated with familial amyotrophic lateral sclerosis. *Journal of Biological Chemistry*, 282(22), 16329-16335.
- Todorovski, T., Fedorova, M., Hennig, L., & Hoffmann, R. (2011). Synthesis of peptides containing 5-hydroxytryptophan, oxindolylalanine, N-formylkynurenine and kynurenine. *Journal of Peptide Science*, 17(4), 256–262.
- Touw, W. G., Baakman, C., Black, J., te Beek, T. A. H., Krieger, E., Joosten, R. P., & Vriend, G. (2014). A series of PDB-related databanks for everyday needs. *Nucleic Acids Research*, 43(D1), D364–D368.
- Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., & Bussi, G. (2014). PLUMED 2: New feathers for an old bird. *Computer Physics Communications*, 185(2), 604–613.
- Tuckerman, M. E. (2007). *Free Energy Calculations: Theory and Applications in Chemistry and Biology* Springer Series in Chemical Physics, 86.
- Valastyan, J. S., & Lindquist, S. (2014). Mechanisms of protein-folding diseases at a glance. *Disease models & mechanisms*, 7(1), 9–14.
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16), 1701–1718.
- Verlet, L. (1967)- Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review*, 159, 98-103.
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9), 1157–1174.
- Wetlaufer, D. B. (1973). Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins. *Proceedings of the National Academy of Sciences*, 70(3), 697–701.
- Williams, T., & Kelley, C. (2011). Gnuplot 4.5: an interactive plotting program. URL <http://gnuplot.info>.

Zheng, W., Gallicchio, E., Deng, N., Andrec, M., & Levy, R. M. (2011). Kinetic Network Study of the Diversity and Temperature Dependence of Trp-Cage Folding Pathways: Combining Transition Path Theory with Stochastic Simulations. *The Journal of Physical Chemistry B*, 115(6), 1512–1523.

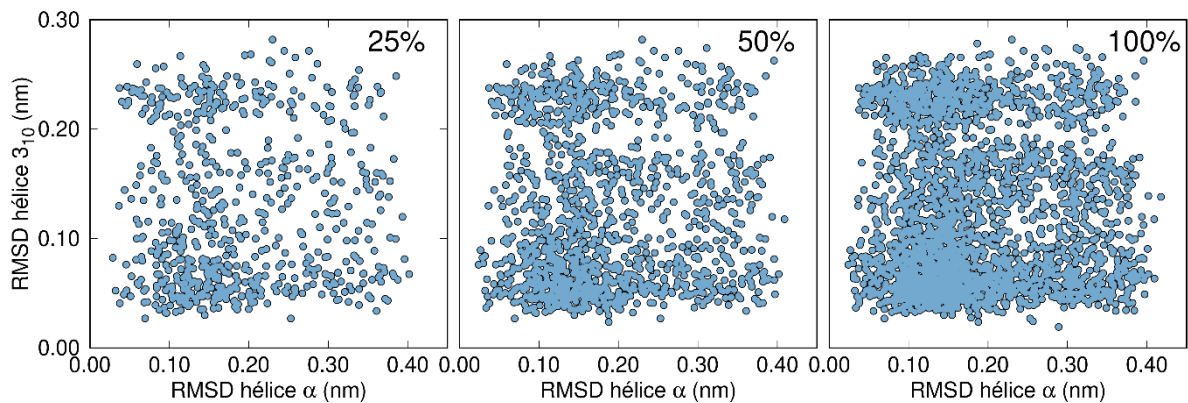
Zhou R. (2003). Trp-cage: folding free energy landscape in explicit water. *Proceedings of the National Academy of Sciences of the United States of America*, 100(23), 13280–13285.

Zhou, C.-Y., Jiang, F., & Wu, Y.-D. (2015). Folding Thermodynamics and Mechanism of Five Trp-Cage Variants from Replica-Exchange MD Simulations with RSFF2 Force Field. *Journal of Chemical Theory and Computation*, 11(11), 5473–5480.

