



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE QUÍMICA

**Caracterización del ensamble conformacional
de proteínas y su dependencia con el pH**

T E S I S

**QUE PARA OBTENER EL TÍTULO DE
QUÍMICA**

PRESENTA

DULCE CONSUELO GUZMÁN OCAMPO

Ciudad Universitaria, CDMX, 2017





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

PRESIDENTE: LILIA ANGÉLICA BERNAL GRACIDA

VOCAL: JUAN RAÚL ÁLVAREZ IDABOY

SECRETARIO: LAURA DOMÍNGUEZ DUEÑAS

1er. SUPLENTE: BRUNO CHRISTIAN LANDEROS RIVERA

2° SUPLENTE: JOSÉ CARLOS PÁEZ FRANCO

ESTE TRABAJO SE REALIZÓ EN LA FACULTAD DE QUÍMICA, UNAM, CU. EDIFICIO "F"
2DO PISO.

ASESOR

LAURA DOMÍNGUEZ DUEÑAS

SUPERVISOR TÉCNICO

RODRIGO AGUAYO ORTIZ

SUSTENTANTE

DULCE CONSUELO GUZMÁN OCAMPO

AGRADECIMIENTOS

A la Universidad Nacional Autónoma de México y a la Facultad de Química por formarme como Química y permitirme ver el mundo desde otras perspectivas.

Agradezco el apoyo financiero al proyecto por parte de la DGAPA mediante el programa PAPIIT (PAPIIT IA RA204716) y la proporción de recursos de Supercómputo por parte de la DGTIC (LANCAD-UNAM-DGTIC-306).

Agradezco al Sistema de Becas para Estudiantes Indígenas (SBEI) por el apoyo que me brindó a lo largo de 5 años; sin su contribución, no habría logrado cumplir mi meta.

ÍNDICE

1. INTRODUCCIÓN	6
2. ANTECEDENTES	7
2.1. Efecto del pH en las proteínas.....	7
2.2. Determinación del pKa mediante el trazado de curvas de titulación.....	8
2.3. Simulación Molecular	9
2.4. Simulaciones de Dinámica Molecular a pH constante	14
2.5. La enfermedad de Alzheimer	23
2.6. Proteínas intrínsecamente desordenadas.....	30
3. Objetivos	36
4. Hipótesis.....	36
5. PROCEDIMIENTO	37
5.1. Detalles generales.....	37
5.2. Preparación del sistema para simulaciones CpHMD.....	37
5.3. Compuestos de referencia	38
5.4. Aminoácidos titulables	38
5.5. pH-REMD.....	38
5.6. Estructura inicial del sistema de la GS.....	39
5.7. Estructura inicial de los sistemas de R2/wt.....	40
5.8. Minimización	40
5.9. Procedimiento de preparación de R2/wt con solvatación implícita	41
5.10. Procedimiento de preparación de R2/wt con solvatación explícita.....	41
5.11. Procedimiento de preparación de la GS.....	41
5.12. Análisis.....	42
6. RESULTADOS Y ANÁLISIS	43
6.1. Primer enfoque: Estudio de los estados de protonación de los aspárticos catalíticos de la enzima γ -secretasa.....	43
6.2. Segundo enfoque: Efecto estructural del pH en un péptido intrínsecamente desordenado.	53
7. CONCLUSIONES.....	65
7.1. Conclusiones particulares.....	66
8. REFERENCIAS	68
9. APÉNDICE.....	75

1. INTRODUCCIÓN

Las proteínas son entidades dinámicas, por lo que caracterizar su ensamble conformacional es un gran reto. La determinación experimental de la estructura de proteínas, además de que tiene grandes limitaciones logísticas, únicamente logra capturar una pequeña muestra de su amplio conjunto de conformaciones. Es por ello que el estudio del ensamble conformacional de las proteínas muchas veces se apoya en el modelado molecular. De esta forma, se puede potenciar la capacidad predictiva y obtener información sobre las propiedades estructurales y dinámicas de las proteínas. El efecto del pH en la estructura de las proteínas es un factor relevante en su caracterización; su importancia radica en los cambios en el estado de protonación de aminoácidos con propiedades ácido-base en su cadena lateral. Estos cambios en los estados de protonación conducen a la ruptura y formación de interacciones que se traducen en cambios conformacionales; a su vez, cambios conformacionales pueden implicar cambios en los estados de protonación debido a alteraciones en el ambiente electrostático del sitio titulable. Esta correlación conformación-protonación no se debe ignorar si se desea llevar a cabo un muestreo correcto de cualquiera de los dos ensambles.

Sumado a la capacidad de la Dinámica Molecular para llevar a cabo el muestreo conformacional; las simulaciones de Dinámica Molecular de intercambio de réplicas a pH constante (*pH-REMD*) permiten llevar a cabo el muestreo de los estados de protonación de manera simultánea. Además, es una técnica que permite superar las barreras energéticas más fácilmente por el paseo de múltiples réplicas en la dimensión de pH.

La γ -secretasa es una proteasa intramembranal que cuenta con dos aspárticos catalíticos en su sitio activo. Tras su catálisis, la γ -secretasa libera los péptidos amiloidogénicos denominados β -amiloides, los cuales se agregan formando oligómeros y fibras tóxicas que dañan a las neuronas. Por otro lado, la proteína tau es una proteína intrínsecamente desordenada encargada de estabilizar a los microtúbulos. Esta proteína tiene en su secuencia un fragmento muy amiloidogénico comúnmente denominado R2/wt que a su vez contiene varios aminoácidos con propiedades ácido-base, y, por tanto, susceptibles al pH.

En este trabajo de tesis, se empleó la metodología de *pH-REMD* para estudiar: **1)** el estado de protonación del sitio activo de la enzima γ -secretasa y **2)** el ensamble conformacional de la proteína tau. Ambas proteínas, estrechamente relacionadas con la enfermedad de Alzheimer.

2. ANTECEDENTES

2.1. Efecto del pH en las proteínas

La estabilidad estructural y el funcionamiento de las proteínas se encuentran relacionados con el pH. El principal efecto del pH en una proteína es el control del equilibrio de protonación de sus sitios titulables, cuya consecuencia más marcada es probablemente el cambio asociado en las cargas e interacciones electrostáticas.¹ Se llama sitio tituable a aquél sitio con algún grupo con propiedades ácido-base según la definición de Brønsted-Lowry, que puede adquirir un estado protonado o desprotonado dependiendo del pH de la disolución. En la Figura 1 se presentan los aminoácidos con cadena lateral con propiedades ácido-base; de aquí en adelante se referirá a estos aminoácidos como titulables. El valor de pH al cual la cadena lateral de un aminoácido tituable tiene en igual proporción estados protonados y desprotonados se le llama pKa de la cadena lateral.

Además del conocido efecto del pH en el estado de protonación de especies con propiedades ácido-base, el valor de pKa de la cadena lateral en los aminoácidos libres es susceptible a cambiar en el entorno de una proteína; esto se debe al cambio en la naturaleza química de su ambiente inmediato. Los principales efectos ambientales que provocan este cambio son de naturaleza electrostática, como la desolvatación, las interacciones coulómbicas y puentes de hidrógeno.²

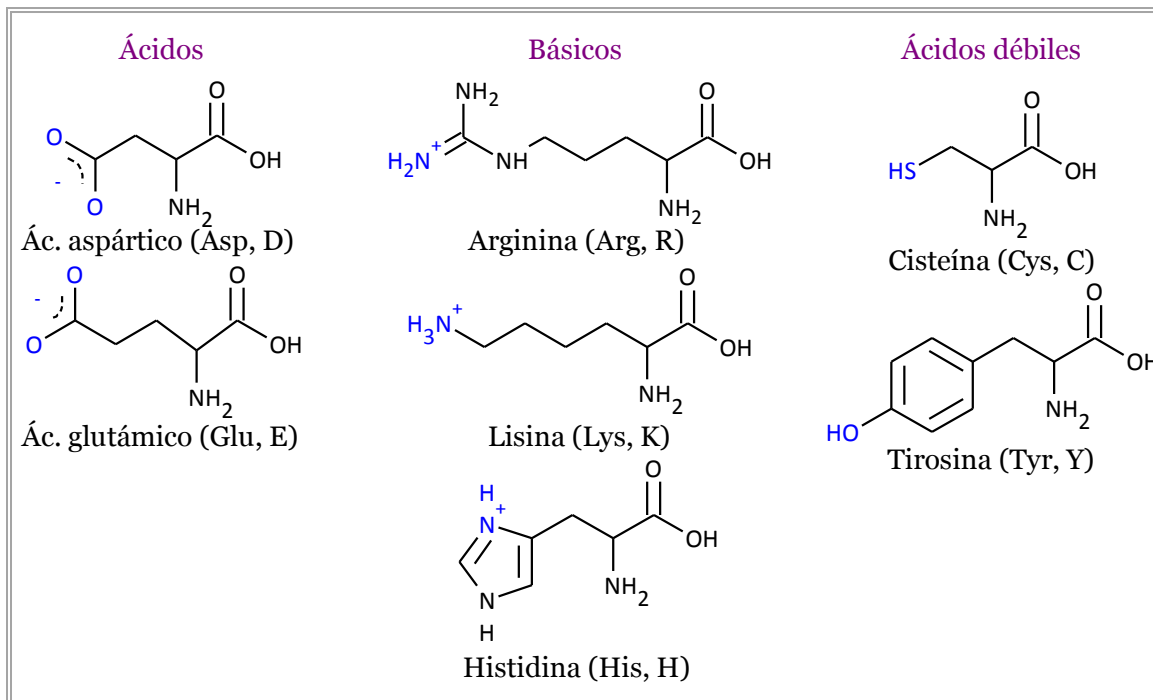


Figura 1. Aminoácidos titulables. Aminoácidos con cadena lateral con propiedades ácido-base.

2.2. Determinación del pKa mediante el trazado de curvas de titulación

Experimentalmente, el pKa de un par ácido-base se determina mediante una titulación volumétrica y el trazado de la curva de volumen vs pH correspondiente. De la definición de la constante de acidez se deriva la ecuación de Henderson-Hasselbalch (HH) para un ácido HA y su base conjugada A^- (ec. 1), para dicha ecuación se cumple que cuando la concentración del ácido y su base son iguales, el pH es igual al pKa.

$$pKa = pH - \log \frac{[A^-]}{[HA]} \quad (1)$$

La ecuación de HH también se puede reescribir en función de la fracción de protonación f_p :

$$pKa = pH + \log \left(\frac{f_p}{1 - f_p} \right) \quad (2)$$

En proteínas, comúnmente, las curvas de titulación se desvían de la forma de HH; por lo que se recurre a la ecuación de Hill. Hill dedujo esta ecuación para determinar el

índice de cooperatividad entre sitios de unión de una enzima y puede ser extrapolada a la unión del protón (ecuación (3)). Utilizando la ecuación de Hill se considera en el cálculo del pKa la interacción entre los sitios titulables.

$$\log\left(\frac{f_p}{1-f_p}\right) = n_H(pKa - pH) \quad (3)$$

En la ecuación (3), n_H es el coeficiente de Hill que refleja el grado de cooperatividad entre los sitios titulables. Cuando n_H es menor que 1, hay cooperatividad negativa, mientras que cuando es mayor que 1, hay cooperatividad positiva. La cooperatividad positiva consiste en que la unión del protón a un sitio facilita la unión de otro protón a otro sitio mediante un aumento en las afinidades de los sitios disponibles. Cuando la unión produce una disminución de afinidad de los sitios no ocupados, se habla de cooperatividad negativa.^{3,4}

A continuación, se extiende la ecuación (3):

$$\log\left(\frac{f_p}{1-f_p}\right) = n_H pKa - n_H pH \quad (4)$$

Mediante el trazado de la gráfica de Hill, representando pH vs $\log(f_p/(1-f_p))$, y llevando a cabo la regresión lineal correspondiente, se obtiene una pendiente $-n_H$, y una ordenada al origen $n_H pKa$. La ecuación (4) también se puede representar de la siguiente manera:

$$f_p = \frac{1}{1 + 10^{-n_H(pKa-pH)}} \quad (5)$$

2.3. Simulación Molecular

La simulación molecular se lleva a cabo con la intención de entender las propiedades de conjuntos de moléculas en términos de su estructura y las interacciones microscópicas entre ellas. Esto sirve como un complemento a los experimentos convencionales que nos permite explicar fenómenos que no se podrían explicar por otras vías. Las dos principales familias de técnicas de simulación son la Dinámica Molecular (DM)

y Monte Carlo (MC) (ver Apéndice A); de las cuales se deriva toda una gama de técnicas híbridas.⁵

La simulación actúa como un puente entre las escalas de longitud y tiempo microscópicas y el mundo macroscópico del laboratorio: mediante ellas se proporciona una estimación de las interacciones entre las moléculas y se obtienen predicciones “exactas” de las propiedades de bulto (Figura 2). También, las simulaciones conectan la teoría con el experimento; se puede probar una teoría realizando una simulación usando el mismo modelo. También se pueden llevar a cabo simulaciones de experimentos, que son difíciles o imposibles en el laboratorio, por ejemplo, a temperaturas o presiones extremas.⁵

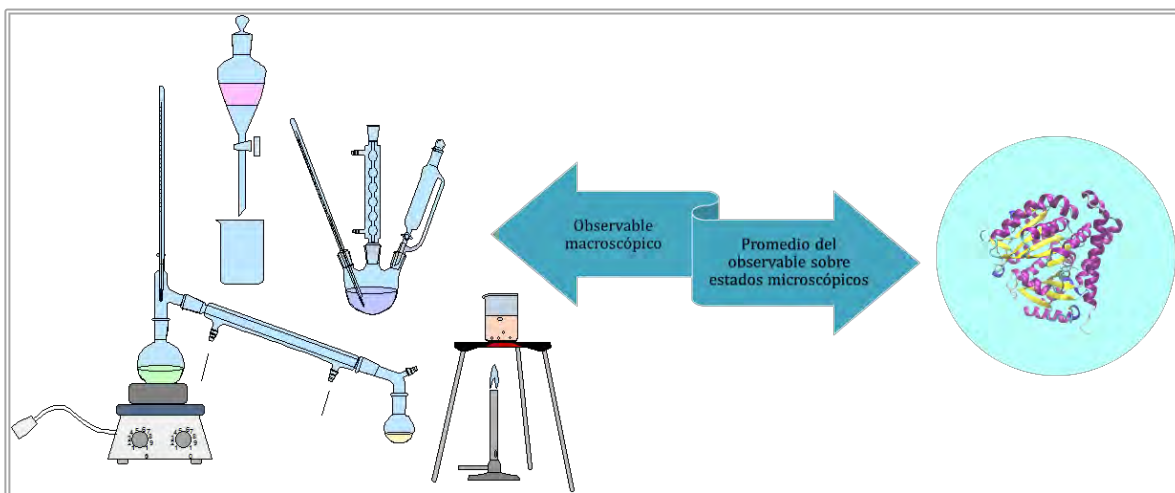


Figura 2. La simulación conecta el mundo microscópico con el mundo macroscópico.

2.3.1. Mecánica molecular y campos de fuerza

Muchos de los sistemas de interés para el modelado molecular son demasiado grandes para ser estudiados mediante mecánica cuántica. Mientras que la mecánica cuántica se ocupa de los electrones en un sistema, la mecánica molecular (MM) ignora el movimiento electrónico y calcula la energía de un sistema sólo en función de las posiciones nucleares; es así como se simplifican significativamente los cálculos y permite trabajar con sistemas con un gran número de átomos. Sin embargo, debido a esta aproximación, la MM no es capaz de estudiar propiedades que dependen de la distribución electrónica en una molécula.⁶

El funcionamiento de la MM se debe absolutamente a la validez de varias suposiciones; entre las cuales la principal es la aproximación de Born-Oppenheimer, sin la cual no sería posible escribir la energía como función de las coordenadas nucleares. La mecánica molecular se basa en un conjunto de modelos simples de las interacciones dentro de un sistema con contribuciones de procesos como el estiramiento de enlaces, la apertura y cierre de ángulos y las rotaciones sobre enlaces.⁶

La simulación de DM consiste en la solución numérica, paso por paso, de las ecuaciones de movimiento de Newton, ecuación (6). Para llevar a cabo este cálculo, es necesario estimar el conjunto de fuerzas \mathbf{f}_i que actúan sobre cada uno de átomos del sistema de interés. Este conjunto de fuerzas se resume en los campos de fuerza.

$$m_i \ddot{\mathbf{r}}_i = \mathbf{f}_i \quad \mathbf{f}_i = -\frac{\partial}{\partial \mathbf{r}_i} U \quad (6)$$

Un campo de fuerza es una expresión matemática que describe la dependencia de la energía de un sistema con respecto a las coordenadas de sus partículas. Consiste en una forma analítica de la energía potencial interatómica, $U(\mathbf{r}^N)$, en la que $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ representa el conjunto completo de $3N$ coordenadas atómicas, y un conjunto de parámetros que la conforman. Los parámetros se obtienen típicamente a partir de cálculos cuánticos *ab initio* o semiempíricos, o por ajuste a datos experimentales tales como difracción de neutrones, rayos X y electrones, resonancia magnética nuclear (RMN), espectroscopía infrarroja, Raman y de neutrones, etc.⁷

Los campos de fuerza más comunes dividen sus términos en cuatro contribuciones esenciales de fuerzas intra- e intermoleculares (ver Figura 3). Las contribuciones intramoleculares corresponden a la desviación de enlaces, ángulos y torsiones dihedrales de sus valores de referencia, mientras que las contribuciones intermoleculares, o no enlazantes, incluyen a las interacciones coulómbicas y de van der Waals. Estos cuatro componentes se representan en la ecuación (7),

$$\begin{aligned}
U(\mathbf{r}^N) = & \sum_{\text{enlaces}} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{\text{ángulos}} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 \\
& + \sum_{\text{torsiones}} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \\
& + \sum_{i=1}^N \sum_{j=i+1}^N \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right)
\end{aligned} \tag{7}$$

$U(\mathbf{r}^N)$ denota la energía potencial que es función de las posiciones (\mathbf{r}) de N partículas. El primer término de la ecuación modela la interacción entre pares de átomos enlazados, en este caso es modelado por una potencial armónico que da el incremento en energía potencial cuando la longitud de enlace l_i se desvía del valor de referencia $l_{i,0}$. El segundo término es una suma sobre todos los ángulos del tipo A-B-C, también modelado utilizando un potencial armónico. El tercer término en la ec. (7) es un potencial torsional que modela el cambio en la energía por efecto de la rotación de enlaces. La cuarta contribución describe las interacciones intermoleculares o de no enlace; este término es calculado para todos los pares de átomos (i y j) que están en diferentes moléculas o que están en la misma molécula, pero separados por al menos tres enlaces. En un campo de fuerza simple el término de no enlace es generalmente modelado por un potencial coulombico para interacciones electrostáticas y un potencial de Lennard-Jones para las interacciones de van der Waals.⁶

En cualquier molécula que contenga más de cuatro átomos en una hilera, es necesario incluir un término diedro o torsional. Por su parte, el doblamiento de ángulos, y en particular el estiramiento de enlaces, son movimientos de alta frecuencia que a menudo no son relevantes para el estudio de las propiedades de interés y pueden ser reemplazados por una aproximación rígida. En cambio, los movimientos torsionales son típicamente cientos de veces menos rígidos y son necesarios para asegurar el grado correcto de rigidez de la molécula y para reproducir los cambios conformacionales mayores debidos a rotaciones sobre enlaces. Por lo tanto, juegan un papel crucial en la determinación de la estructura local de una macromolécula o la estabilidad relativa de diferentes conformaciones moleculares.⁷

Las interacciones de Van der Waals entre dos átomos surgen del equilibrio entre fuerzas repulsivas y atractivas. La repulsión se debe a la superposición de las nubes electrónicas de ambos átomos, mientras que las interacciones entre los dipolos inducidos resultan en un componente atractivo que varía como $r^{-6,7}$

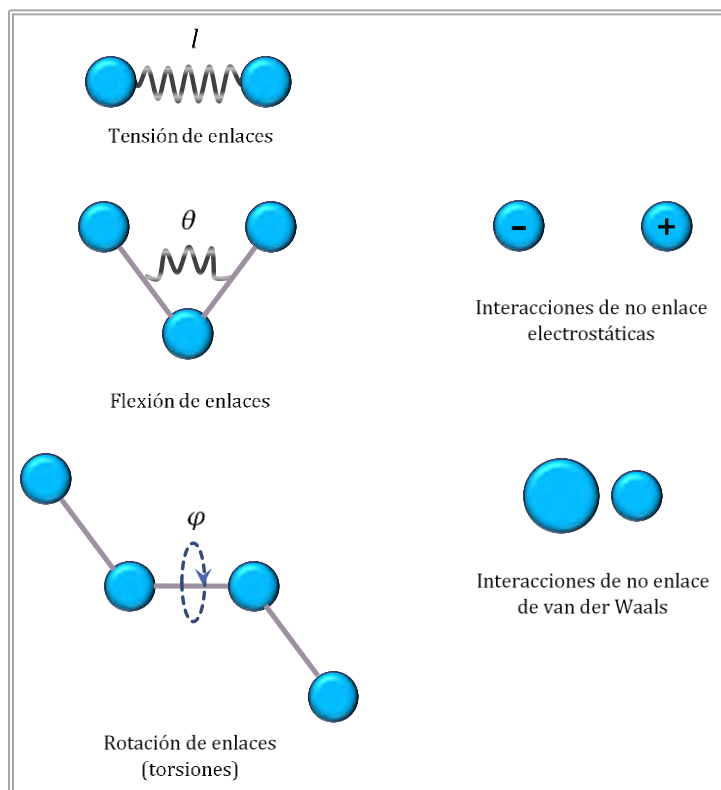


Figura 3. Representación esquemática de los cuatro componentes clave de un campo de fuerza de mecánica molecular.