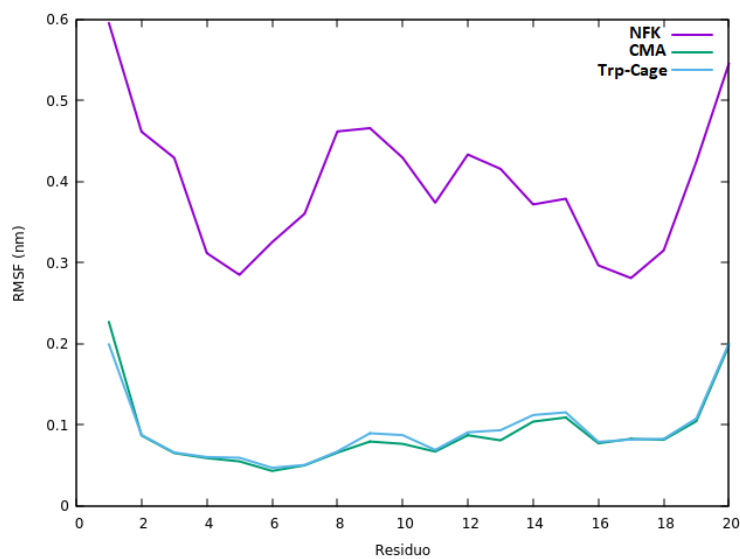
## Anexo



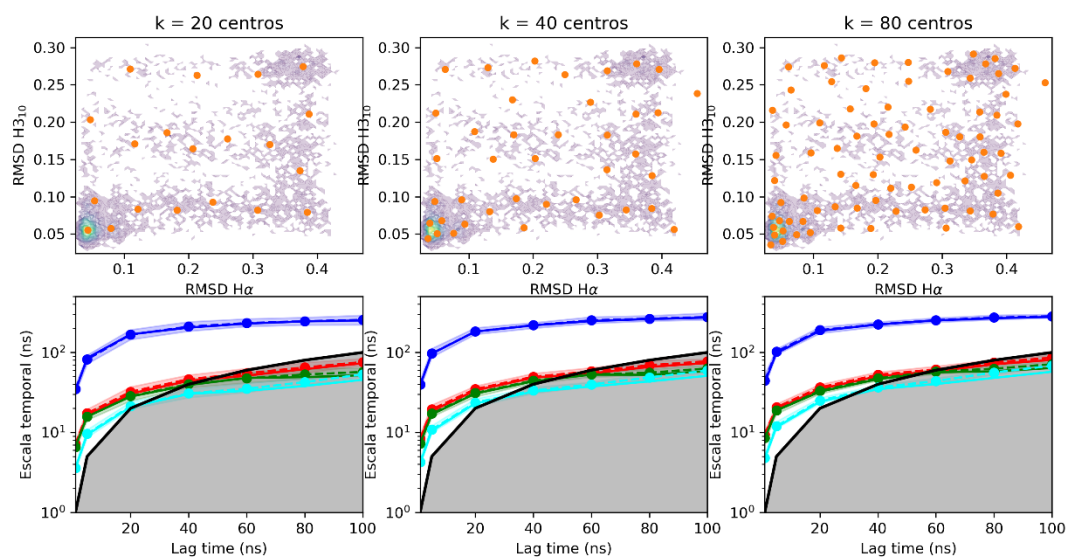
**Figura A1.** Muestreo considerando el 25% (izquierda), 50% (centro) y 100% (derecha) de las conformaciones de cada trayectoria, para todas las trayectorias simuladas de CMA.



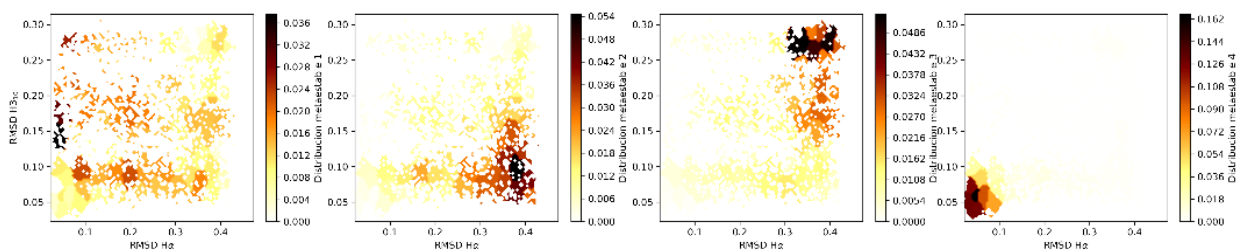
**Figura A2.** Muestreo considerando el 25% (izquierda), 50% (centro) y 100% (derecha) de las conformaciones de cada trayectoria, para todas las trayectorias simuladas de NFK.



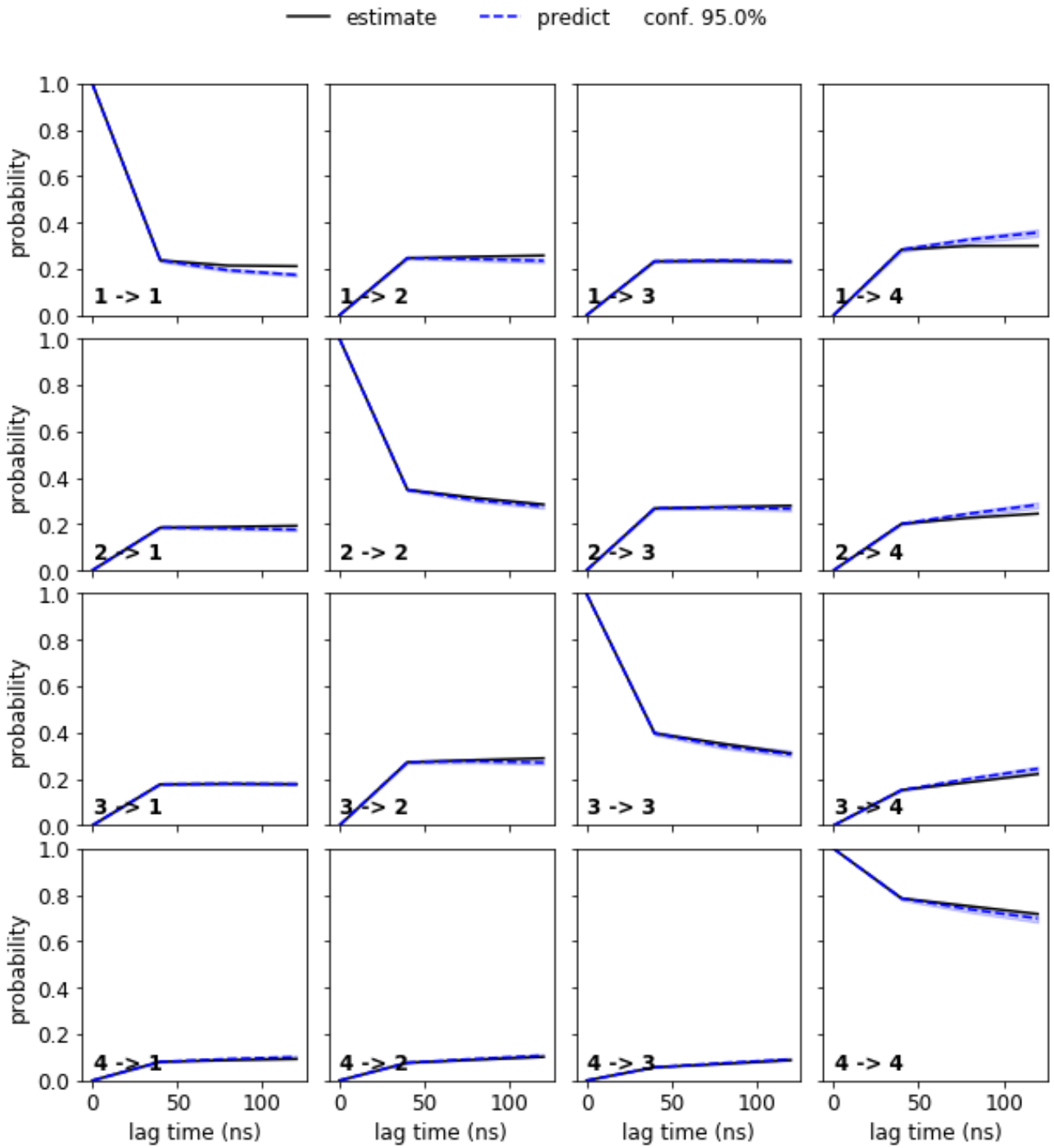
**Figura A3.** RMSF por residuo de Trp-Cage (azul), NFK (lila) y CMA (verde).