2.3.2. Algoritmo de Dinámica molecular

Como se ha mencionado anteriormente, la simulación de DM es una técnica que produce una trayectoria dinámica para un sistema compuesto de N partículas por integración de las ecuaciones de movimiento de Newton (ec. (6)). Se necesita un conjunto de condiciones iniciales (posición y velocidad para cada partícula), un buen modelo para representar las fuerzas que actúan entre las partículas y definir las condiciones de frontera a emplear. Dadas las condiciones iniciales, se resuelve el sistema de ecuacio-

nes (6), que es un sistema de N ecuaciones diferenciales acopladas no lineales de segundo orden que no tiene solución exacta, por lo que tiene que ser resuelto numéricamente, paso a paso (8), utilizando un algoritmo de integración apropiado.⁷

$$\mathbf{r}_i(t_0) \rightarrow \mathbf{r}_i(t_0 + \Delta t) \rightarrow \mathbf{r}_i(t_0 + 2\Delta t) \rightarrow \dots \mathbf{r}_i(t_0 + n\Delta t) \quad (8)$$

Las propiedades deseables en un integrador son:^{5,7-9}

- Calcular con poca frecuencia las fuerzas, ya que es un cálculo muy costoso.
- Estabilidad para pasos de tiempo cortos y largos; ya que las ecuaciones son “rígiditas”, el integrador debe ser capaz de hacer frente a escalas de tiempo cortas y largas.
- Calcular con exactitud las propiedades dinámicas, especialmente en los tiempos cercanos a los tiempos de correlación de las propiedades de interés.
- Permanecer en la hipersuperficie de energía constante durante tiempos largos, con el fin de llevar a cabo un muestreo correcto del ensamble. Conservar energía e impulso.
- Conservar el volumen de espacio de fase.

Todas estas propiedades tienden a favorecer al algoritmo de Verlet (ver Apéndice A) de una forma u otra.

2.4. Simulaciones de Dinámica Molecular a pH constante

En las simulaciones habituales de dinámica molecular se emplea un conjunto de estados de protonación fijo; sin embargo, esta aproximación sufre de varias limitaciones. Primero, para elegir un estado “representativo” de un sitio titulable es necesario conocer el valor de su pKa; los sitios pueden exhibir valores de pKa atípicos, el pKa de un grupo en una proteína puede diferir por varias unidades de pH del aminoácido correspondiente en solución. Cuando el pKa de un grupo es cercano al pH del disolvente, no es adecuado elegir sólo un estado de protonación como representativo, puesto que en tales condiciones el equilibrio entre los dos estados no favorece significativamente a alguno. Otra limitación es que incluso si es posible determinar un estado de protonación representativo o predominante, esta consideración desprejaría los efectos entrópicos asociados con las fluctuaciones vacío/ocupado de los sitios, cuya contribución al

equilibrio de protonación puede ser significativa.¹⁰ Quizá la limitación más importante de la aproximación de estados de protonación constantes es que desacopla la dependencia dinámica del pKa y el estado de protonación sobre la conformación.¹¹

Los métodos de Dinámica Molecular a pH constante (*CpHMD*, por sus siglas en inglés) fueron desarrollados con el fin de correlacionar la conformación de la proteína con su estado de protonación.¹² Las simulaciones de *CpHMD* son simulaciones de Dinámica Molecular en las que el pH es uno de los parámetros termodinámicos externos, como la temperatura y la presión. El enfoque habitual de la DM es adecuado para abordar los aspectos conformacionales, pero es incapaz de tratar los aspectos de protonación de una manera realista.¹¹

Se han desarrollado diferentes enfoques para tratar los eventos de protonación. El enfoque más fundamental consiste en incluir en el Hamiltoniano la asociación y disociación de protones, esta aproximación requiere el uso de métodos cuánticos y semi-clásicos que resultan útiles para estudiar la transferencia del protón, pero no el equilibrio de los sitios titulables con la disolución acuosa. Otra ruta para tratar los efectos del pH en una simulación es considerar sólo los estados de protonación discretos de interés y calcular la diferencia de energías libres. Esto se puede hacer con métodos de cálculo de energías libres basados en MM. Este enfoque sólo es factible para estudiar moléculas pequeñas ya que una molécula con n sitios titulables tiene 2^n estados de protonación. Alternativamente, se pueden usar Hamiltonianos más simplificados, orientados a la electrostática, como los que representan el disolvente como una rejilla de dipolos de Langevin o como un continuo dieléctrico descrito por la teoría de electrostática de continuo (*CE*).¹¹ Las energías obtenidas por estos medios se pueden usar para muestreo Monte Carlo u otros métodos aproximados; sin embargo, surge otro inconveniente, el soluto se trata como rígido, por lo que la reorganización estructural debida a la carga es ignorada. En contraste con la DM, el uso de Hamiltonianos simplificados es adecuado para muestrear estados de protonación, pero no reflejan los aspectos conformacionales de una manera realista; por ello, el desarrollo de metodologías que permitan el muestreo simultáneo de protonación y conformación se ha enfocado en la combinación de la DM y los Hamiltonianos simplificados.

Se han propuesto varios modelos para llevar a cabo DM a pH constante con estados de protonación dinámicos. Los primeros emplearon parámetros de protonación que permitieran variar los estados de protonación de forma continua;^{1,13} sin embargo, un modelo continuo de estados de protonación no es equivalente al muestreo de estados discretos, especialmente en el caso de grupos titulables fuertemente acoplados, donde las curvas de titulación son altamente no lineales. El modelo continuo, puede conducir a estimaciones inapropiadas de observables físicos. Es por ello que el desarrollo de modelos para *CpHMD* se ha orientado hacia la implicación de estados de protonación discretos.¹⁴

En 1997, Baptista *et al.*¹ propusieron un método de titulación implícita para simulaciones de DM a pH constante en el que las conformaciones y los estados de protonación se muestrean simultáneamente durante la simulación. El método se basó en un potencial de fuerza media (*PMF*) que utiliza estados de protonación fraccionales obtenidos con un enfoque *CE*. La razón de ser de esta propuesta fue tomar ventaja de las metodologías y restringirlas a lo que pueden hacer mejor; es decir, MM/DM para el muestreo conformacional en un estado de protonación fijo, y *CE* para muestrear estados de protonación con una conformación del soluto fija e intentar combinarlos en un estado de protonación fijo. Sin embargo, dos aspectos del método no fueron del todo satisfactorios; el primero fue la necesidad práctica de actualizar el *PMF* a intervalos de tiempo periódicos y no en todos los pasos de tiempo de la simulación, y el segundo fue la suposición de que la probabilidad de un estado de protonación particular depende principalmente de la conformación de la proteína, descuidando su dependencia de la configuración del disolvente. Debido a esto, en 2002, Baptista *et al.*¹¹ propusieron un método estocástico para simulaciones de *CpHMD* denominado *titulación estocástica*, en el que estas suposiciones no fueran necesarias.

En el Apéndice A se aborda la teoría sobre la que se soporta la titulación estocástica.

2.4.1. Algoritmo de CpHMD con solvatación implícita GB

El método propuesto, emplea DM con solvatación *GB*, con un muestreo periódico de estados de protonación con MC. Entre los pasos de MC, el sistema evoluciona como una simulación de DM estándar. El esquema de muestreo y su justificación son los mismos que describen Baptista *et al.* (ver Apéndice A), con excepción de que no hay relajación del disolvente porque la simulación es conducida en disolvente implícito.

Para el muestreo MC de estados de protonación de un residuo titulable a un valor de pH dado, la energía libre de transición es calculada de acuerdo con:

$$\Delta G_{MC} = k_B T (pH - pK_{a,ref}) \ln 10 + \Delta G_{elec} - \Delta G_{elec,ref} \quad (9)$$

donde k_B es la constante de Boltzmann, T es la temperatura, $pK_{a,ref}$ es el pK_a del compuesto de referencia apropiado, ΔG_{elec} es el componente electrostático de la energía libre calculado para el grupo titulable en la proteína, y $\Delta G_{elec,ref}$ es el componente electrostático de la energía libre de transición del compuesto de referencia. Esta ecuación se basa en la separación de la energía libre de transición en sus componentes electrostático y no electrostático. La energía libre de transición no electrostática comprende todas las contribuciones de energía libre no tomadas en cuenta en la electrostática *GB*, incluyendo la energía libre mecanocuántica de enlace y la energía libre de solvatación del protón. Debido a la dificultad para realizar el cálculo del componente no electrostático, se aproxima que este tiene el mismo valor independientemente del ambiente electrostático. Bajo esta aproximación, se introduce un compuesto de referencia con pK_a conocido para cancelar la porción no electrostática de la energía libre de transición. ΔG_{elec} es la diferencia entre los potenciales calculados con las respectivas cargas para los estados protonado y desprotonado.

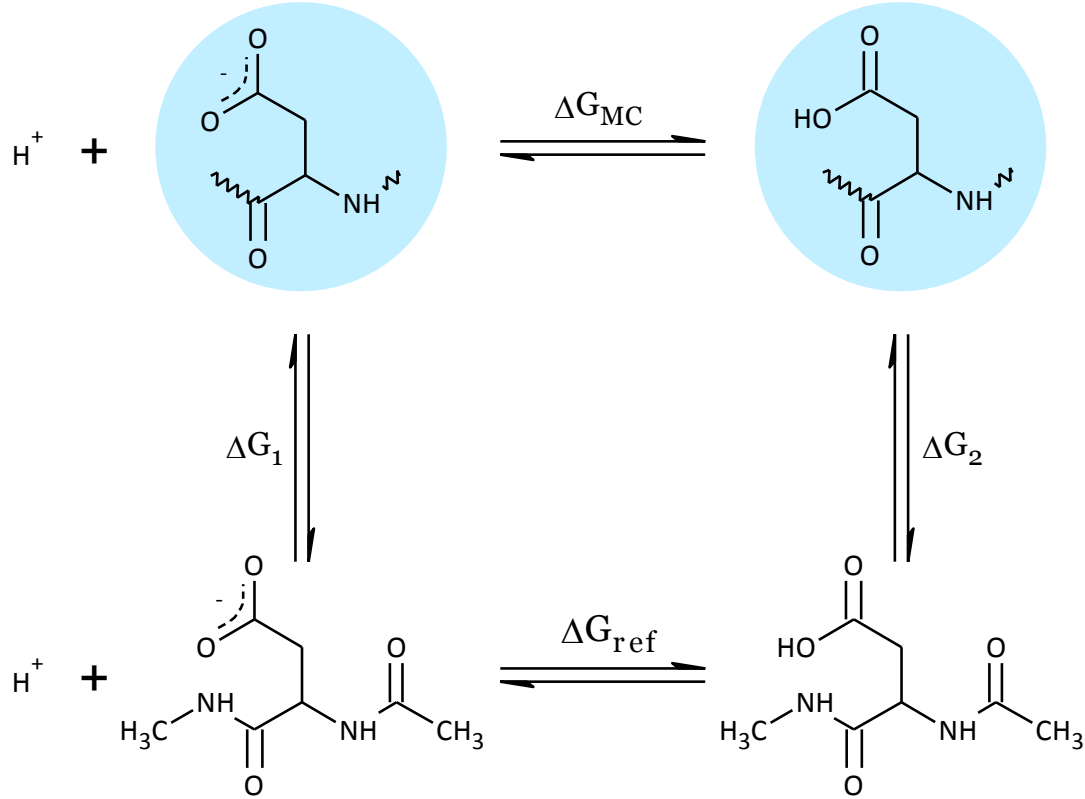


Figura 4. Ciclo termodinámico utilizado en simulaciones de *CpHMD*. ΔG_{MC} es la energía utilizada como base para aplicar el criterio de Metrópolis.

De la ecuación (9), para la transición del estado desprotonado a protonado, se puede asignar ΔG_{ref} y desglosar ΔG_{elec} , en consistencia con la Figura 4, de la siguiente forma:

$$\Delta G_{ref} = k_B T (pH - pK_{a,ref}) \ln 10 - \Delta G_{elec,ref} \quad (10)$$

$$\Delta G_{elec} = \Delta G_2 - \Delta G_1 \quad (11)$$

donde ΔG_1 y ΔG_2 son las diferencias de potenciales del aminoácido en la proteína con respecto al de referencia en los estados desprotonado y protonado respectivamente. La ecuación (9), se puede escribir como:

$$\Delta G_{MC} = \Delta G_{ref} + \Delta G_2 - \Delta G_1 \quad (12)$$

La energía total de transición ΔG_{MC} es usada como la base para aplicar el criterio de Metropolis¹⁵ (ver Apéndice A) para determinar si la transición será aceptada. Si es aceptada, la simulación continúa con el grupo titulable en el nuevo estado de protonación; si no, con el estado de protonación no cambiado. La probabilidad de transición $w(d \rightarrow p)$ está definida por

$$w(d \rightarrow p) = \begin{cases} 1, & \text{para } \Delta G_{MC} \leq 0 \\ \exp(-\beta \Delta G_{MC}), & \text{para } \Delta G_{MC} > 0 \end{cases} \quad (13)$$

En la implementación de este método en AMBER, para cada residuo que se titula, existe un 25% de probabilidad de que se produzca una denominada titulación multisitio con un residuo vecino; es decir, el cambio propuesto implicará cambios en el estado de protonación tanto del residuo como de su vecino. Dos residuos titulables se consideran "vecinos" si dos átomos de hidrógeno titulantes están a 2 Å o menos uno del otro. La inclusión de saltos de estado de protonación de multisitio es importante para sistemas que tienen residuos titulables que interactúan estrechamente. Sin estos movimientos multisitios, las transferencias de protones entre residuos adyacentes titulables implicados en un puente de hidrógeno nunca ocurrirían debido a la gran penalización de interrumpir la interacción añadiendo otro protón o eliminando el protón implicado en el puente de hidrógeno.

Swalis *et al.*¹⁶ se basaron en la definición del sistema y el ensamble de Baptista *et al.*¹¹, en el que las coordenadas del sistema se pueden separar en grados de libertad de soluto y disolvente, implementaron el tratamiento del disolvente explícitamente en los pasos de DM de *CpHMD* en AMBER. En la Figura 5 se muestra el diagrama de flujo del proceso de *CpHMD* de dicha implementación. Estos autores demostraron que la solvatación explícita ayuda a aliviar el problema de muestreo de configuraciones no físicas, que puede tener impacto en la precisión de los valores de pKa predichos cuando el disolvente es modelado implícitamente durante las simulaciones de *CpHMD*.

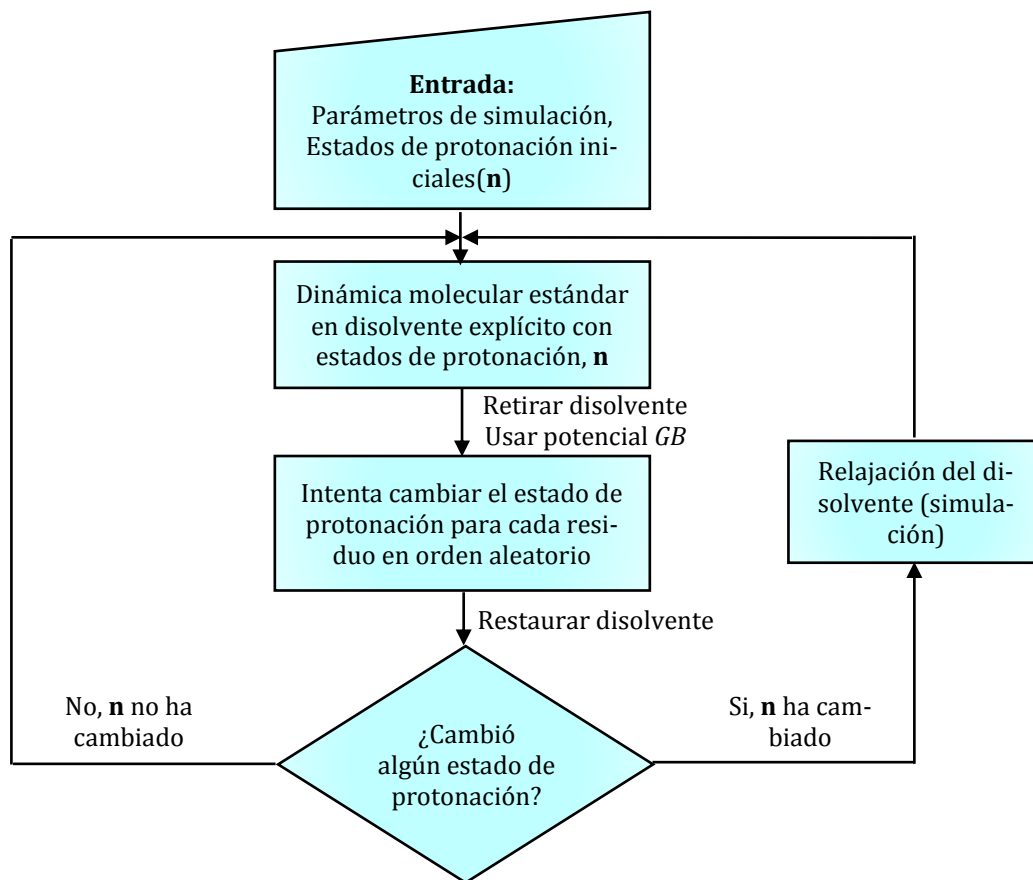


Figura 5. Diagrama de flujo para simulaciones a pH constante en disolvente explícito.

a) Intercambio de réplicas en el espacio pH

A lo largo del desarrollo de metodologías para el modelado preciso del espacio de protonación, ha sido ampliamente discutida la necesidad de un mejor muestreo del espacio conformacional. Es por ello que se han propuesto muchos métodos teóricos para superar las barreras de energía libre en el espacio conformacional; entre ellos, están bien establecidos los métodos de ensamble generalizado, tales como la DM de intercambio de réplicas (*REMD*), el templado simulado (*ST*), el conjunto multicanónico, la metadinámica y la caminata aleatoria en el espacio ortogonal (*OSRW*). En algunos de estos métodos, además del espacio conformacional, el sistema realiza una caminata aleatoria en el espacio de energía o de temperaturas, reduciendo barreras conformacionales.¹⁷

Para explicar el acoplamiento de la protonación y el muestreo conformacional, Khandogin *et al.*¹⁸ combinaron la Dinámica Molecular a pH constante con protonación

continúa empleando el método *REMD* (REX-CpHMD). En 2010, Meng y Roiberg¹² utilizaron un método híbrido combinando *REMD* de temperatura (*T-REMD*, en el cual las réplicas corren a diferentes temperaturas) y *CpHMD* con protonación discreta. Posteriormente, este mismo grupo¹⁷ desarrolló un método de *pH-REMD* en el cual se intentan intercambios de réplicas vía intercambio de Hamiltonianos donde las coordenadas atómicas son intercambiadas periódicamente.