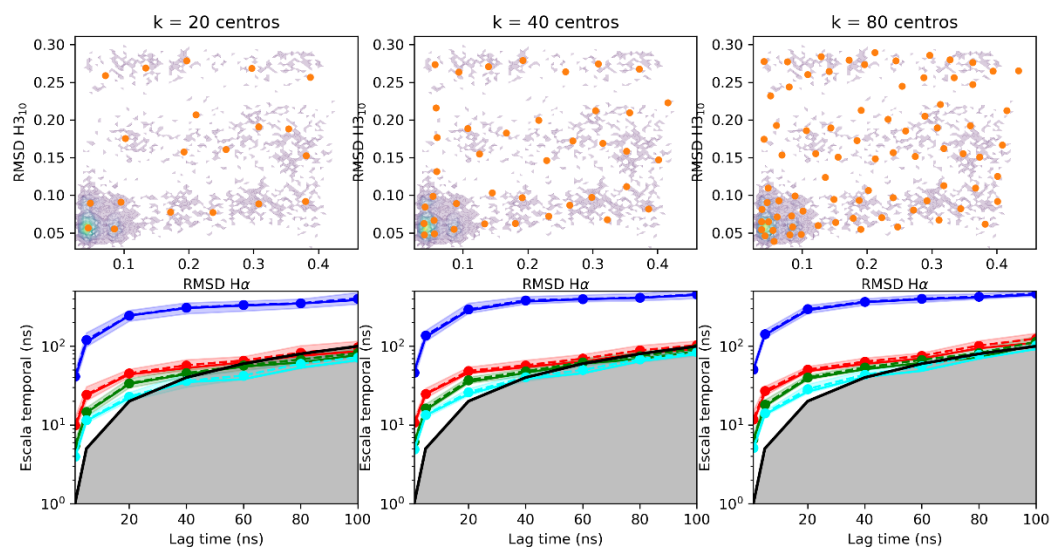
**Figura A4.** Parte superior: centros de los clústeres de Trp-Cage sobre la densidad de población en función de los RMSD de la hélice  $3_{10}$  y la hélice  $\alpha$ . Parte inferior: Escalas temporales para los 4 procesos más lentos (colores y en orden temporal ascendente: azul, rojo, verde, cian, morado).



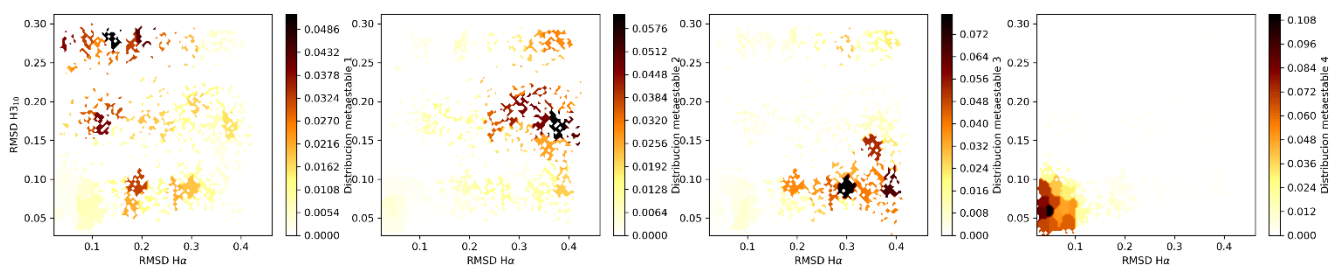
**Figura A5.** Distribuciones metaestables en función de los RMSD de la hélice  $3_{10}$  y la hélice  $\alpha$  de Trp-Cage.



**Figura A6.** Test CK para un modelo Markoviano de 4 macroestados para Trp-Cage.



**Figura A7.** Parte superior: centros de los clústeres de CMA sobre la densidad de población en función de los RMSD de la hélice  $3_{10}$  y la hélice  $\alpha$ . Parte inferior: Escalas temporales para los 4 procesos más lentos (colores y en orden temporal ascendente: azul, rojo, verde, cyan, morado).