En 2012, Swails y Roitberg¹⁹ evaluaron la eficiencia del método *pH-REMD* propuesto por Itoh *et al.*²⁰ en el cual se lleva a cabo un intercambio periódico entre los pHs de la disolución de las réplicas (ver Figura 6), lo que permite intercambios entre réplicas más eficientes que su variante que emplea *H-REMD*.¹⁷ La evaluación demostró que este método mejora significativamente el muestreo en comparación con *CpHMD*¹⁴ en el estudio realizado para la lisozima, sistema comúnmente utilizado como punto de referencia para los cálculos de pKa. Las curvas de titulación generadas con *pH-REMD* contienen menos ruido y convergen más rápido que las obtenidas por *CpHMD*, lo que sugiere que el *pH-REMD* es una herramienta poderosa para llevar a cabo predicciones de pKa. A continuación, se presenta una breve explicación de la teoría en la que se basa el método de Itoh *et al.*²⁰

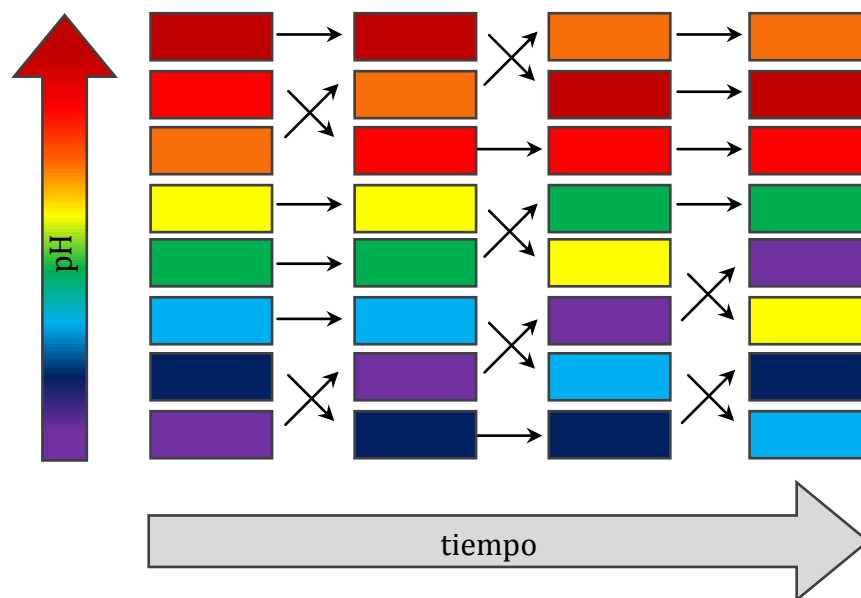


Figura 6. Esquema de intercambio de réplicas que muestra 8 réplicas a diferentes pHs. Los bloques representan las trayectorias del DM, mientras que las flechas representan los intercambios entre las réplicas.

i. *Método de pH-REMD*

Consideramos M réplicas no interactuantes de un sistema, donde la réplica i ($i = 1, \dots, M$) tiene un vector de coordenadas q_i y un vector de momento p_i a una temperatura T y un valor de pH de pH_l ($l = 1, \dots, M$) en el ensamble gran canónico de protones. Además, el sistema tiene N^t residuos titulables y N_i^P se protonan en la réplica i . Los intercambios de valores de pH entre réplicas se llevan a cabo de modo que se satisface la condición del balance detallado para todos los intercambios. Para un intercambio entre la réplica i a pH_l y la réplica j a pH_m , la condición del balance detallado está definida por

$$P(X_i^l)P(X_j^m)w(X_i^l, X_j^m \rightarrow X_i^m, X_j^l) = P(X_i^m)P(X_j^l)w(X_i^m, X_j^l \rightarrow X_i^l, X_j^m), \quad (14)$$

donde P es la probabilidad de equilibrio, $X_i^l \equiv (q_i, p_i, N_i^P, pH_l)$, y w es la probabilidad de transición. En el ensamble gran canónico de protones, la probabilidad de equilibrio P con una energía cinética K y una energía potencial V a la temperatura T está dada por

$$P(X_i^l) = \frac{1}{\Xi} \exp\left(-\beta(K(p_i, N_i^P) + V(q_i, N_j^P) + k_B T pH_l \ln 10 N_i^P)\right), \quad (15)$$

donde Ξ es la función de partición gran canónica. De las ecuaciones (14) y (15), la razón de las probabilidades de transición es calculada de

$$\frac{w(X_i^l, X_j^m \rightarrow X_i^m, X_j^l)}{w(X_i^m, X_j^l \rightarrow X_i^l, X_j^m)} = \exp(-\Delta), \quad (16)$$

y

$$\Delta \equiv \ln 10 (pH_m - pH_l)(N_i^P - N_j^P) \quad (17)$$

y la probabilidad de intercambio se obtiene mediante:

$$w(X_i^l, X_j^m \rightarrow X_i^m, X_j^l) = \begin{cases} 1, & \text{para } \Delta \leq 0 \\ \exp(-\Delta), & \text{para } \Delta > 0 \end{cases} \quad (18)$$

Obsérvese que la relación de aceptación de los intercambios de pH entre réplicas disminuye exponencialmente con la diferencia de los dos valores de pH según se ve en estas ecuaciones.

2.5. La enfermedad de Alzheimer

En 1907, Alois Alzheimer describió dos alteraciones patológicas en el cerebro de una paciente que padecía de demencia. Estas dos lesiones representan las características patognomónicas de la enfermedad, y su observación durante el examen *post mortem* sigue siendo necesaria para el diagnóstico definitivo de la enfermedad de Alzheimer (EA). Alzheimer describió una "sustancia peculiar" que existe como depósitos extracelulares en regiones específicas del cerebro, que ahora se conocen como placas amiloides. No fue hasta mediados de la década de los 80s que se descubrió que estas placas se componen de agregados de un pequeño péptido llamado β -amiloide ($A\beta$). La segunda alteración, conocida como ovillos neurofibrilares (*NFT*), consiste en agregados de la proteína tau en la región intraneuronal, los cuales se forman a partir de un aumento en el nivel de fosforilación (hiperfosforilación) de los aminoácidos involucrados en la interacción normal con los microtúbulos.²¹

Los péptidos $A\beta$ se forman por la escisión secuencial de la proteína precursora amiloide (*APP*) por dos proteasas: β - y γ -secretasa. Primero, la β -secretasa escinde a la *APP* (en su sitio escindible β) removiendo el ectodominio, reduciéndola al fragmento C-terminal con 99 aminoácidos (C99). Posteriormente, la enzima γ -secretasa, mediante su componente proteolítico presenilina, corta a C99 dentro de su dominio transmembranal, produciendo el fragmento *ACID* (dominio intracelular de la *APP*) y liberando al péptido $A\beta$ de la membrana celular (ver Figura 7).^{22,23}

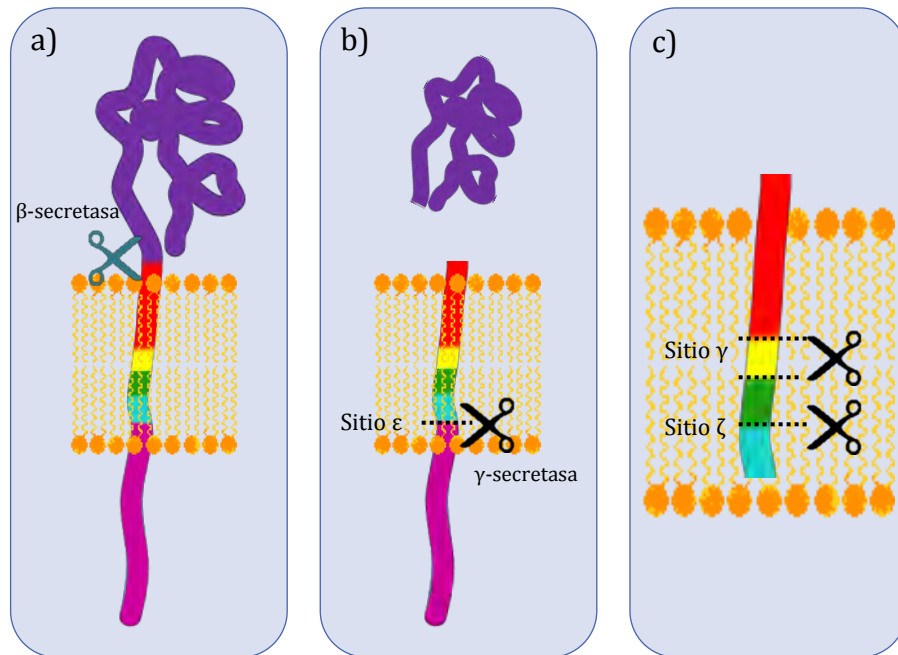


Figura 7. Escisión secuencial de APP por a) β-secretasa (✂) en el sitio escindible β, liberando el ectodominio y por b) γ-secretasa (✂) en el sitio ε produciendo el fragmento AICD y c) en los sitios ζ y γ para producir los diferentes péptidos Aβ.

La variación natural en el proceso de escisión del C99 por la γ-secretasa conduce a la producción de diferentes isoformas del péptido Aβ de diferentes longitudes. La isoforma predominante en pacientes sanos es Aβ40, la cual se compone de 40 aminoácidos y no se considera tóxica; mientras que la segunda isoforma predominante es Aβ42, que tiene 42 aminoácidos, siendo la más propensa a agregarse de manera patológica.^{22,23}

Se ha identificado la escisión secuencial de tres sitios diferentes del C99 implicados en la formación de los péptidos Aβ, denominados sitios de escisión ε, ζ y γ (ver Figura 7). La primera escisión se produce en el sitio ε, entre la Leu49 y la Val50 para generar Aβ49 y liberar el fragmento AICD. El producto Aβ49 se escinde en el sitio ζ entre Val46 y Ile47 para producir Aβ46 y un tripéptido. Finalmente, el péptido Aβ46 se escinde en el sitio γ para producir Aβ43 o Aβ40, que liberan uno y dos tripéptidos respectivamente. Sin embargo, los diferentes patrones de escisión de C99 también pueden conducir a la formación de productos Aβ38 y Aβ42.²³

La cantidad total de Aβ42, así como la proporción de Aβ42/Aβ40, se cree que son mediadores clave de la patogénesis de Aβ, una hipótesis fuertemente apoyada por el

hecho de que casi todas de las más de 200 mutaciones autosómicas dominantes de la enfermedad de Alzheimer familiar (*FAD*) en las presenilinas 1 y 2 y la *APP* aumentan la relación $A\beta_{42}/A\beta_{40}$ (ver www.alzforum.org/mutations).²² Debido a esta evidencia, se sugiere que la inhibición de la producción de todas las isoformas de $A\beta$ o su modulación específica son opciones viables para el tratamiento de la enfermedad de Alzheimer.^{21,23,24}

2.5.1. Proteasas intramembranales

Las proteasas intramembranales (*IMPRs*) escinden otras proteínas transmembranales dentro de la bicapa lipídica. Estas enzimas cumplen importantes papeles biológicos y se encuentran en casi todos los rincones de los sistemas de membrana biológica de las células eucariotas.

El agua necesaria para la proteólisis es escasa en las membranas lipídicas, por lo que inicialmente la noción de proteólisis intramembranal fue recibida con escepticismo. Hoy día, se sabe que los residuos catalíticos de las *IMPRs* están inmersos profundamente dentro de la membrana lipídica, ya que se han resuelto las estructuras de miembros representativos de las cuatro familias de *IMPRs*; sin embargo, el mecanismo de reconocimiento y escisión del sustrato es aún poco comprendido.²⁵

2.5.2. Estructura de la γ -secretasa y su sitio activo.

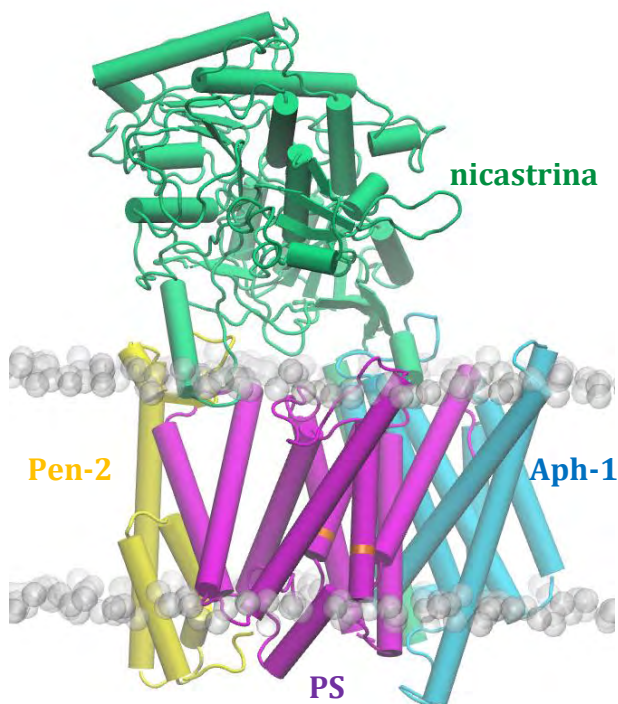


Figura 8. Complejo de la γ -secretasa. Nicastrin(verde), Pen-2 (amarillo), PS(violeta) y Aph-1 (cian).

(TMs 7-9) respectivamente. La nicastrina es el componente más grande de la γ -secretasa, tiene sólo una hélice transmembranal y contiene un dominio extracelular grande y altamente glicosilado (**ECD**), el cual se ha propuesto como esencial para el reconocimiento del sustrato, y para la maduración y ensamblaje de la GS. La Pen-2 es el componente más pequeño, la cual se encarga de estabilizar el complejo y facilitar la maduración autocatalítica de la presenilina y la actividad de la γ -secretasa. La Aph-1 es la subunidad requerida para el autoensamblaje de la γ -secretasa por su interacción con la PS y la hélice TM de la nicastrina.^{23,26}

El sitio activo de la γ -secretasa se encuentra en la subunidad catalítica presenilina y contiene dos aspárticos catalíticos localizados en los motivos de proteasa intramembranal VTD₂₅₇ en TM6 y LGLGD₃₈₅ en TM7. La identificación de estos aspárticos como catalíticos y, por lo tanto, de la actividad de proteasa de la GS, se consiguió mediante estudios de mutagénesis dirigida. Estos estudios mostraron que la mutación de cualquiera

La γ -secretasa (**GS**) es un complejo proteico transmembranal que consta de cuatro componentes o subunidades: presenilina (**PS**), *presenilin enhancer 2* (**Pen-2**), nicastrina y *anterior pharynx-defective 1* (**Aph-1**) (ver Figura 8). La presenilina actúa como la subunidad catalítica; contiene nueve hélices transmembranales (**TMs**), con dos aspárticos catalíticos en TM6 y TM7. Durante el ensamblaje de la γ -secretasa, la presenilina sufre una escisión autocatalítica para producir los fragmentos amino y carboxilo terminal, **NTF** (TMs 1-6) y **CTF**

de estos aspárticos (D257 o D385), inhibe la autoescisión de la PS y, por tanto, la actividad de la GS.²⁷

Las proteínas transmembranales son difíciles de cristalizar debido a su naturaleza anfifílica. Relativamente pocas proteínas de membrana han sido estudiadas en su totalidad por cristalografía de rayos X, y la determinación de su estructura tridimensional se lleva a cabo por enfoques alternativos, como la microscopía electrónica, análisis de difracción de electrones, en conjunto con otras técnicas como RMN, técnicas de secuenciación y clonación de ADN y del modelado por homología.²⁸

A pesar de que la γ -secretasa obtuvo relevancia desde el reconocimiento de su participación en la formación de los péptidos $A\beta$, su estructura tridimensional se logró determinar por criomicroscopía electrónica en el año 2015.²⁹⁻³¹ La ausencia de información tridimensional ralentizó por varios años el estudio de esta enzima; incluso ahora, aunque se ha obtenido mayor información estructural, no se tiene resolución comparable con la de la cristalografía de rayos X y su estudio a nivel atómico continúa siendo un reto.

2.5.3. Determinación del estado de protonación de la díada catalítica de las aspartil proteasas

Para conseguir un modelo del sitio activo de GS y proponer el mecanismo de reacción de la enzima que pueda servir como guía para el diseño de fármacos para detener o controlar el progreso de la enfermedad de Alzheimer (EA) es importante: (1) identificar los estados de protonación de los aspárticos catalíticos, así como (2) localizar la posición preferencial de sus protones.

En principio, los estados de protonación de los residuos ionizables embebidos en una proteína se pueden sondear directamente localizando las coordenadas de los átomos de hidrógeno en experimentos de difracción. Por ejemplo, Coates *et al.* han aplicado la difracción de neutrones y rayos X en la endotiapepsina, también una aspartil proteasa, para determinar los estados de protonación de la díada catalítica y para estudiar su mecanismo de acción.^{32,33} Sin embargo, la resolución necesaria para poder determinar las coordenadas de los átomos de hidrógeno en las estructuras cristalográficas

debe ser muy alta. Esta, ha sido una limitación incluso para proteínas solubles, para el caso de una proteína tal como la GS, que está embebida en la membrana, la dificultad de cristalizarla hace nula esta opción.

El estado de protonación y la configuración de la díada catalítica de la familia de las aspartil proteasas han sido ampliamente estudiados, en la Figura 9 se representan las configuraciones posibles. Una de las enzimas más extensamente estudiadas de esta familia es la proteasa VIH-1, que cristaliza como un homodímero con una díada catalítica simétrica. En el 2000, Piana *et al.*³⁴ llevaron a cabo un estudio usando simulaciones de dinámica molecular *ab initio*, en el que propusieron una configuración monoprotónada en la que los aspárticos comparten el átomo de hidrógeno forman un puente de hidrógeno de barrera baja (*LBHB*). En un estudio posterior, propusieron un estado diprotónado cuando el inhibidor pepstatina A se encuentra unido a la VIH-1.³⁵

Friedman y Caflisch³⁶ llevaron a cabo múltiples simulaciones de DM con disolvente explícito de la plasmepsina II libre e inhibida, a partir de las cuales obtuvieron que la protonación de Asp214 favorece una estructura estable del sitio activo y describieron que el estado de protonación de los aspárticos catalíticos tienen una fuerte influencia sobre una cadena lineal de puentes de hidrógeno con las cadenas laterales adyacentes.

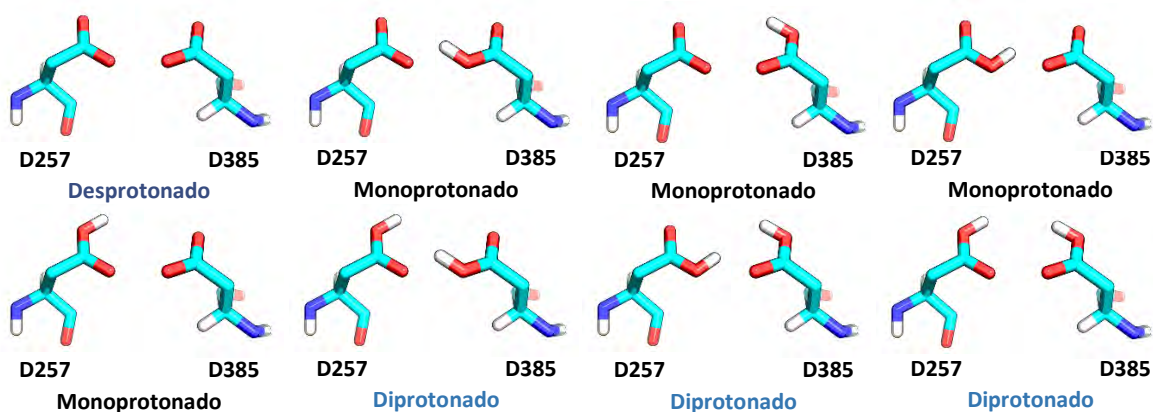


Figura 9. Configuraciones posibles del estado de protonación de la díada catalítica.

a) Estudios sobre la asignación de estados de protonación de la GS

Los estados de protonación y el mecanismo de escisión de la β -secretasa (BACE-1) han sido ampliamente estudiados y se ha obtenido mucha información de ello (ver Apéndice A). Sin duda, el soporte del estudio de la BACE-1 son sus estructuras cristalográficas, por lo que la GS representa un reto aún mayor. Que la GS sea una proteína transmembranal no sólo complica su estudio experimental, sino que la membrana tampoco puede ser ignorada en el estudio computacional.