**Figura A8.** Distribuciones metaestables en función de los RMSD de la hélice  $3_{10}$  y la hélice  $\alpha$  de CMA.



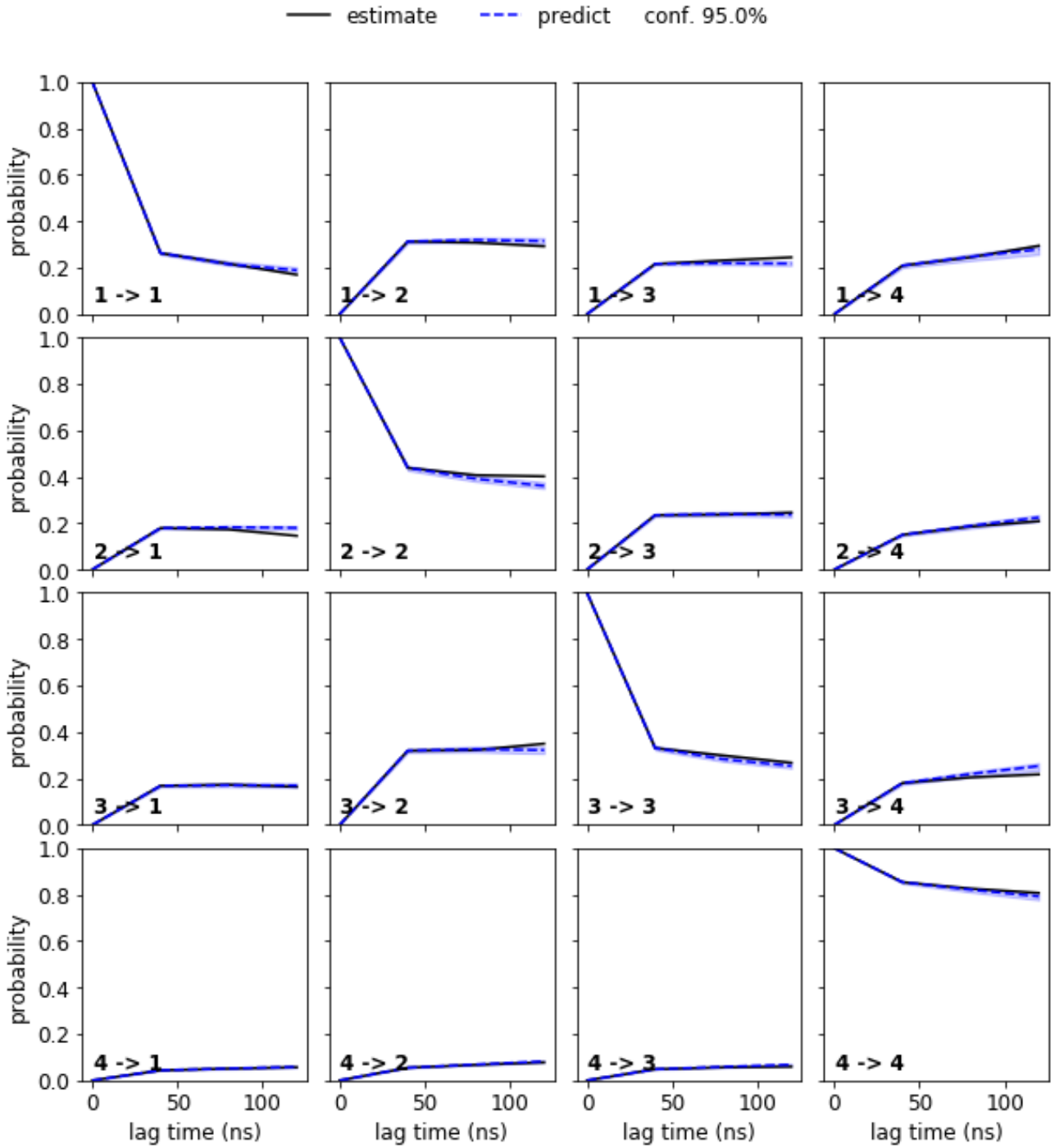
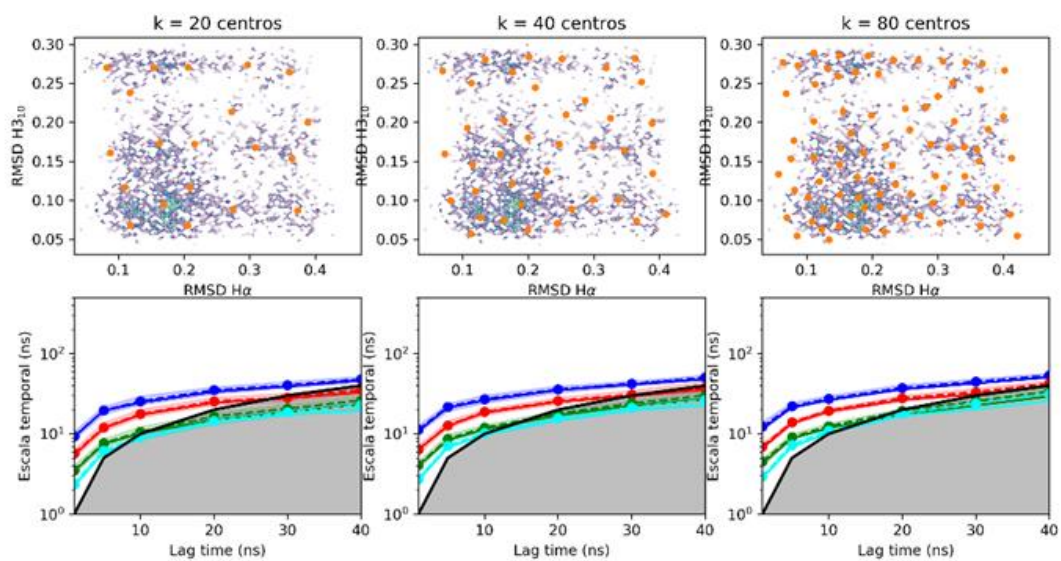
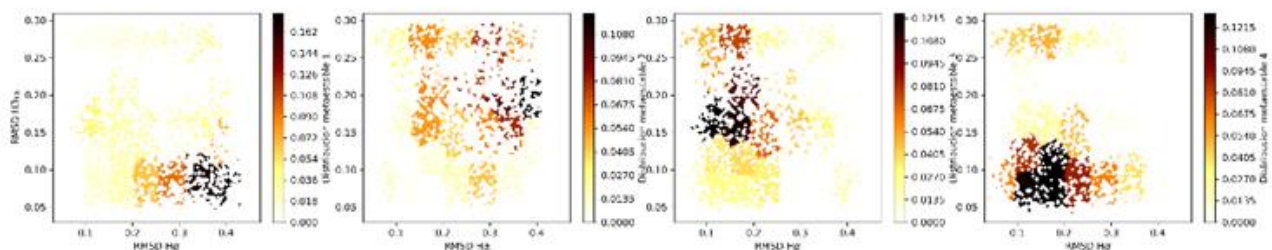