Se ha reportado un estudio computacional del mecanismo de escisión del C99 por la presenilina,³⁷ en el cual, en ausencia de estructuras cristalográficas de la PS, la BACE-1 fue elegida como un modelo. Aunque el mecanismo general de las aspartil proteasas intramembranales y no intramembranales pueda ser el mismo, las interacciones son dependientes de la secuencia, la estructura 3D y el ambiente electrostático; y es el estudio de estas interacciones el que puede dar respuestas acerca de la formación de los péptidos $A\beta$.

b) Estudios recientes de las aspartil proteasas utilizando metodologías de DM a pH constante

En secciones anteriores se han planteado las múltiples ventajas de utilizar la simulación de Dinámica Molecular a pH constante, en las que quizá, la más importante, es que se mantiene la correlación entre los estados de protonación y la conformación. Esta correlación no ha sido considerada (al menos en su totalidad) en la mayoría de los métodos con los que se ha abordado el estudio de la asignación de estados de protonación en las aspartil proteasas.

La VIH-1 y la BACE-1 han sido estudiadas por esta técnica, ambas en combinación con la técnica de intercambio de réplicas para ampliar el espacio conformacional.

McGee *et al.*³⁸ utilizaron pH-REMD para simular la VIH-1 libre y la acoplada con 12 inhibidores distintos, para determinar los pKas de la díada Asp en cada caso. Este estudio provee evidencia de que los aspárticos catalíticos están primariamente en una configuración monoprotónada, en forma libre y unida, en el intervalo de pH en el que

ocurren la mayoría de los experimentos y simulaciones. Los pKas determinados para la díada catalítica de la enzima libre fueron 1.29 y 7.32.

Ellis y Shen³⁹ emplearon la metodología de DM a pH constante, con un modelo de estados de protonación continuo, para elucidar el mecanismo conformacional dependiente del pH que regula la actividad de la BACE-1; ellos determinaron pKas de 4.1 y 1.9 para Asp32 y Asp228 respectivamente.

La díada catalítica de la VIH-1 es simétrica, por lo que sus aspárticos son indistinguibles y McGee *et al.* calcularon pKas generales de la díada y representan, por lo tanto, un pKa de configuración diprotonada→monoprotonada y otro de monoprotonada→desprotonada. Para la BACE-1, Ellis y Shen calcularon los pKas de los aspárticos de manera independiente.

2.6. Proteínas intrínsecamente desordenadas

La introducción de proteínas biológicamente activas, pero esencialmente no estructuradas desafió uno de los, hasta entonces, pilares fundamentales de la biología de proteínas, el paradigma de estructura-función. Este concepto afirma que una función específica de una proteína está determinada por su única y rígida estructura tridimensional. Esta idea, fue formulada por Emil Fischer en 1894 como un modelo de llave-cerradura para explicar la especificidad de la hidrólisis enzimática de glucósidos, y fue reforzada por estructuras cristalográficas en las que el sitio activo y el sustrato se encuentran rígidos.⁴⁰ Claramente, la consideración de una proteína como una entidad cristalina rígida es una simplificación excesiva, ya que incluso las proteínas más estables y bien plegadas son sistemas dinámicos que poseen diferentes grados de flexibilidad conformacional.⁴¹

A pesar de sus limitaciones, el modelo de llave-cerradura fue responsable de la creación de la ciencia de proteínas moderna. Sin embargo, la acumulación de evidencia experimental de la existencia de una gran cantidad de dominios proteicos y proteínas enteras carentes de estructura ordenada en condiciones fisiológicas justificaron la reevaluación del paradigma estructura-función.⁴⁰

Hay que tener en cuenta que una gran parte de las estructuras de proteínas que se depositan en el *Protein Data Bank* (PDB) tienen porciones de su secuencia ausentes en las estructuras determinadas (regiones con densidad electrónica perdida) debido a la incapacidad del átomo, cadena lateral, residuo o región no observada de dispersar los rayos X coherentemente, causadas por su naturaleza flexible o desordenada.^{41,42}

Además de las proteínas ordenadas, que poseen regiones desordenadas de longitud variable, la literatura contiene numerosos ejemplos de proteínas biológicamente activas con estructuras flexibles. Por lo tanto, hay otra clase de proteínas y regiones proteicas funcionales que contienen fragmentos más o menos dinámicos. Se caracterizan incluso por una completa o casi completa falta de estructura ordenada en condiciones fisiológicas (al menos *in vitro*), que parece ser un aspecto crítico para la función *in vivo* de estas proteínas. Estas proteínas y regiones proteicas intrínsecamente desordenadas (*IDPs* e *IDRs*, respectivamente), no tienen una estructura definida al equilibrio y existen como conjuntos heterogéneos de confórmeros, de tal manera que ningún conjunto único de coordenadas o ángulos de Ramachandran es suficiente para describir sus propiedades conformacionales.

2.6.1. Funciones y características generales

Históricamente, las *IDRs* fueron pensadas como uniones flexibles entre las regiones más ordenadas, pero estudios más recientes han revelado que pueden mejorar la función de la proteína, interactuar con un gran número de proteínas y ligandos, y permitir que un gran número de modificaciones postraduccionales tomen lugar. Las *IDRs* también pueden contribuir a la regulación de la función de la proteína, permitiendo la exposición transitoria de pequeños grupos de aminoácidos al agua y otras moléculas, equilibrando así los comportamientos dicotómicos de proteínas y poblando conformaciones altamente heterogéneas que se interconvierten rápidamente.⁴³

Debido a la gran complejidad de las *IDRs* e *IDPs*, Van der Lee *et al.*⁴⁴ realizaron un interesante trabajo en el que sintetizan y proporcionan una visión general de las diversas clasificaciones de regiones y proteínas intrínsecamente desordenadas que se han presentado en la literatura desde el inicio de estudios sistemáticos sobre su función.

Discuten enfoques basados en función, elementos funcionales, estructura, secuencia, interacciones de proteínas, evolución, regulación y propiedades biofísicas.

Las secuencias de las *IDPs* son muy diferentes a las de las proteínas ordenadas. Por ejemplo, se ha señalado que la combinación del bajo contenido de residuos hidrofóbicos y una alta carga neta representa un rasgo característico de algunas *IDPs* (denominadas nativamente desplegadas, con estructuras en forma de serpentina) (Figura 10b).⁴⁵ Por lo tanto, las proteínas compactas y las *IDPs* extendidas se pueden distinguir con base en su carga neta y su hidrofobicidad.⁴⁵ La mayoría de las *IDPs* son polianfolitos, la fracción de residuos cargados discrimina entre polianfolitos débiles y fuertes, de manera que los anfolitos débiles forman glóbulos. Las preferencias estructurales de los polianfolitos fuertes se determinan por una combinación de las fracciones de residuos cargados y las distribuciones de secuencia lineal de residuos opuestamente cargados (Figura 10a).⁴⁶

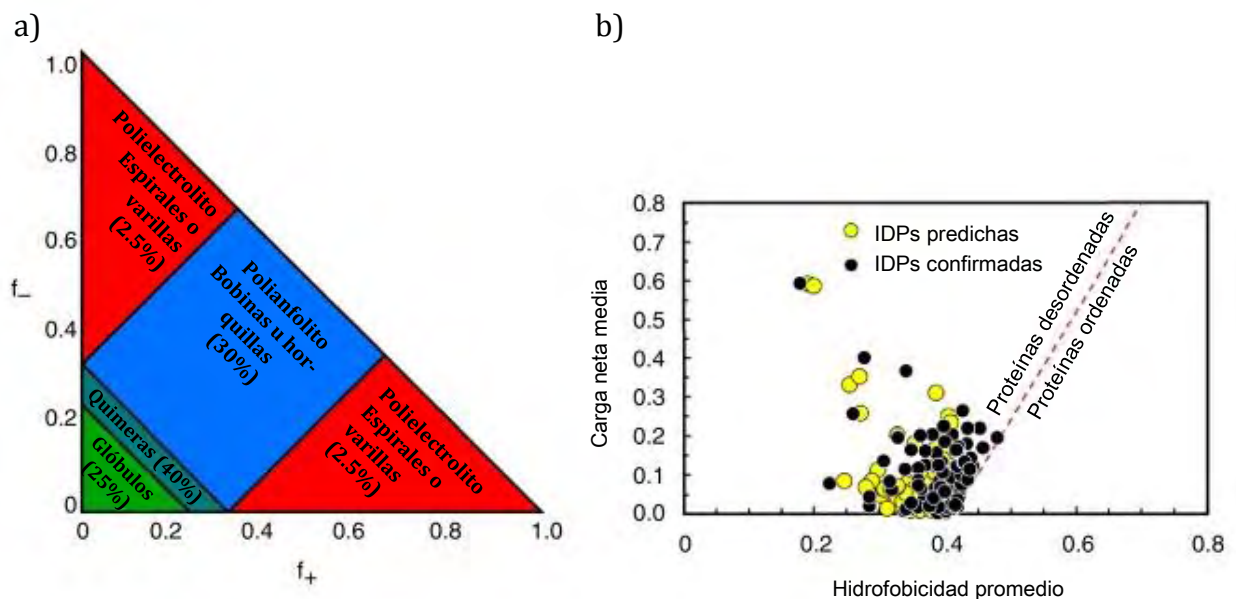


Figura 10. a) Las *IDPs* pueden clasificarse por la cantidad de aminoácidos cargados positivamente (f_+) y negativamente (f_-) que contienen, lo que puede ayudar a predecir sus comportamientos. b) La mayoría de las proteínas desordenadas tienden a presentar cargas netas relativamente altas y muy poca hidrofobicidad. Figura adaptada de la referencia ⁴³.

2.6.2. Colaboración entre estudios experimentales y computacionales

Las técnicas experimentales más comunes utilizadas para caracterizar las *IDPs* implican el uso de medidas *SAXS* (dispersión de rayos X de ángulo pequeño) y *FRET* (transferencia de energía de resonancia de Förster) o RMN. Sin embargo, mediante estas técnicas se obtienen promedios; es por ello que, para la caracterización de los ensambles conformacionales transitorios de las *IDPs*, los estudios experimentales se apoyan en el modelado computacional.^{41,43,44,47}

Para el estudio de la estructura de las *IDPs* se utilizan diversas herramientas computacionales que incorporan enfoques basados en el conocimiento. Alternativas emergentes a estos enfoques incluyen enfoques *de novo*, que combinan campos de fuerza empíricos con simulaciones de Dinámica Molecular o de Monte Carlo. Al igual que con los métodos basados en el conocimiento, las estructuras *de novo* se validan mediante cálculos retroactivos a observables experimentales. Sin embargo, una ventaja neta es que estos métodos generan conjuntos de Boltzmann y otros parámetros dinámicos que pueden correlacionarse con las mediciones experimentales.^{43,44}

En el caso de la simulación molecular, hay dos cuestiones críticas a considerar; la precisión de los campos de fuerza que se utilizan y la calidad del muestreo. Se requieren protocolos de muestreo mejorados para obtener el ensamble conformacional de una *IDP*. Esto es importante porque los cambios conformacionales de las *IDPs* pueden ocurrir en escalas de tiempo de hasta milisegundos, imposibles de muestrear en simulaciones de DM convencionales. Es por ello que, protocolos como *Replica-Exchange Molecular Dynamics (REMD)* y *Metadynamics* han sido particularmente exitosos en este sentido, y se han utilizado para caracterizar las estructuras monoméricas de múltiples *IDPs*.⁴³

La optimización del campo de fuerza sigue siendo un reto y un campo activo de investigación. Si bien muchos modelos existentes de proteínas y agua son adecuados para las simulaciones de proteínas globulares a temperatura ambiente, donde los residuos hidrofóbicos están protegidos del agua; en cambio, las simulaciones de *IDPs* extendidas requieren mayor precisión en los efectos de solvatación, ya que la exposición

al agua es común. Los campos de fuerza de proteínas que siguen siendo parametrizados con modelos de agua antiguos resultan en *IDPs* compactas e hidrofóbicas que sobrestiman las interacciones proteína-proteína comparadas con el experimento.^{43,47}

2.6.3. Una proteína asociada al microtúbulo: Tau

Los microtúbulos, componentes clave del citoesqueleto, son polímeros proteicos largos, filamentosos y tubulares, esenciales para todas las células eucariotas. Son cruciales para el desarrollo y el mantenimiento de la forma celular, para el transporte de vesículas, mitocondrias y otros componentes a través de las células, en la señalización y la división celular. Los microtúbulos se componen de heterodímeros de α - y β -tubulina dispuestos en forma de finos tubos filamentosos que pueden tener varios micrómetros de longitud.⁴⁸

La dinámica de los microtúbulos en las neuronas es modulada por varias proteínas llamadas proteínas asociadas a los microtúbulos.⁴⁹ Una importante proteína asociada a los microtúbulos es la proteína tau, que promueve la formación de microtúbulos axonales, los estabiliza para impulsar el crecimiento neurítico.

2.6.4. La hiperfosforilación y agregación de tau

La proteína tau es una fosfoproteína que normalmente contiene de 2 a 3 moles de fosfato por mol de proteína. En 1986, se descubrió que en el cerebro de pacientes que padecían la enfermedad de Alzheimer se encuentra tau 3 veces más fosforilada que lo normal.⁵⁰⁻⁵² Esta proteína tau hiperfosforilada, se coagrega con la proteína tau normal, generando filamentos helicoidales emparejados insolubles (PHFs). La agregación de la proteína tau en estos filamentos, que forman los ovillos neurofibrilares, conduce a la destabilización de los microtúbulos, provocando su despolimerización.⁵³

Se han identificado dos motivos hexapeptídicos en la proteína tau, PHF6 y PHF6*, que probablemente median la interacción intermolecular entre moléculas de tau para formar oligómeros asociados con el proceso de agregación.⁵⁴

Larini *et al.*⁵⁵ investigaron el autoensamblaje del fragmento peptídico más amiloideogénico de la proteína tau, $_{273}\text{GKVQIINKKLDL}_{284}$ (R2/wt), empleando una combinación de simulaciones y experimentos, con énfasis en caracterizar los primeros pasos de su agregación (Figura 11 y 12). Este péptido contiene la región PHF6* (VQIINK). Ellos emplearon el método T-REMD para caracterizar el amplio ensamble conformacional de este péptido. De su estudio, surgieron dos familias estructurales primarias: una familia de conformaciones extendidas y una familia de conformaciones compactas. La familia de conformaciones compactas incluyó estructuras estabilizadas mediante enlaces de hidrógeno (principalmente horquillas y, en menor medida, hélices) o por puentes salinos entre K274 y D283, más frecuentes entre los pares de residuos adyacentes K280 y D283 o K281 y D283. Las conformaciones compactas por puentes de hidrógeno se estabilizan por razones entálpicas; sin embargo, los ensambles de puente salino se estabilizan tanto por contribuciones energéticas como entrópicas, ya que son posibles múltiples conformaciones del *backbone* para cualquier puente salino (Figura 11). La competencia entre la formación de puentes salinos y de puentes de hidrógeno es la razón por la que este péptido es intrínsecamente desordenado, incapaz de encontrar una estructura única. Las conformaciones estabilizadas por puentes de hidrógeno se interrumpen continuamente por la formación de puentes salinos y viceversa.^{55,56}

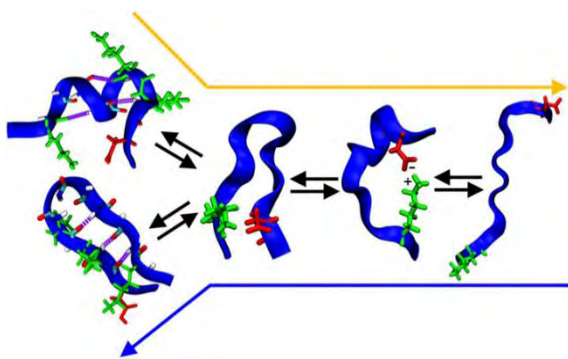


Figura 11. Las conformaciones más probables de R2/wt dependen de las condiciones ambientales. Figura adaptada de la referencia ⁵⁶.

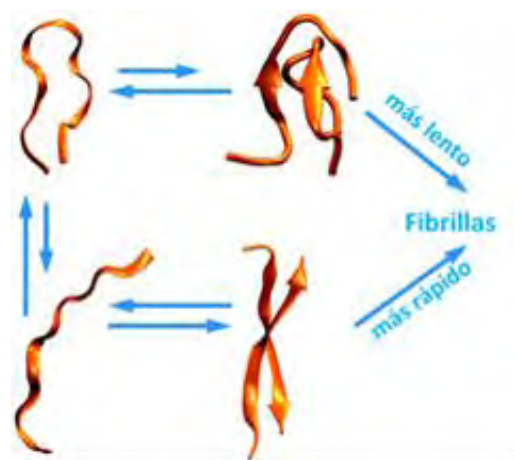


Figura 12. Esquema de la vía de agregación de R2/wt propuesta por Larini *et al.*⁵⁵

Se ha propuesto que el monómero adopta principalmente una conformación compacta, que es energéticamente más favorable debido a la presencia de enlaces de hidrógeno y puentes salinos y que, en estas condiciones, la agregación se ve obstaculizada debido a la necesidad de romper y reformar estos enlaces de hidrógeno y puentes salinos (Figura 12).

3. Objetivos

1) Determinar el estado de protonación de la díada catalítica de γ -secretasa, por el método de muestreo simultáneo de conformación y estados de protonación, pH-*REMD*.

2) Caracterizar el ensamble conformacional de un fragmento de la proteína tau (R2/wt), empleando el pH como una variable del sistema que permita potenciar el muestreo de su ensamble conformacional.

4. Hipótesis

○ Al alcanzar el equilibrio en la conformación activa de la GS, la titulación de los aminoácidos aspárticos catalíticos estará acoplada debido a su cercanía y su estrecha interacción.

○ Si se permite el cambio en los estados de protonación del fragmento amiloidogénico de tau; entonces se identificará la transición entre distintas posibles conformaciones del péptido R2/wt, debido a la ruptura y formación de puentes salinos.

○ Si las conformaciones del péptido R2/wt dependen de los estados de protonación de aminoácidos con propiedades ácido-base; entonces, el intercambio de réplicas de distintos pHs permitirá obtener un ensamble conformacional más completo y adecuado.

5. PROCEDIMIENTO

5.1. Detalles generales

Todas las simulaciones se realizaron en AMBER 16⁵⁷. Los sistemas se parametrizaron utilizando el campo de fuerza ff14SB⁵⁸. El modelo *GB* propuesto por Onufriev *et al.*⁵⁹ (*igb* =2 para los programas de AMBER) con concentración de sal 0.1 M (basada en Debye-Hückel) fue usado para las simulaciones con solvatación implícita y para la evaluación de los intentos de cambio de estados de protonación. Las longitudes de enlace que incluyen hidrógeno se restringieron usando el algoritmo *SHAKE*⁶⁰. El paso de tiempo para todas las simulaciones fue de 2 fs. Las simulaciones a pH constante se llevaron a cabo a 300 K, ya que los cálculos para los compuestos de referencia se realizaron a esta temperatura.

El método *CpHMD* implementado en AMBER fue desarrollado por Mongan *et al.*¹⁴ y corre las simulaciones con el modelo de solvatación implícita *GB* y Swails *et al.*¹⁶ para simulaciones con disolvente explícito. Este método emplea transiciones de Monte Carlo entre estados de protonación discretos y a intervalos periódicos durante la simulación. Para realizar las simulaciones con solvatación explícita, se sigue el diagrama de flujo de la Figura 5. El tiempo de relajación del disolvente τ_{rlx} empleado para todas las simulaciones es de 200 fs, que de acuerdo con Swails *et al.* es suficiente para alcanzar la relajación del disolvente.

5.2. Preparación del sistema para simulaciones *CpHMD*

AMBER provee definiciones para titular las cadenas laterales de ASP, GLU, HIS, LYS, TYR y CYS. Si se desea titular ASP, GLU o HIS, es necesario reemplazar los nombres de los aminoácidos en el archivo PDB por AS4, GL4 e HIP. Esto asegura que en el archivo de topología habrá un hidrógeno definido en cada posición de protonación posible.

Todas las topologías fueron preparadas empleando el módulo de *tLeap* y leyendo el archivo de comandos *leaprc.constph* que carga el campo de fuerza de AMBER para proteínas ff14SB y agrega las librerías especiales de los residuos carboxílicos y las modificaciones al campo de fuerza que definen un átomo de hidrógeno en cada posición

protonable (*syn*- y *anti*- para ambos oxígenos) junto con torsiones impropias para evitar que se giren entre sí. También ajusta los radios de solvatación *GB* a *mbondi2*, que corresponde al modelo empleado para la parametrización de los compuestos de referencia.