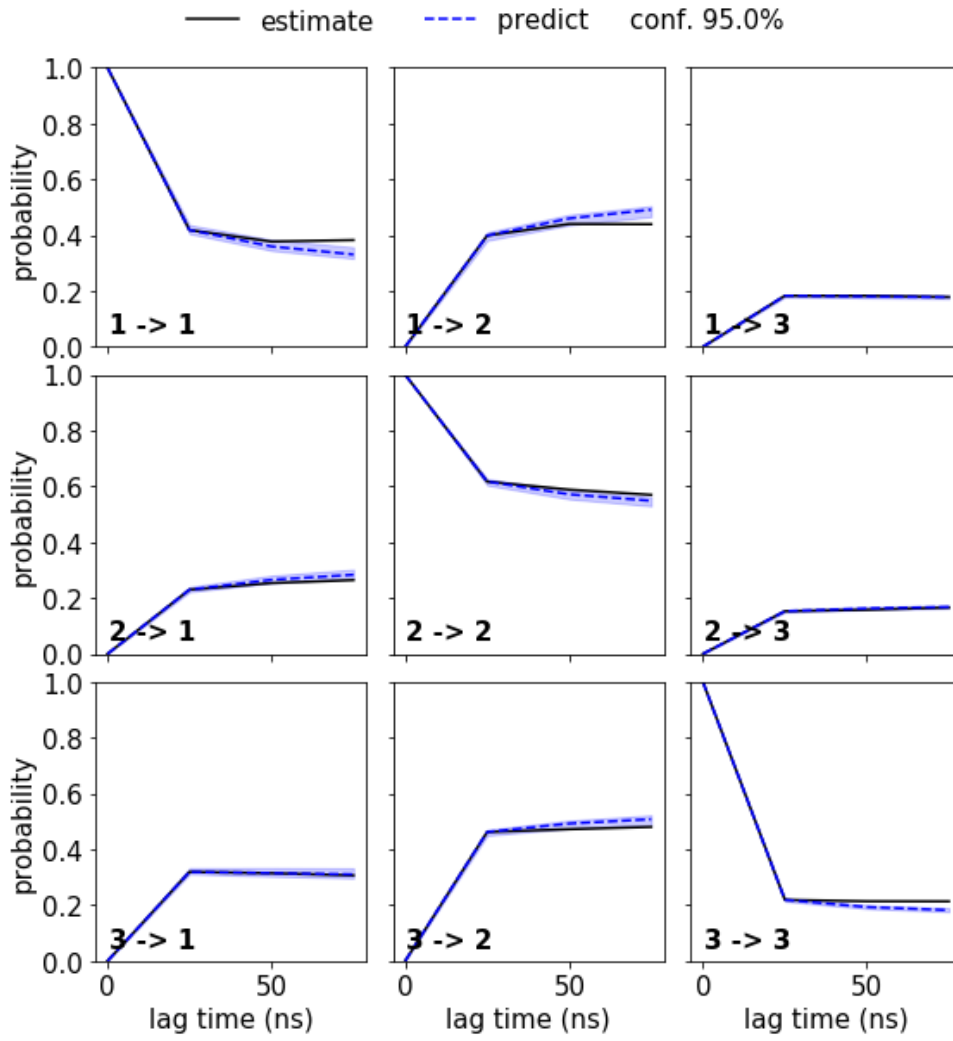
Figura A9. Test CK para un modelo Markoviano de 4 macroestados para CMA.