5.3. Compuestos de referencia

Los valores pKa absolutos en disolución son difíciles de calcular; son imposibles usando campos de fuerza clásicos. Como resultado, cada método *CpHMD* basado en física utiliza la idea de un compuesto modelo cuyo pKa experimental es fácil de medir con un alto nivel de precisión. Se añade entonces un parámetro empírico (la energía de referencia) para que *CpHMD* reproduzca los valores experimentales de pKa de estos compuestos.¹⁶ Los compuestos de referencia tienen la secuencia ACE-X-NME.

5.4. Aminoácidos titulables

Para cada sistema, se definen los aminoácidos que se desean titular; se construye un archivo que contiene los estados de protonación posibles y sus energías relativas. El método *CpHMD* implementado en AMBER utiliza estados de protonación discretos. Para los carboxilos, los conjuntos de cargas de los estados de protonación se definen de tal manera, que sólo uno de los cuatro hidrógenos pueda tener una carga distinta de cero a la vez.¹⁴ El radio intrínseco de solvatación de los átomos de oxígeno es reducido de 1.5 a 1.3 Å, para las simulaciones con solvatación explícita, y así compensar el efecto de tener 2 hidrógenos virtuales en cada oxígeno.

5.5. pH-REMD

Swails *et al.*¹⁹ implementaron en AMBER el método de pH-REMD descrito por Itoh *et al.*²⁰. Este método consiste en que un número de réplicas son simuladas a diferentes pHs, y periódicamente emplea un paso de transición de Monte Carlo en el que se intenta un intercambio entre el pH de las réplicas. En AMBER, la simulación de intercambio de réplicas sólo es posible a volumen constante.

Los cambios de estados de protonación fueron intentados cada 10 fs en los sistemas con solvatación implícita y cada 200 fs en los sistemas con solvatación explícita. Los intentos de intercambio de réplicas se llevaron a cabo cada 200 fs.

5.6. Estructura inicial del sistema de la GS

La determinación de los estados de protonación de los aminoácidos aspárticos catalíticos de la GS se realizó a partir de la estructura 5FN2 del PDB (*Protein Data Bank*). Esta estructura fue resuelta con el inhibidor DAPT, el cual estabiliza el complejo proteico manteniéndolo en una conformación cerrada, y muestra una distancia entre los aspárticos catalíticos de 3.9 Å.

Como conformación inicial, se partió de la estructura del complejo proteico con una membrana bilipídica de POPC en una caja ortogonal de agua TIP3P⁶¹ con iones cloruro y sodio 0.15 M, obtenida de 1 μ s de simulación de dinámica molecular de grano grueso de nuestro grupo de investigación. El fragmento extracelular de nicastrina fue removido con *PyMOL*⁶² para reducir el tamaño del sistema, ya que se ha demostrado que no tiene efecto en la dinámica de la PS. Se utilizó *charmmlipid2amber.py* para renombrar los tipos de átomos y los residuos de la membrana y se empleó *pdb4amber* para adecuar el archivo de coordenadas de la proteína (PDB) para ser reconocidos por *tleap*.

Se eligieron como aminoácidos titulables a D257, K265, E273, K380 y D385; el par catalítico y otros aminoácidos con propiedades ácido-base cercanos al sitio activo. Los nombres de residuo del par catalítico fueron reemplazados de ASP a AS4 en el archivo PDB.

Se leyó el archivo de comandos *leaprc.constph* en *tleap* y se cargó el campo de fuerza de lípidos *lipid14*⁶³ para definir la membrana. El espacio vacío, como resultado de la remoción del fragmento extracelular de nicastrina, se llenó con moléculas de agua para generar la topología y las coordenadas finales del sistema.

5.7. Estructura inicial de los sistemas de R2/wt

A partir de la secuencia GKVQIINKKLDL, se construyó el péptido con *tLeap*, amidando y acetilando el C- y N-terminal respectivamente y asignando el aspártico como AS4.

Para la simulación del péptido con solvatación implícita *GB* se generaron la topología y las coordenadas del péptido construido. Se guardaron las coordenadas también en un archivo PDB.

Para la simulación del péptido con solvatación explícita, el péptido fue solvatado con un *buffer* de agua TIP3P de 10 Å alrededor de la proteína en una caja octaédrica truncada. Se agregaron 23 cloruros y 21 sodios en posiciones aleatorias para proveer un ambiente iónico y neutralizar el sistema con los estados de protonación iniciales.

Con *PyMOL* se colocaron dos péptidos tau en forma de “X”, se generaron la topología y las coordenadas iniciales para la simulación con solvatación *GB* y se añadieron 19 sodios para neutralizar el sistema.

En los cuatro sistemas se titularon todos los aminoácidos titulables.

5.8. Minimización

Para todas las simulaciones con solvatación explícita, la energía de la estructura inicial fue minimizada con 1000 pasos del algoritmo de *steepest descent* seguido por 4000 pasos de gradiente conjugado y una restricción en la cadena principal de 10 kcal/mol·Å². Para los dos sistemas de R2/wt el *cutoff* de interacciones de no enlace utilizado fue de 8 Å y de 10 Å para la GS.

La regulación de la temperatura en la simulación se llevó a cabo para todos los sistemas con la dinámica de Langevin.

Todas las simulaciones pH-*REMD* del péptido de tau se llevaron a cabo con 14 réplicas empleando un intervalo de pH de 1-14, separadas por una unidad de pH. La frecuencia de colisión de la dinámica de Langevin aplicada fue de 10 ps⁻¹.

5.9. Procedimiento de preparación de R2/wt con solvatación implícita

Todo el protocolo se llevó a cabo con la solvatación *GB* y un *cutoff* de 1000 Å.

Para los dos sistemas de R2/wt simulados con solvatación implícita, la minimización se llevó a cabo con 1000 pasos de gradiente conjugado. Las estructuras minimizadas se calentaron a volumen constante por 67 ps, variando linealmente de 10 a 300 K con una restricción en la cadena principal de 1 kcal/mol·Å² y una frecuencia de colisión de 5 ps⁻¹; posteriormente la temperatura se mantuvo a 300 K durante 33 ps. El sistema calentado se relajó en un ensamble NVT con una restricción de la cadena principal de 0.1 kcal/mol·Å² y una frecuencia de colisión de la dinámica de Langevin de 10 ps⁻¹.

5.10. Procedimiento de preparación de R2/wt con solvatación explícita

Desde la minimización hasta pH-*REMD* se aplicaron restricciones en las posiciones de la cadena principal de 10 kcal/mol·Å². El *cutoff* de interacciones de no enlace utilizado para todo el protocolo fue de 8 Å. La estructura minimizada fue calentada a volumen constante, variando linealmente de 10 a 300 K durante 400 ps, con una frecuencia de colisión de 5 ps⁻¹. Se equilibró a 300 K en las mismas condiciones por 1 ns. La estructura obtenida, se equilibró en un ensamble NPT, con una frecuencia de colisión de 1 ps⁻¹.

5.11. Procedimiento de preparación de la GS

Los valores de pH elegidos para la simulación de pH-*REMD* de este sistema fueron de 2 a 9 separados por una unidad de pH.

La estructura minimizada se calentó a volumen constante de 0 a 100 K durante 5 ps con una frecuencia de colisión de 1 ps⁻¹, un *cutoff* de 10 Å, y una restricción sobre la cadena principal y los lípidos de 10 kcal/mol·Å². La estructura a 100 K se calentó hasta

300 K durante 100 ps con escalamiento de presión anisotrópica con tiempo de relajación de 2 ps, frecuencia de colisión de Langevin de 1 ps^{-1} y una restricción de $10 \text{ kcal/mol} \cdot \text{Å}^2$ sobre la cadena principal y la membrana.

Se llevó a cabo la relajación de los lípidos de la membrana (se les liberó la restricción) en un ensamble *a*NPT por 500 ps. Posterior a ello, se llevó a cabo la relajación de los lípidos durante 1 ns en condiciones de presión semi-isotrópica y con la tensión superficial de interfase en el plano *xy*. Se liberó a la proteína bajo las mismas condiciones durante 1 ns y se llevaron a cabo 8 equilibrios independientes NVT con los lípidos y la proteína sin restricciones; las estructuras obtenidas fueron las iniciales para pH-*REMD*. Para el conjunto de simulaciones pH-*REMD*, se utilizó una frecuencia de colisión de 5 ps^{-1} en la dinámica de Langevin.

5.12. Análisis

Todo el análisis de las trayectorias (*RMSD*, *RMSF*, puentes de hidrógeno, distancias, radio de giro y número de contactos) se llevó a cabo utilizando *cpptraj*⁶⁴ excepto el análisis de *clustering* que se llevó a cabo con “*gmx_cluster*” de GROMACS 5.1.2⁶⁵. El cálculo del *RMSD* para todos los sistemas se hizo sobre el *backbone* (N, C α ,C). El cálculo del *RMSF* se llevó a cabo por aminoácido. El criterio utilizado para la formación de puentes de hidrógeno es el estándar en *cpptraj*, distancia A-D menor que 3 Å y ángulo A-H-D mayor que 135°. Para el *clustering* se comparó el *backbone* sin tomar en cuenta el C- y N-terminal, con un *cutoff* de *RMSD* entre vecinos para el monómero de 1.4 Å y para el dímero de 2.6 Å. Las series temporales de fracciones protonadas se calcularon utilizando *cphstats* de AmberTools 16. Los promedios acumulativos y por trozos de las fracciones protonadas se calcularon cada nanosegundo y la ventana utilizada para el cálculo de promedios corrientes fue de 5 ns. El criterio de contacto utilizado fue de 3 Å. Los ajustes de las curvas de titulación fueron hechos con *gnuplot* usando el algoritmo de mínimos cuadrados no lineales de Marquardt-Levenberg.

6. RESULTADOS Y ANÁLISIS

6.1. Primer enfoque: Estudio de los estados de protonación de los aspárticos catalíticos de la enzima γ -secretasa.

6.1.1. Simulación de pH-REMD de GS durante 27 ns, en un intervalo de pH de 2 a 9

a) Monitoreo del cambio estructural general de la proteína

El *RMSD* (*Root Mean Square Deviation*) utilizando como referencia las coordenadas iniciales, es un parámetro que permite monitorear el cambio conformacional durante una simulación. En la Figura 13, se representa el *RMSD* del *backbone* peptídico en el que se identifica una gran desviación a tiempos iniciales. Esta tendencia es común debido a la respuesta del sistema a las nuevas condiciones asignadas, en este caso la asignación de pHs. La variación del *RMSD* es dependiente de cada sistema, hay sistemas que llegan a una estructura de equilibrio más rápido que otros. En este caso, pasados los tiempos iniciales, el *RMSD* de las réplicas es creciente con una pendiente pequeña, asociado al muestreo simultáneo de conformación y estados de protonación.

Otra medida cuantitativa habitualmente utilizada en la simulación es el *RMSF* (*Root Mean Square Fluctuation*). Este parámetro permite evaluar la fluctuación o el movimiento por aminoácido durante la dinámica (o bien, el conjunto de átomos deseado). En la Figura 14 se representa el *RMSF* para todos los pHs de muestreo, de los aminoácidos cercanos al sitio activo de la GS. En esta figura se observa que la región carboxilo terminal intracelular de TM6 es la que presenta mayor fluctuación debido a que este segmento se encuentra expuesto al medio acuoso.

Es posible identificar una dependencia entre la desviación estructural y la dinámica de los aminoácidos con respecto al pH, tanto el *RMSD* como el *RMSF* son mayores cuanto mayor es el pH. Hay una desviación en la tendencia del *RMSD* a pH 9, que se aproxima a la de pH 2. En el *RMSF* se tiene la particularidad de que se superpone a pH 2 y 3, habiendo un pequeño salto entre pH 3 y pH 4.

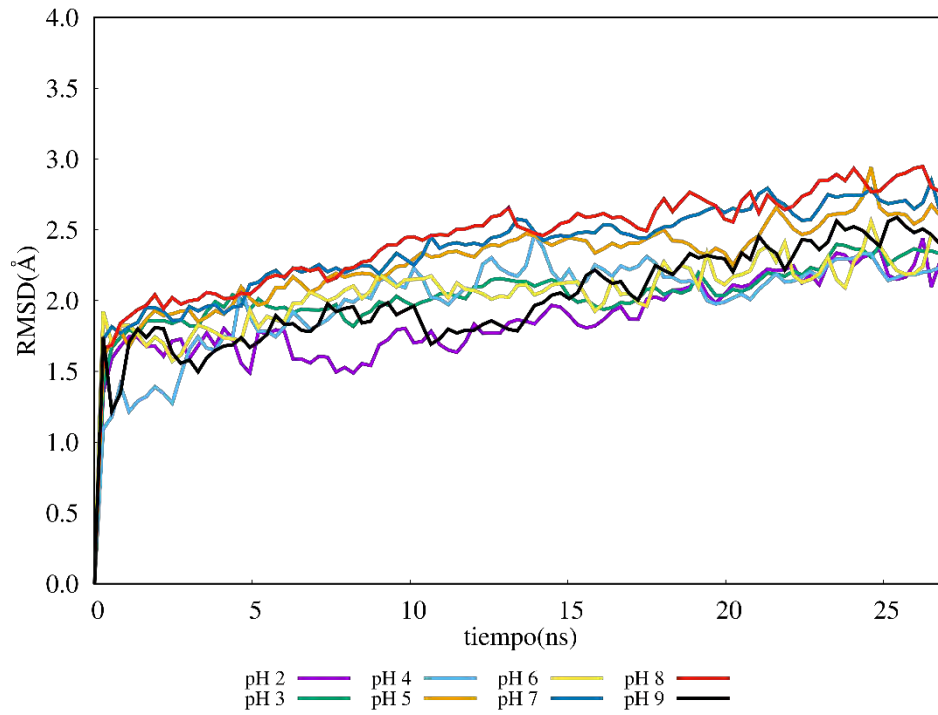


Figura 13. Desviación cuadrática media del *backbone* (N, C α ,C) de GS a lo largo de la trayectoria para cada pH. El cambio estructural tiene una tendencia creciente con el pH, excepto a pH 9(negro), cuyo *RMSD* se aproxima al de la réplica a pH 2.

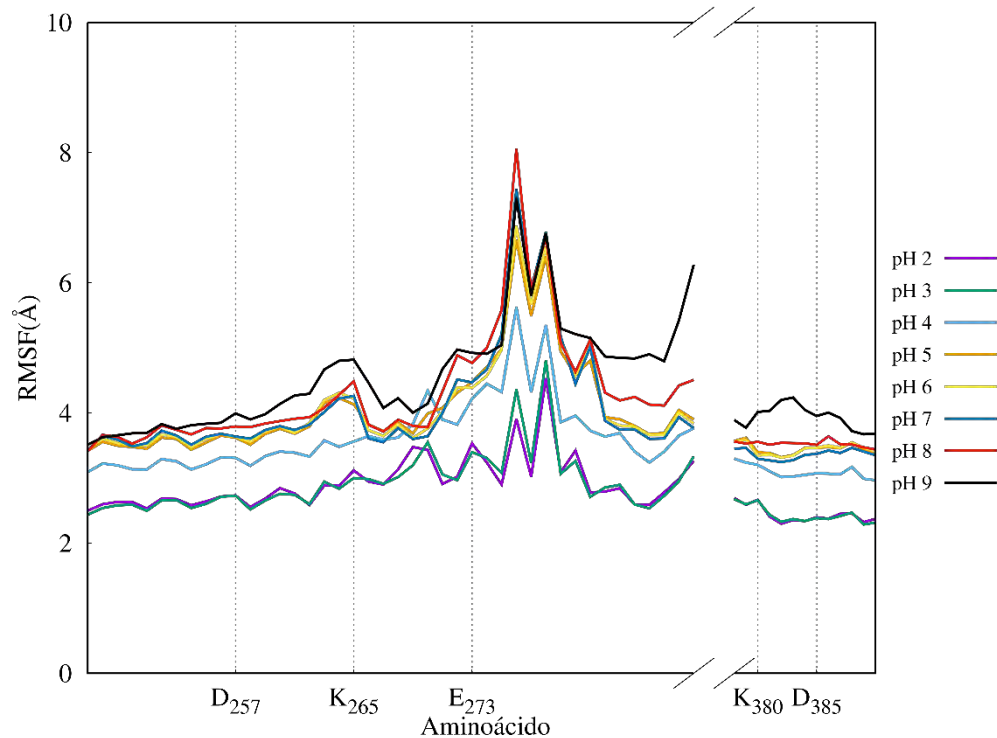


Figura 14. Fluctuación cuadrática media por aminoácido de la GS alrededor de los aminoácidos titulados. Se observa una dependencia directa entre la fluctuación y el pH.

b) Monitoreo de los estados de protonación y cálculo de los pKas

Para el monitoreo de los estados de protonación se calcularon las fracciones protonadas por promedios corrientes (ver Apéndice A). Mediante las series temporales obtenidas, se puede identificar el cambio en los estados de protonación a lo largo de la trayectoria.

En la Figura 15, es posible identificar que el aminoácido D385 sólo se encuentra parcialmente protonado a pH 2, mientras que el D257 se comporta con una acidez relativamente menor. El objetivo de la simulación a pH constante consiste en alcanzar un equilibrio entre la conformación y los estados de protonación. Las fracciones protonadas de ambos aspárticos es creciente al final de la simulación, por lo que se puede observar que no alcanza un equilibrio durante el tiempo simulado; sólo es posible predecir que, con más tiempo de simulación, los aspárticos presentarían una menor acidez. También, a lo largo de la simulación, la basicidad del aminoácido K380 es menor que el

del aminoácido K265, evidenciando el efecto del ambiente inmediato en el comportamiento ácido-base de residuos del mismo tipo.

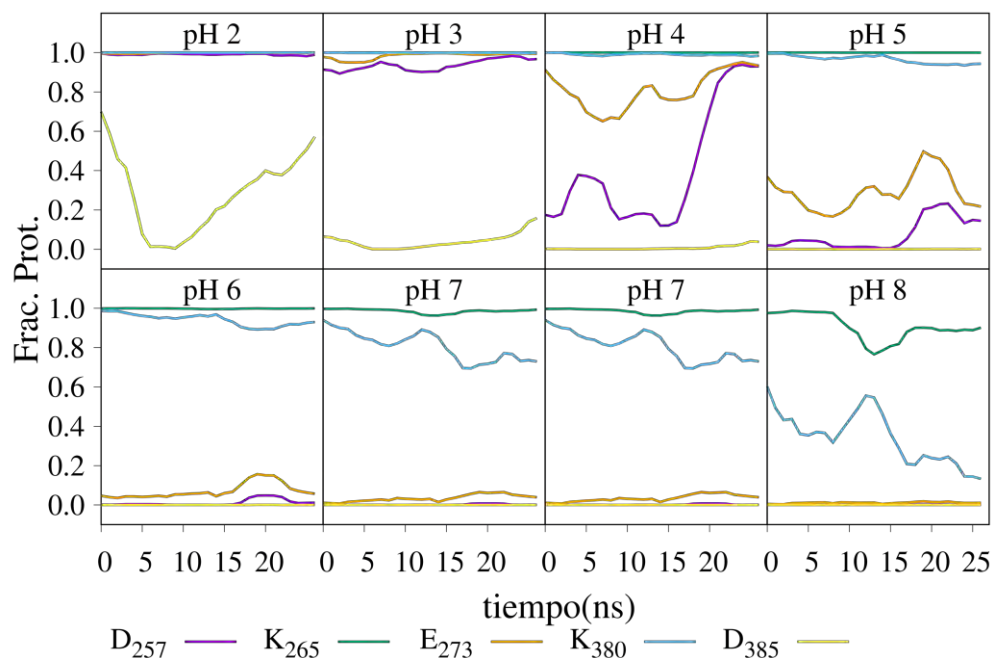


Figura 15. Series temporales de las fracciones de protonación calculadas por promedios corrientes para la GS. La fracción de protonación se calculó cada nanosegundo con un tamaño de ventana de 5 ns. El D385(amarillo) tiene una acidez mayor que el D257(violeta) a lo largo de toda la trayectoria,

El pKa es una constante que permite predecir el estado de protonación representativo de una especie ácido-base a un valor de pH dado y facilita la comparación del comportamiento ácido-base entre especies. La determinación del pKa de las cadenas laterales de aminoácidos se lleva a cabo por ajuste a la ecuación de Hill (ecuación (5)), que considera la cooperatividad entre sitios titulables; que es algo común en proteínas.