**Figura A10.** Parte superior: centros de los clústeres de NFK sobre la densidad de población en función de los RMSD de la hélice  $3_{10}$  y la hélice  $\alpha$ . Parte inferior: Escalas temporales para los 4 procesos más lentos (colores y en orden temporal ascendente: azul, rojo, verde, cian, morado).



**Figura A11.** Distribuciones metaestables en función de los RMSD de la hélice  $3_{10}$  y la hélice  $\alpha$  de NFK.



**Figura A12.** Test CK para un modelo Markoviano de 3 macroestados para NFK.

Radio de giro (nm)			
	Trp-Cage	NFK	CMA
Estado 1	0.72 ± 0.02	0.82 ± 0.09	0.72 ± 0.02
Estado 2	0.80 ± 0.08	-	0.81 ± 0.07
Estado 3	0.81 ± 0.10	0.85 ± 0.10	0.82 ± 0.10
Estado 4	0.83 ± 0.07	0.84 ± 0.11	0.82 ± 0.07
Estado 5	0.86 ± 0.09	0.86 ± 0.10	0.85 ± 0.08
Estado 6	0.83 ± 0.09	0.88 ± 0.10	0.87 ± 0.11

**Tabla A1.** Radio de giro de Trp-Cage, NFK y CMA en cada uno de sus estados.

% Clúster más poblado			
	Trp-Cage	NFK	CMA
Estado 1	99,9	4	99,9
Estado 2	19	-	11
Estado 3	11	6	8
Estado 4	4	13	18
Estado 5	2	4	3
Estado 6	8	5	3

**Tabla A2.** Porcentaje Clúster más poblado para cada estado de Trp-Cage, NFK y CMA.

% SASA Total			
	Trp-Cage	NFK	CMA
Estado 1	70 ± 3	75 ± 6	66 ± 3
Estado 2	74 ± 5	-	74 ± 5
Estado 3	75 ± 7	78 ± 6	75 ± 7
Estado 4	77 ± 6	78 ± 8	77 ± 5
Estado 5	79 ± 6	79 ± 6	78 ± 6
Estado 6	76 ± 7	82 ± 7	80 ± 7

**Tabla A3.** Superficies accesibles al disolvente total de Trp-Cage, NFK y CMA.

Distancia (nm)	
Estado 1	0.31 ± 0.1
Estado 2	-
Estado 3	0.31 ± 0.1
Estado 4	0.4 ± 0.1
Estado 5	0.4 ± 0.1
Estado 6	0.4 ± 0.1

**Tabla A4.** Interacción entre el Oxígeno y el Hidrogeno del backbone del residuo NFK6 en cada uno de sus estados.