Debido a que la fracción protonada del aminoácido D385 sólo es significativa a pH 2, no es posible trazar su curva de titulación y, por lo tanto, su pKa. En la Figura 16, se representa la evolución temporal del pKa y el coeficiente de Hill del aminoácido D257 obtenidos por el ajuste no lineal de la curva de titulación a la ecuación de Hill para cada tiempo. En los últimos nanosegundos de la simulación, el pKa del aminoácido D257 se estabiliza a pH 4.6; sin embargo, las series temporales de las fracciones de protonación indican que se trata de una estabilización aparente. En la Figura 17 se representa la

curva de titulación del aminoácido D257 con la fracción protonada correspondiente para los últimos 5 ns de la simulación.

El coeficiente de Hill durante la mayor parte de la simulación varía entre 1 y 2. Este intervalo es consecuencia de la cooperatividad positiva, es decir, la titulación de otro aminoácido favorece la titulación del D257. Con base en la Figura 15, las posibilidades sólo descartan al aminoácido D385, que es el único de los aminoácidos titulados que no se encuentra protonado a los valores de pH alrededor del pKa determinado.

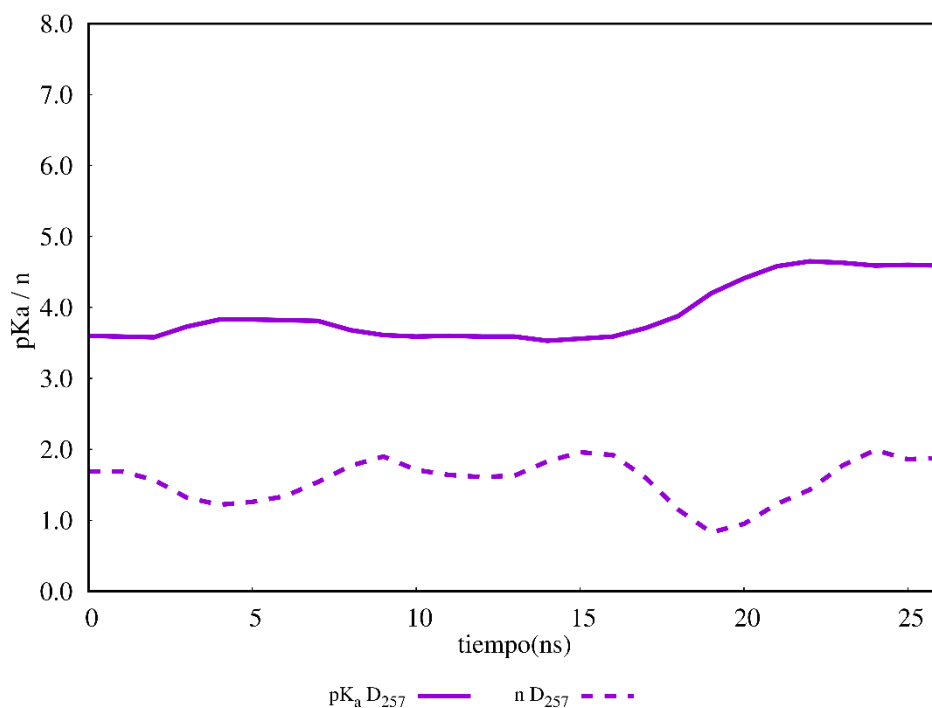
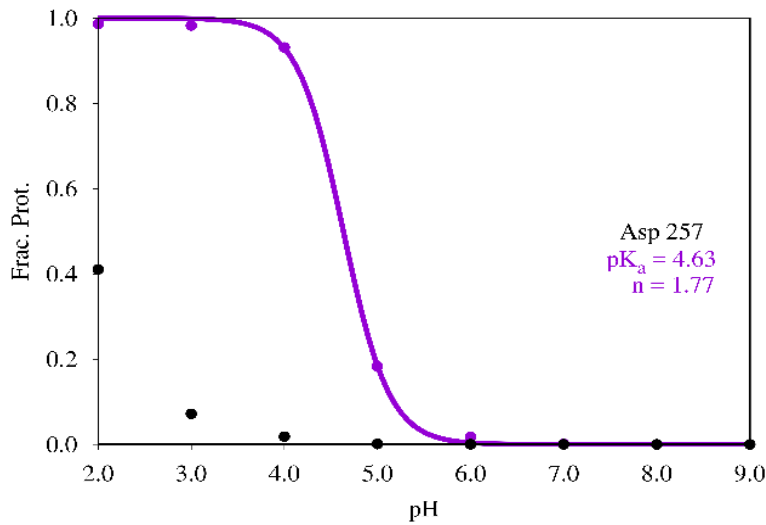


Figura 16. Evolución del pKa del aminoácido D₂₅₇ con el tiempo, calculado con la ecuación de Hill para la GS.



	Parámetros	Error asintótico estándar (%)
pK_{a257}	4.63	0.38
n_{257}	1.77	4.20

Figura 17. Curva de titulación del D257 en los últimos 5 ns de la simulación.

Mediante estudios de rayos X de resolución atómica y de difracción de neutrones de resolución media de un complejo de endotiapepsina con un inhibidor análogo al estado de transición, se determinó que el Asp32 (para esta proteasa) es interpretado como ionizado, mientras que el Asp215 como protonado. Y, aunque la determinación de la estructura de neutrones no tuvo evidencia de un hidrógeno entre los oxígenos internos de los aspárticos, la distancia determinada entre estos átomos es de 2.93 Å.³³

De los estudios experimentales y teóricos de diferentes aspartil proteasas, se recupera que su actividad está asociada al estado monoprotonado y a la distancia entre los carboxilos que se encuentran correlacionados. También, se ha determinado experimentalmente mediante un ensayo *in vitro* para la escisión de la GS, que su pH óptimo es de 6.8, siendo activa de pH 6.0 a 8.4.⁶⁶ Este resultado es consistente con un ensayo *in vivo* (en el cerebro de una rata), en el que se observó que el pH óptimo es alrededor de 7.⁶⁷ Por lo que se esperaría que, de acuerdo con las tendencias de las fracciones protonadas obtenidas, con más tiempo de simulación se obtendrían pKas un poco mayores a los que se obtuvieron con el tiempo de simulación realizado en este trabajo.

c) Interacción entre los aspárticos catalíticos

La tendencia en los valores de las distancias entre los oxígenos de los carbonilos con respecto al pH (Figura 18) es similar a la tendencia que observamos en el *RMSF*, y al igual que en este, las distancias promedio a pH 2 y pH 3 son las más cortas y se superponen. En el último tercio de la simulación, la distancia también disminuye a pH 4 hasta superponerse. En algunos tiempos, la distancia promedio a pH 5 disminuye, mientras que la distancia promedio a pHs mayores se mantiene aproximadamente constante.

A pH 4 y 5, la tendencia de la distancia entre los oxígenos carboxílicos de los aspárticos, concuerda con la de la formación del puente de hidrógeno carboxilo-carboxilo. Durante la última parte de la simulación a pH 4, también se observó una mayor probabilidad de formación de este puente de hidrógeno y una disminución de la distancia entre los oxígenos carboxílicos de los aspárticos. Alrededor de los 20 ns a pH 5, también se encontró que hay una disminución en la distancia, conduciendo la formación del puente de hidrógeno.

Comparando las Figura 19 y 20, es posible identificar que la unión por puente de agua entre los aspárticos cambia por el puente de hidrógeno carboxilo-carboxilo al final de la trayectoria.

En la Figura 15, se observa que a pH 3 se tiene la mayor separación entre las fracciones protonadas de los aspárticos. Aunque las distancias a pH 3 tienen la misma tendencia que a pH 2, se puede observar menor interacción entre los carboxilos (ver las Figura 19 y 20), esto coincide a su vez con la menor interacción entre el D385 y la R269 (Figura 22). Una explicación a ello podría ser que sólo hay interacción por carga sin que se forme el puente de hidrógeno. Esta particularidad también se observa en la Figura 21 en la que la formación del puente de hidrógeno no decae de la forma esperada.

Con base en las distancias entre los oxígenos carboxílicos y la probabilidad de unión por puente de agua; se puede concluir que a pHs altos aumentan las distancias y, por tanto, la probabilidad de interacción disminuye.

En la Figura 23, se representan capturas del estado conformacional y la protonación finales de la díada catalítica de pH 2 a 5. A pH 2 se tiene una configuración diprotonada, en la que los aspárticos se encuentran relativamente alejados y no interactúan directamente; mientras que, a pH 3 la monoprotonación permite la formación de un puente de hidrógeno entre carboxilos, siendo éste el estado con mayor cercanía entre los aspárticos. A pH 4, en el que a tiempos finales, la configuración de la díada es mayormente didesprotonada, los carboxilos se encuentran separados, como es posible identificar en los tiempos finales de la Figura 18. Comparado con pH 4, a pH 5 la configuración es prácticamente didesprotonada, lo que incrementa la solvatación de los aspárticos por agua.

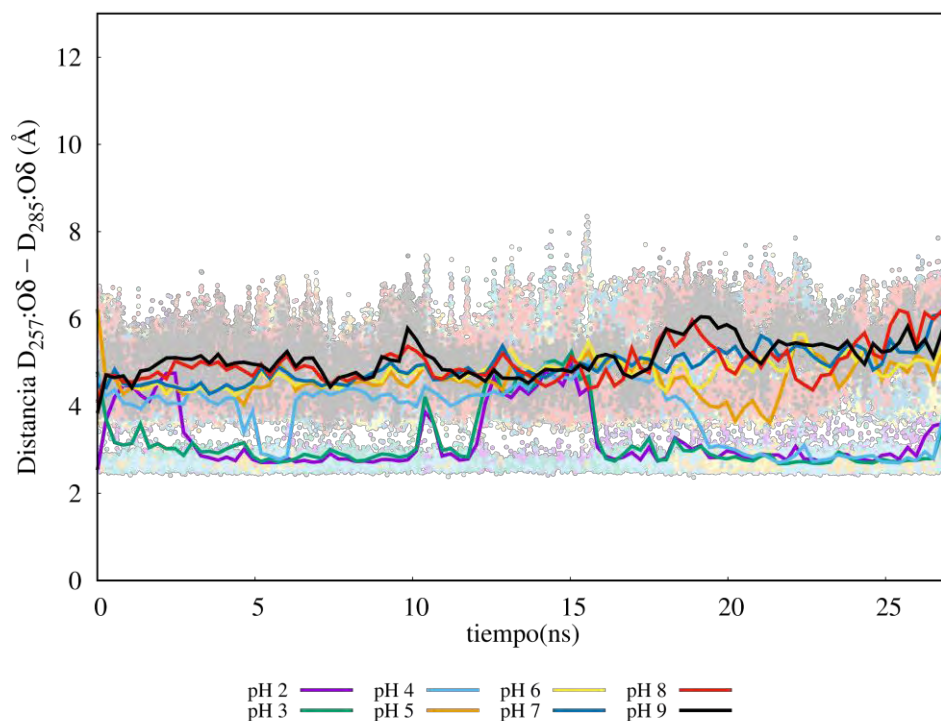


Figura 18. Distancia mínima entre los oxígenos de los aspárticos catalíticos para la GS. Las distancias más cortas se tienen a pH 2 y 3 (violeta y verde), en la última parte de la simulación a pH 4 la distancia disminuye considerablemente (azul claro).

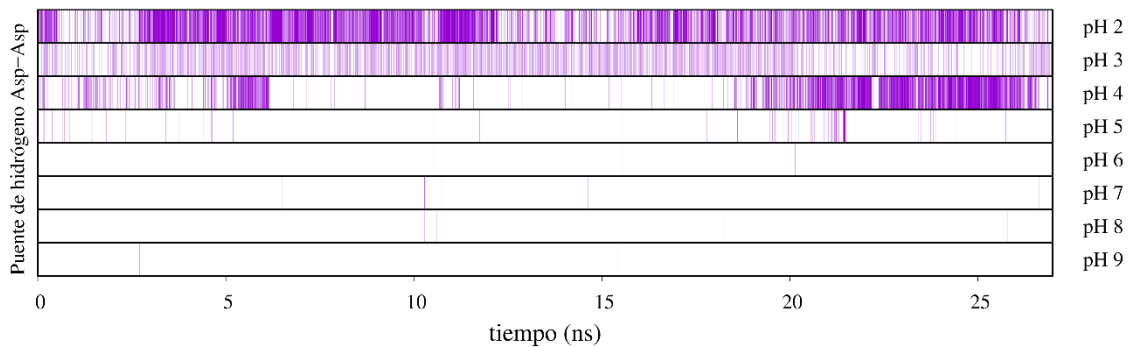


Figura 19. Formación de puentes de hidrógeno entre los carbonilos de los aspárticos catalíticos a lo largo de toda la trayectoria para la GS. Hay una mayor proporción de formación de puentes de hidrógeno a pH 2, mientras que en la última parte de la simulación se comienza a favorecer la formación de este puente de hidrógeno a pH 4.

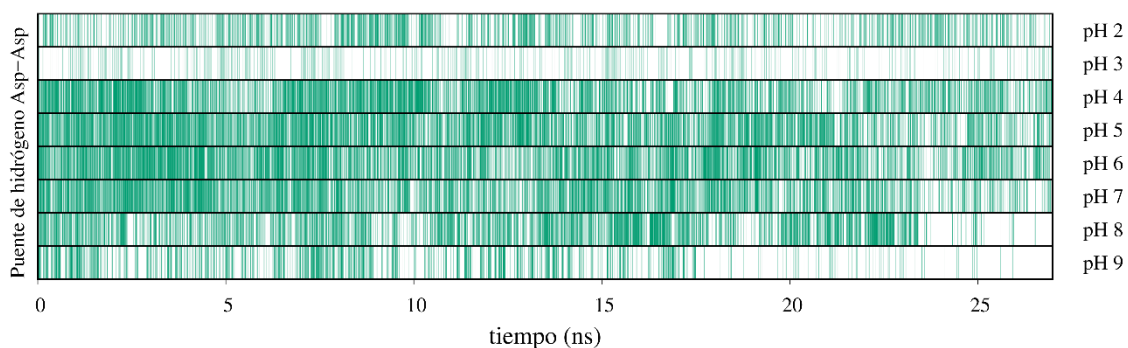


Figura 20. Formación de puentes con moléculas de agua entre los carbonilos de los aspárticos catalíticos a lo largo de toda la trayectoria para la GS. La interacción por puente de agua disminuye en los valores de pH más altos.

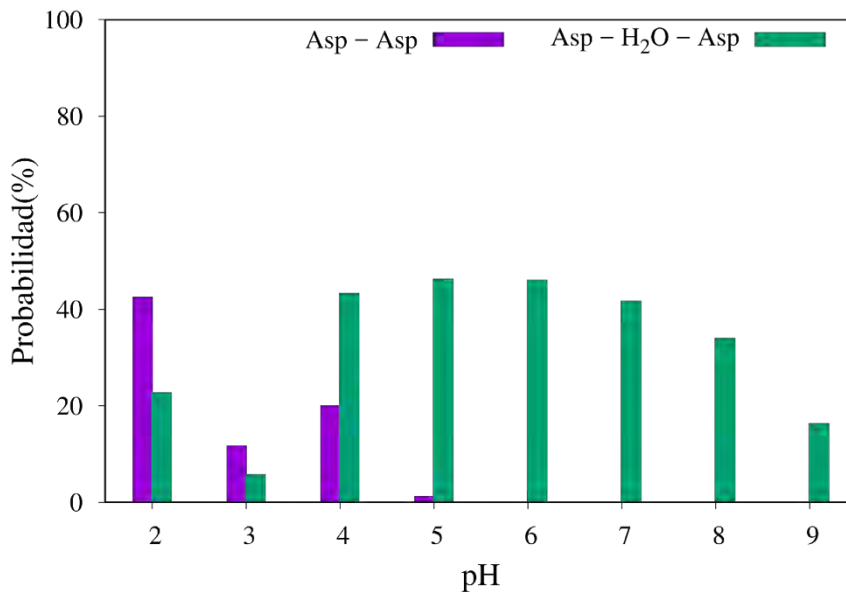


Figura 21. Formación de puentes de hidrógeno entre los carbonilos de los aspárticos catalíticos durante toda la trayectoria de la GS. La mayor proporción de formación de puentes Asp-Asp(violeta) se da a pH 2; sin embargo, se obtiene un segundo máximo a pH 4.

d) Interacción de los aspárticos catalíticos con otros aminoácidos cercanos al sitio activo

Ambos aspárticos interactúan con la R269 por puentes de hidrógeno; el D385 forma puentes de hidrógeno en casi toda la trayectoria con la R269, llegando a formar 2 puentes de hidrógeno a la vez. Justamente la mayor separación entre las fracciones de protonación de los aspárticos coincide con un desequilibrio en el modo de interacción con la R269.

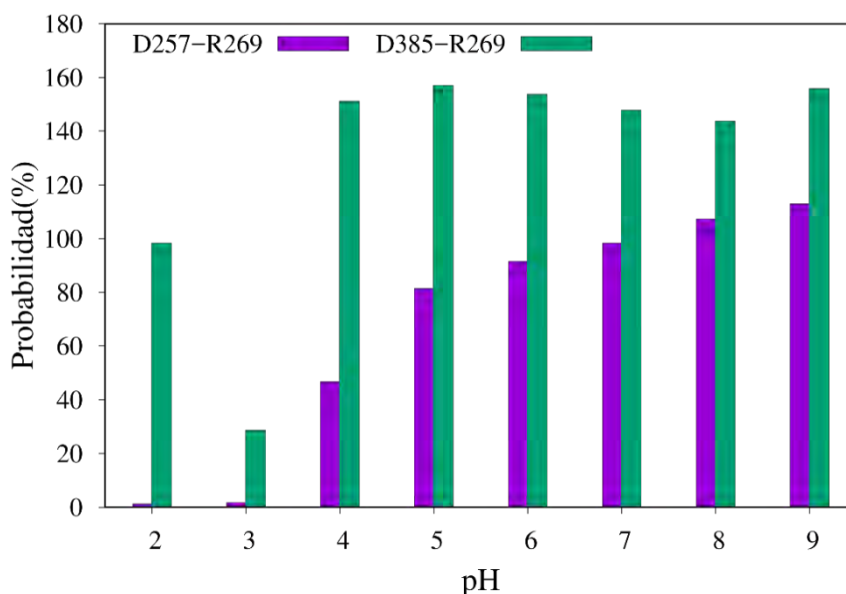


Figura 22. Formación de puentes de hidrógeno de los aspárticos catalíticos con la R269 de la GS. Todos los puentes de hidrógeno se forman entre las cadenas laterales, la probabilidad puede superar el 100% ya que se pueden formar más de un puente de hidrógeno.

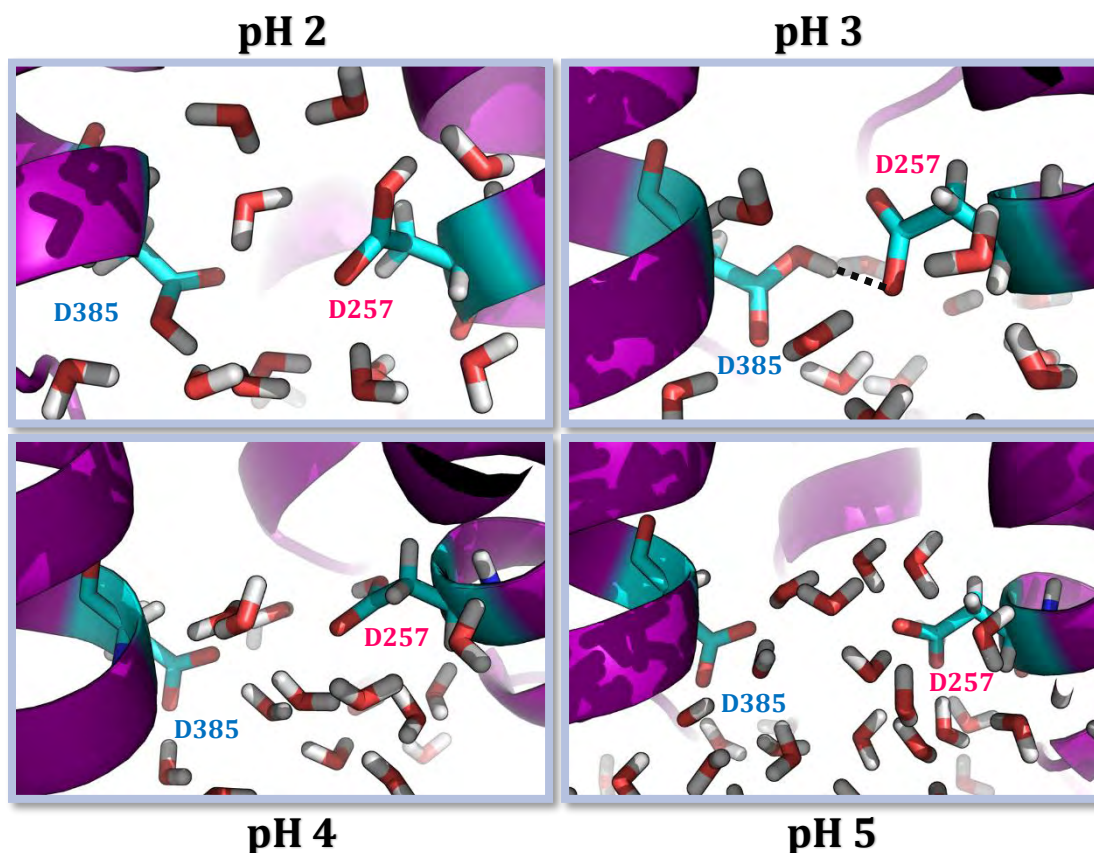


Figura 23. Imagen de la estructura del sitio catalítico de la GS en último paso de la simulación y a diferentes pHs. A pH 3 se tiene la mayor cercanía entre los aspárticos, debido a la unión por un puente de hidrógeno

6.2. Segundo enfoque: Efecto estructural del pH en un péptido intrínsecamente desordenado.

6.2.1. Primer caso de estudio. Caracterización del ensamble conformacional del R2/wt y su dependencia del pH.

a) Monitoreo del cambio conformacional del péptido durante la simulación

Debido a su naturaleza desordenada, se espera que la estructura del péptido cambie constantemente durante la simulación; en la Figura 24 se representa el *RMSD* a cada pH, el cuál fluctúa durante toda la simulación. En cuanto al *RMSF* (Figura 25), en la secuencia KK y D se observa una ligera dependencia con respecto al pH.

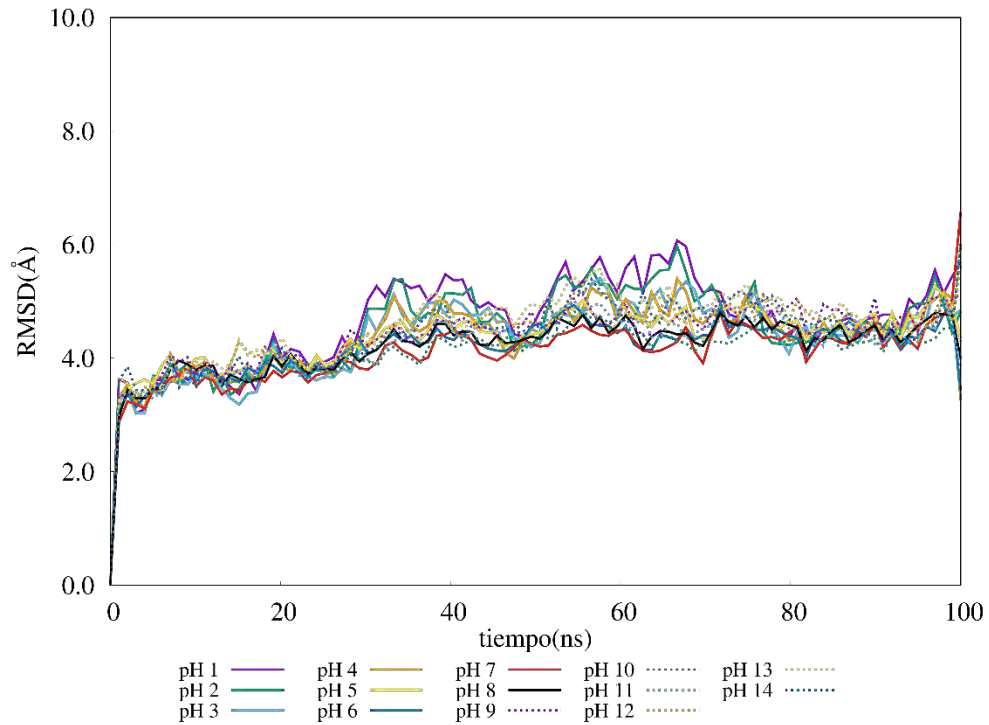


Figura 24. Desviación cuadrática media del *backbone* (N, C α ,C) del R2/wt a lo largo de la trayectoria para cada pH.

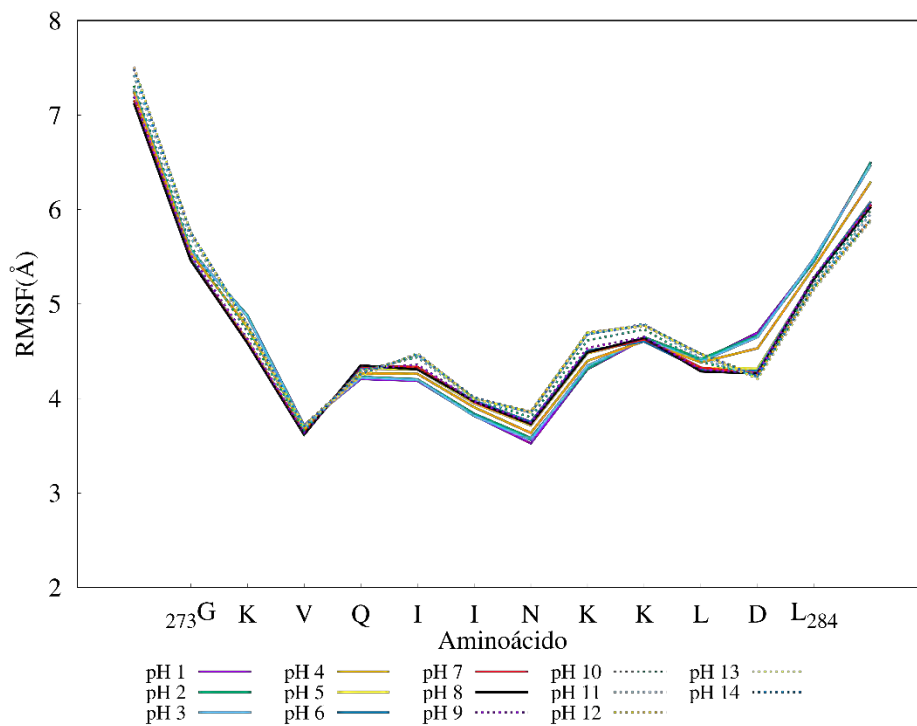


Figura 25. Fluctuación cuadrática media por aminoácido del R2/w.

b) Monitoreo de las fracciones de protonación de R2/wt

En el Apéndice B se pueden consultar las series temporales de las fracciones protonadas y las curvas de titulación de los aminoácidos titulables del péptido en ambas condiciones de solvatación. A diferencia de la GS, las fracciones protonadas son bastante estables en el intervalo de tiempo de simulación y se determinaron pKas de 3.89 y 4.11 para el aspártico y [10.15, 10.22, 10.29] y [10.19,10.21,10.26] para las lisinas, con solvatación implícita y explícita respectivamente.

c) Caracterización conformacional de R2/wt

Como ha sido mencionado anteriormente, Larini *et al.*⁵⁵ llevaron a cabo un estudio del autoensamblaje del péptido R2/wt utilizando la metodología de T-REMD (REMD de temperatura) con 62 réplicas con temperaturas de 290-492 K. De los resultados de su estudio plantearon dos principales familias conformacionales: de estructuras compactas y extendidas.

Debido a que la competencia entre la formación de puentes salinos es la razón de que el R2/wt sea un péptido intrínsecamente desordenado; se consideró caracterizar su ensamble conformacional mediante pH-REMD como una buena alternativa. Los métodos de simulación a pH constante permiten considerar la constante formación y ruptura de puentes salinos, lo que implica una mejora en el muestreo y en el modelado de este tipo de interacciones.

En la Figura 26 y C3 (Apéndice C), se representan las densidades de población de conformaciones del péptido con solvatación explícita e implícita (GB) respectivamente. Un radio de giro pequeño, asociado con una distancia de N- a C-terminal corta, corresponde a una estructura compacta, mientras que un radio de giro grande, asociado a una distancia de extremo a extremo larga, representa una estructura extendida. Con solvatación de agua explícita se obtuvo una distribución más amplia de conformaciones que con solvatación GB; además, con el modelo de solvatación GB las densidades altamente pobladas corresponden a estructuras más compactas que con agua explícita. Las interacciones con el disolvente usualmente permiten compensar la ruptura de interacciones proteína-proteína, lo que en este caso permite obtener estructuras más extendidas.

A pHs extremos (1-3 y 11-14), en los que prácticamente no existen puentes salinos debido a que el aspártico está protonado o las lisinas desprotonadas, hay un aumento en la densidad de población de las estructuras compactas, que están favorecidas por la formación de puentes de hidrógeno del *backbone*. Este aumento de puentes de hidrógeno intramoleculares se refleja en la Figura 27. A estos pHs, la probabilidad de formación de puentes de hidrógeno es mayor (~50%) mientras que en el intervalo de pH 3-10 es de alrededor de 40%.

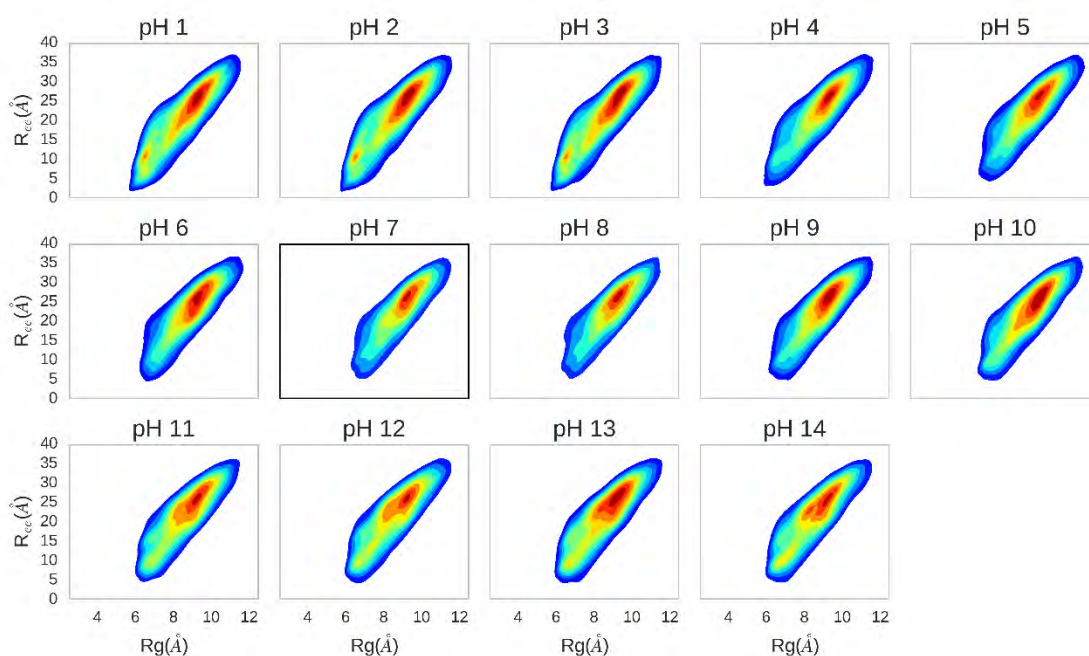


Figura 26. Densidades de población de conformaciones del péptido. R_{ce} es una medida de la distancia de extremo a extremo del péptido (N-terminal a C-terminal) y R_g es el radio de giro sobre el centro de masa. Las densidades altamente pobladas se indican en rojo, mientras que la población disminuye hacia azul.

Comparando con los resultados del grupo de Shea (Larini *et al.* y Levine *et al.*), mediante pH-REMD en el intervalo de pH 1-14, se obtuvo una mayor densidad de población en conformaciones más extendidas. Esto puede asociarse a que mediante pH-REMD con réplicas en todo el intervalo de pH, es más fácil superar la barrera energética asociada a la formación de puentes salinos. Esto también se puede obtener con T-REMD, ya que permite superar más fácilmente las barreras energéticas mediante la variación de temperaturas; pero en este caso, el pH tiene un impacto más grande ya que

se consideran pHs en los que la formación de estos puentes salinos es muy poco probable por el estado neutro de alguno de los aminoácidos involucrados. Esto permite tener durante toda la simulación, réplicas intercambiables con diferentes estados.

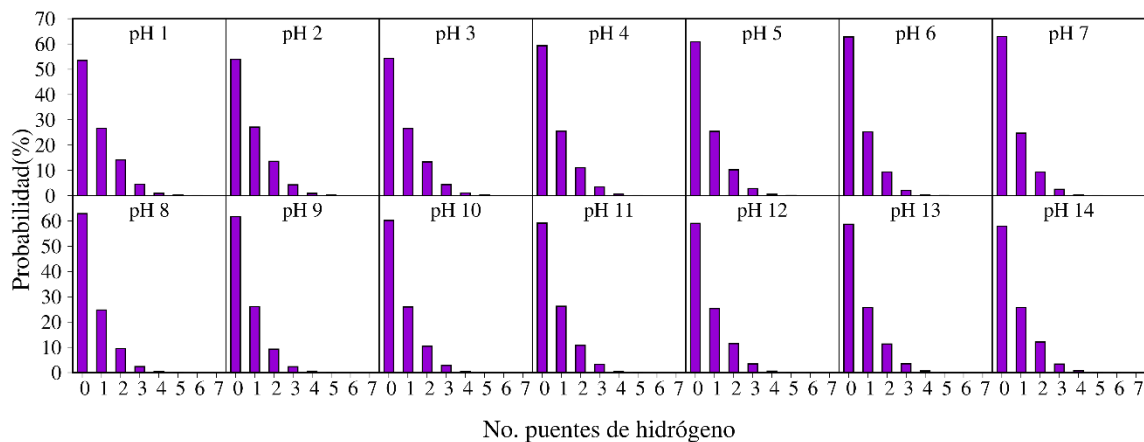


Figura 27. Probabilidades de formación de puentes de hidrógeno intramoleculares (C=O a N-H) sobre toda la trayectoria.

En consecuencia de la obtención de estructuras más extendidas, en la Figura 27 se obtuvo una menor probabilidad de formación de puentes de hidrógeno intramoleculares con respecto al grupo de Shea.

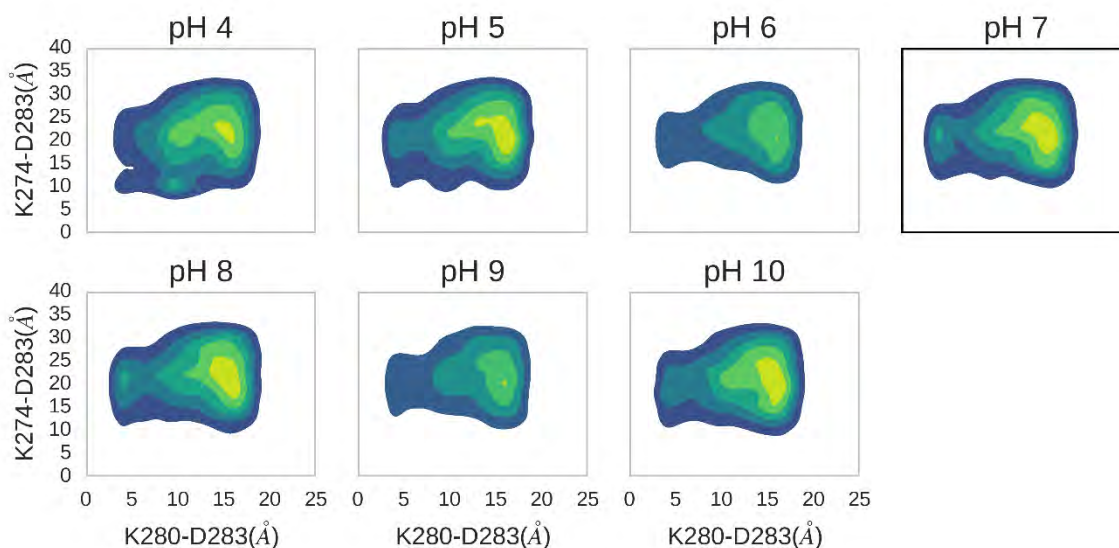
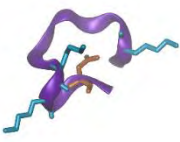

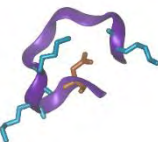
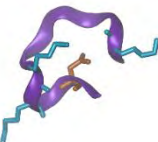



Figura 28. Densidades de población de puentes salinos monoméricos, representados por la proximidad relativa entre aspártico y lisinas. La densidad disminuye de amarillo a verde oscuro.

En la Figura 28 se representan las distancias K-D, en las que es posible distinguir que la formación de puentes salinos entre K280-D283 está favorecida sobre los K274-D283, otra diferencia con el estudio previamente mencionado; por supuesto, estas diferencias están correlacionadas. A pH 4 aparece densidad asociada a la cercanía entre K274-D283, eso se debe a que es el pH en el que hay mayor transición entre los estados de protonación del aspártico, lo que a su vez provoca la mayor ruptura y formación de puentes salinos.

El análisis de *clustering* permitió obtener las estructuras más representativas del ensamble conformacional. En la Figura 29 se representan las figuras más representativas a cada pH; se tienen tres estructuras diferentes, a pH 1-4 se obtuvo una estructura compacta con poca formación de puentes de hidrógeno en el *backbone*, y con una disposición espacial en forma de "C". A pH 5-9 la estructura más representativa es una conformación extendida con un giro en la sección más polar del péptido. La tercera estructura es la más representativa de pH 10-14, consiste en una estructura compacta con forma de "U". En el Apéndice E, se pueden consultar las 3 estructuras más representativas a cada pH, que, en general, son variantes de las tres estructuras principales además de algunas conformaciones más helicoidales.

pH 1	pH 2	pH 3	pH 4	pH 5
				
5.02%	4.85%	4.56%	2.40%	1.31%

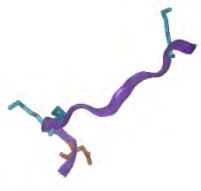








pH 6	pH 7	pH 8	pH 9	pH 10
				
1.33%	1.22%	1.44%	1.41%	2.02%
pH 11	pH 12	pH 13	pH 14	
				
2.85%	3.49%	3.39%	3.41%	

Figura 29. Estructuras más representativas del monómero simulado con agua explícita a cada pH. Las cadenas laterales del aspártico y las lisinas son representadas en color naranja y cian, respectivamente.

6.2.2. Segundo caso de estudio. El efecto del pH en la dimerización del péptido R2/wt

a) Monitoreo del cambio conformacional del péptido durante la simulación

Generalmente se puede identificar la estabilidad de dímero en su *RMSD*, cuando es estable se mantiene prácticamente constante; mientras que cuando se disocia hay grandes variaciones por la separación de los monómeros. En este caso, en la Figura 30 se tiene una desviación estructural promedio bastante regular; a diferencia, en los modelos con solvatación implícita (Apéndice D) se observó mayor fluctuación en el *RMSD*.

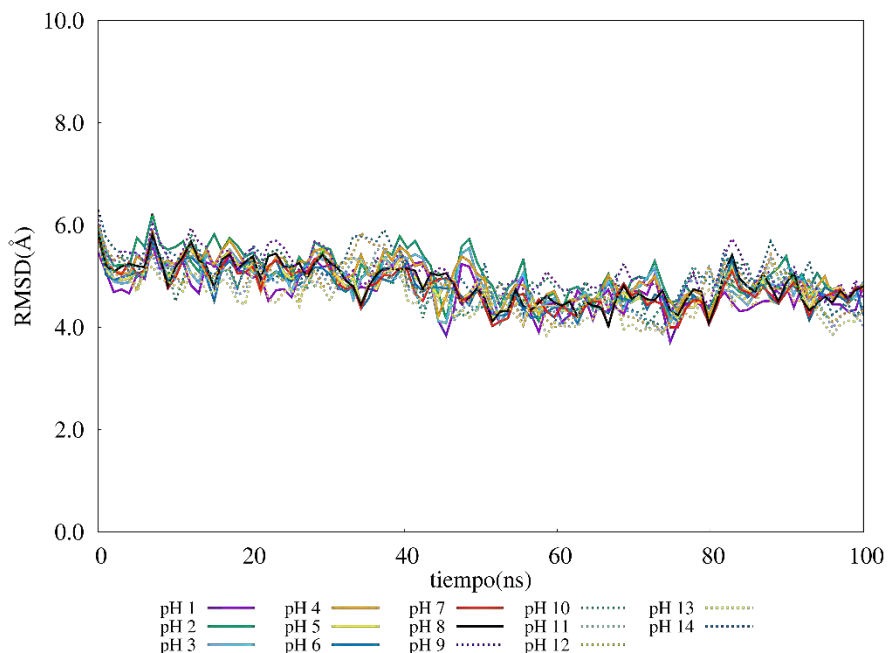


Figura 30. Desviación cuadrática media del *backbone* (N, C α ,C) del dímero del R2/wt a lo largo de la trayectoria a cada pH.

En la representación del *RMSF* (Figura 31) no hay tendencias particulares con dependencia del pH. En el fragmento PHF6* (VQIINK) no se observa alguna variación anormal; a diferencia, en el modelo de solvatación implícita (Apéndice D) la fluctuación cambia mucho con los diferentes pHs. La relevancia de este fragmento consiste en que es particularmente propenso a agregarse, por lo que variaciones en el *RMSF* en esta zona pueden implicar inestabilidad por efecto del pH.

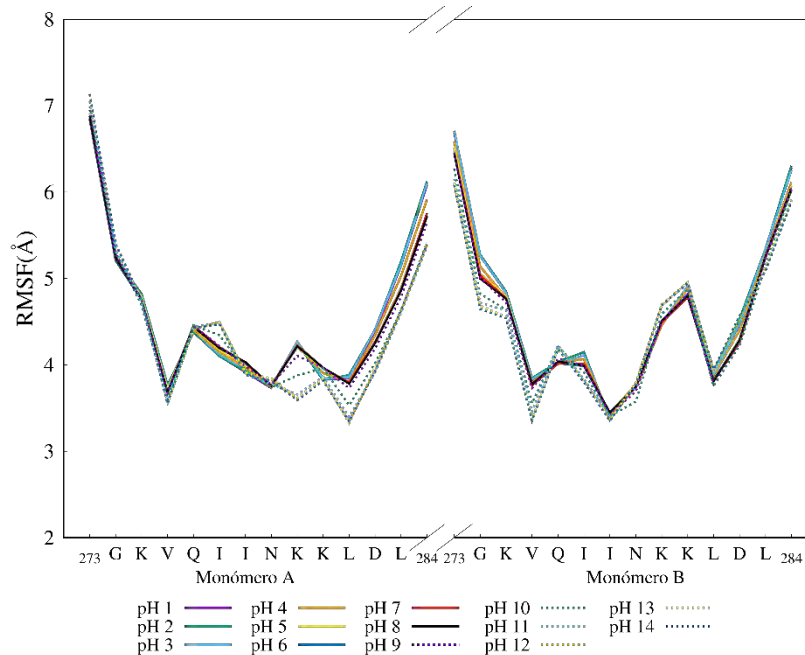


Figura 31. Fluctuación cuadrática media por aminoácido del dímero del R2/wt.

b) Monitoreo de las fracciones de protonación del R2/wt

En el Apéndice B se pueden consultar las series temporales de las fracciones protonadas y las curvas de titulación de los aminoácidos titulables del dímero en ambas condiciones de solvatación. Los pKas son parecidos a los valores obtenidos en los monómeros, las fracciones protonadas fluctúan más comparado con los primeros debido, principalmente, a la interacción entre los aminoácidos titulables de las dos cadenas.

c) Caracterización conformacional del dímero del R2/wt y su estabilidad con el pH

En la Figura 32, se evalúa el efecto del pH en la estabilidad del dímero del péptido tau. En dicha figura, se puede identificar que la disminución del pH favorece la disociación, debido a que hay menos contacto entre los fragmentos PHF6* de los monómeros. En el Apéndice F se puede consultar la representación gráfica del número de contactos entre las cadenas completas, así como las figuras correspondientes a la simulación con solvatación GB.

Con solvatación *GB*, se obtuvo una mayor disociación por efecto del pH. La presencia de agua favorece la interacción entre las regiones no polares por lo que estabiliza al dímero.

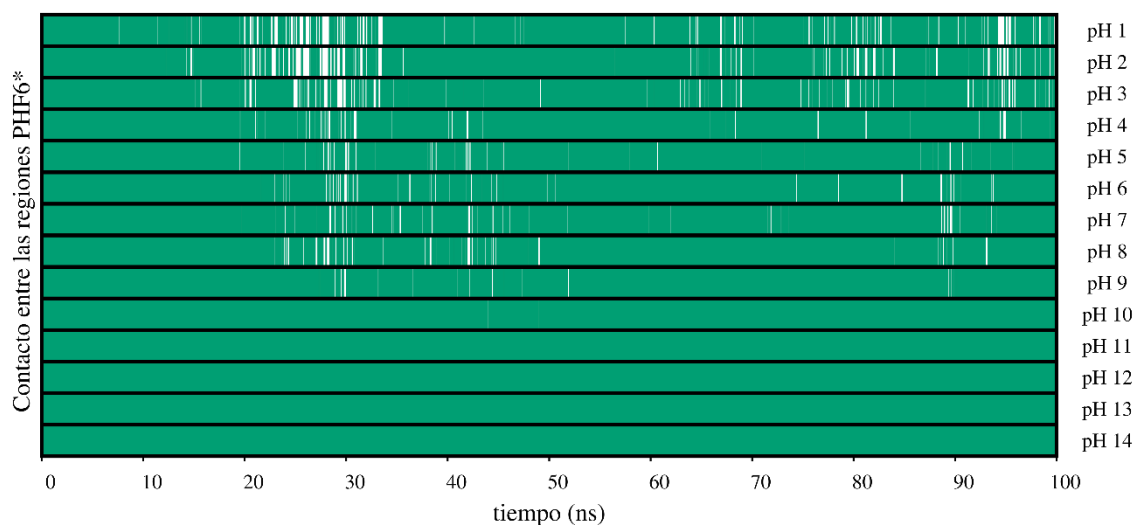


Figura 32. Contacto entre las regiones PHF6* de los monómeros A y B a cada pH. La dimerización está favorecida por el pH.

En la Figura 33 y 34, es posible distinguir cómo a pHs extremos, se tiene una complementariedad en la que un monómero pasa más tiempo en una sola conformación, mientras que la otra cadena presenta una mayor transición entre conformaciones. A pHs intermedios, ambas cadenas se pasean en una forma semejante entre conformaciones. En las estructuras más representativas del dímero presentadas en la Figura 35 también es posible distinguir esta tendencia, en la que un monómero está extendido y el otro tiene una conformación más compacta a pHs básicos o ácidos, mientras que a pHs neutros ambos tienen estructuras medianamente extendidas. En el Apéndice E, se pueden consultar las 3 estructuras más representativas a cada pH.

Como se representa en la Figura 35, la dimerización conduce a conformaciones más ordenadas, característica de los agregados de la proteína tau del tipo de estructura β -cruzada. Debido a ello, se llevó a cabo un análisis de estructura secundaria, cuyos resultados se presentan en la Figura 36 y en la cual se identifica cómo el pH favorece el orden, tanto en la formación de hélices como de láminas β .

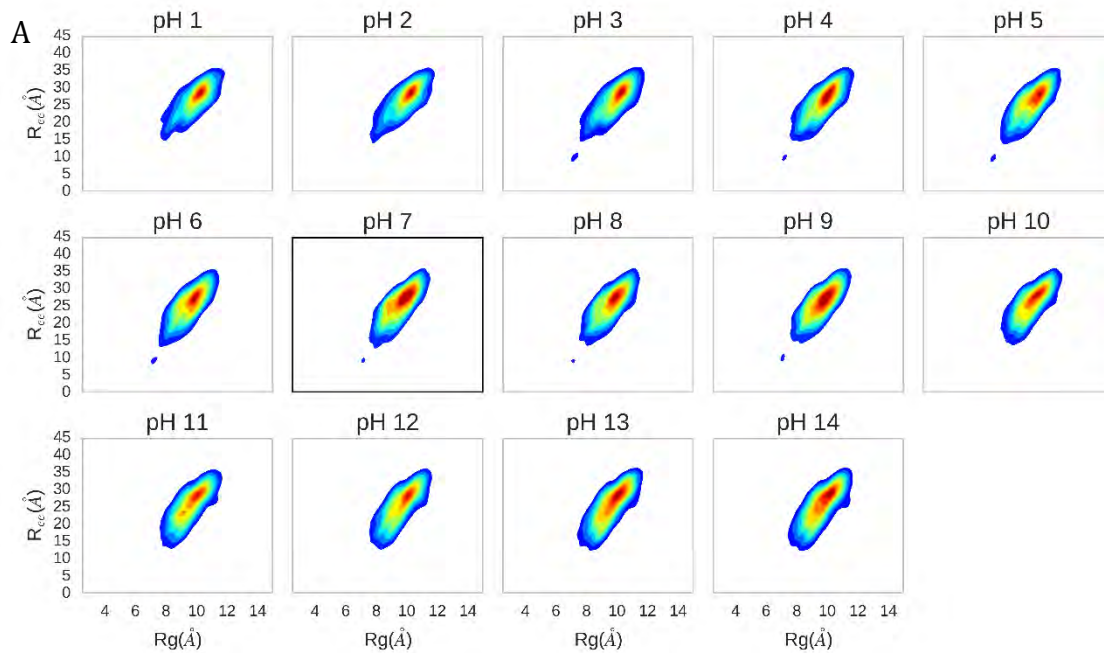


Figura 33. Densidades de población de conformaciones del monómero A. R_{ee} es una medida de la distancia de extremo a extremo del péptido y R_g es el radio de giro sobre el centro de masa.

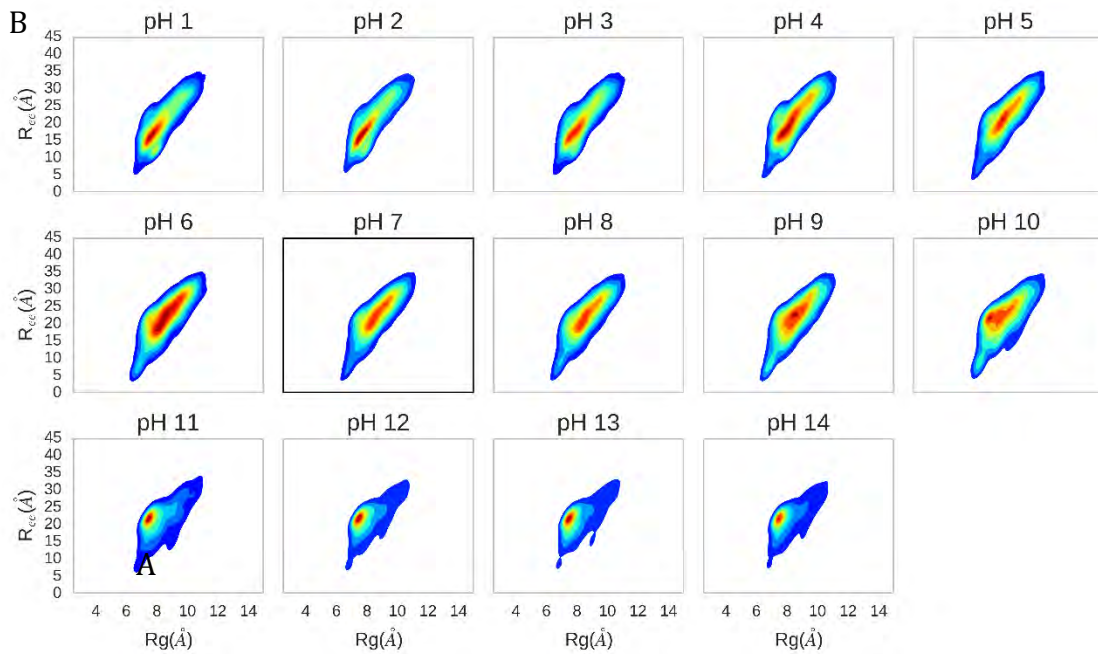


Figura 34. Densidades de población de conformaciones del monómero B. R_{ee} es una medida de la distancia de extremo a extremo del péptido y R_g es el radio de giro sobre el centro de masa.


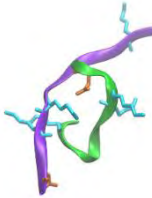
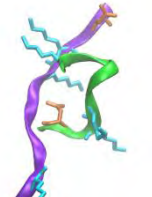





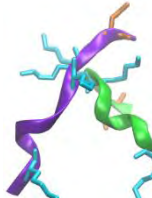




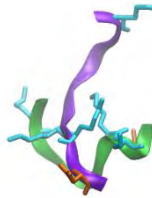
pH 1	pH 2	pH 3	pH 4	pH 5
 2.25%	 2.09%	 1.98%	 2.60%	 3.62%
pH 6	pH 7	pH 8	pH 9	pH 10
 3.37%	 3.20%	 3.92%	 4.90%	 6.21%
pH 11	pH 12	pH 13	pH 14	
 4.23%	 4.73%	 5.50%	 5.00%	

Figura 35. Estructuras más representativas del dímero simulado con agua explícita a cada pH. Las cadenas laterales de los aspárticos y las lisinas son representadas en color naranja y cian respectivamente.

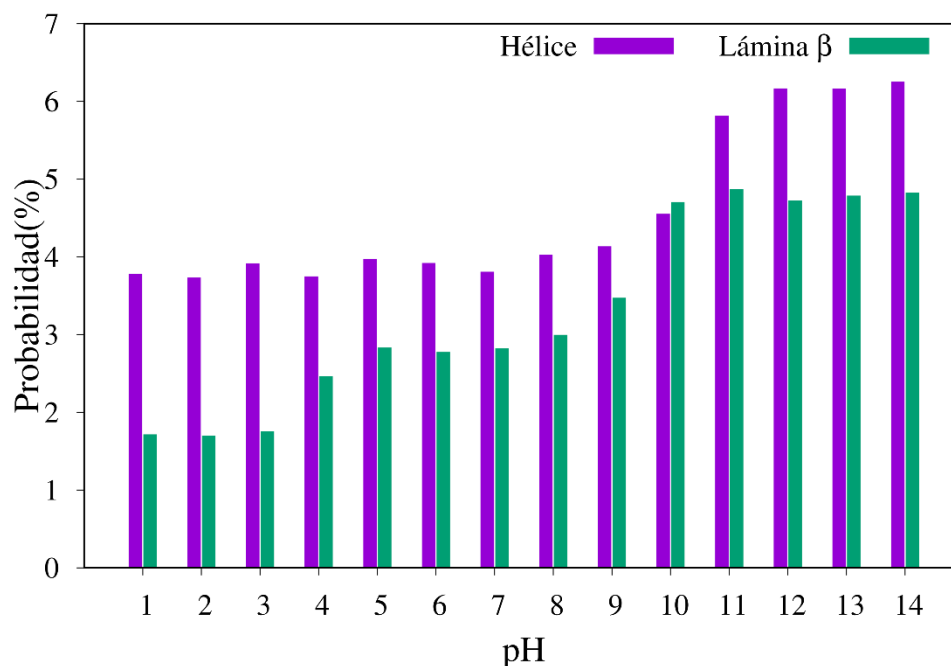


Figura 36. Probabilidad de formación de hélices y láminas β en el dímero del R2/wt. El análisis de estructura secundaria se llevó a cabo utilizando DSSP⁶⁸. En hélice están contenidas α , 3-10 y π -hélices y en lámina β , paralelas y antiparalelas.

7. CONCLUSIONES

Se llevó a cabo el muestreo conformacional y de estados de protonación de los aspárticos catalíticos de la GS y de un péptido de la proteína tau.

En la DM a pH constante, se considera la correlación conformación-protonación mediante el muestreo simultáneo de ambos ensambles, liberando al sistema de la sobresimplificación de estados de protonación constante. Este sistema de perturbación permite monitorear la evolución conformacional asociada a la formación y ruptura de interacciones electrostáticas, que a su vez se asocian a los cambios de carga de los aminoácidos titulables.

La dinámica de las interacciones por efecto de la protonación está fuertemente asociada a la activación de enzimas. En la GS, es posible identificar que el equilibrio de protonación acciona en forma distinta para los dos aspárticos catalíticos, permitiendo el reordenamiento de la configuración (conformación-protonación) del sitio activo. La

diferencia de acidez relativa conduce a diferentes estados dependientes del pH, caracterizados por la distancia e interacción entre la díada catalítica.

La inclusión del pH, como uno de los parámetros termodinámicos del sistema, potencia la caracterización del ensamble conformacional de *IDPs* (como tau), en los que el estado conformacional está dirigido por la competencia entre la formación de puentes salinos y puentes de hidrógeno. La permisión del cambio de estados de protonación considera los efectos entrópicos asociados con la formación y ruptura de puentes salinos, liberando restricciones del muestreo conformacional. Además, mediante simulaciones a pH constante, se puede caracterizar la estabilidad y conformación de dímeros en un intervalo de pH.

7.1. Conclusiones particulares

7.1.1. Estudio de los estados de protonación de los aspárticos catalíticos de la GS

- Se identificó una mayor acidez relativa del aminoácido D385 con respecto al aminoácido D257 en el sitio catalítico de la GS, con un pKa final de 4.63 para D385 y un pKa indeterminado en el intervalo de pH muestreado para el D257.
- Se obtuvo que la interacción por cercanía y puentes de hidrógeno es mayor cuando el par aspártico se encuentra en un estado monoprotonado.
- Las fracciones protonadas de ambos aspárticos tuvieron una tendencia creciente a tiempos finales de la simulación.
- Las lisinas K265 y K380 presentaron diferente basicidad relativa, siendo la lisina K380 la menos básica.

7.1.2. Caracterización estructural del péptido tau (273-284)

- Se caracterizó el ensamble conformacional del péptido considerando la ruptura y formación de puentes salinos y puentes de hidrógeno por efecto del pH. De pH 1

a 14, las estructuras más representativas obtenidas concuerdan con las estructuras de la propuesta de superposición de ensamblajes del grupo de Shea.^{55,56}

- Se encontraron tres estructuras representativas en el intervalo de pH simulado. A pHs ácidos, de 1 a 4, se obtuvieron estructuras compactas con poca formación de puentes de hidrógeno en el *backbone*, y con una disposición espacial en forma de C. A pHs intermedios, de 5 a 9, la estructura más representativa es una conformación extendida que presenta un giro en la región más polar del péptido. A pHs básicos, de 10 a 14, la estructura más representativa obtenida es una estructura compacta en forma de U.

- Se obtuvieron estructuras más extendidas para el péptido tau con respecto al grupo de Shea. Esto es consecuencia de que el muestreo de los estados de protonación permite superar más fácilmente las barreras energéticas de la ruptura de puentes salinos.

7.1.3. Dimerización del péptido tau (273-284)

- Se obtuvo que a pHs extremos se tiene una complementariedad en la que un monómero ocupa predominantemente una conformación, mientras que la otra cadena presenta una mayor transición entre conformaciones. A pHs intermedios, ambas cadenas tienen cambios conformacionales similares. En las estructuras más representativas del dímero se puede distinguir esta tendencia, en la que a pHs ácidos o básicos un monómero está extendido y el otro tiene una conformación más compacta, mientras que a pHs neutros, ambos tienen estructuras medianamente extendidas.

- Se identificó que la dimerización de tau es favorecida con el aumento del pH. A su vez, ésta dimerización conduce a conformaciones más ordenadas y característica de los agregados de la proteína tau, una estructura β -cruzada. Mediante un análisis de estructura secundaria se obtuvo que el pH favorece el orden, tanto en la formación de hélices como de láminas β .

8. REFERENCIAS

1. Baptista, A. M., Martel, P. J. & Petersen, S. B. Simulation of protein conformational freedom as a function of pH: Constant pH molecular dynamics using implicit titration. *Proteins Struct., Funct., Genet.* **27**, 523 (1997).
2. Kessel, A. & Ben-Tal, N. *Introduction to proteins : structure, function, and motion.* (CRC Press, 2011).
3. Onufriev, A., Case, D. A. & Ullmann, G. M. A novel view of pH titration in biomolecules. *Biochemistry* **40**, 3413–3419 (2001).
4. James, N. W. The Hill equation revisited: uses and misuses. *FASEB J.* **11**, 835–841 (1997).
5. Allen, M. Introduction to molecular dynamics simulation. *Comput. Soft Matter From Synth. Polym. to Proteins* **23**, 1–28 (2004).
6. Leach, A. R. *Molecular modelling : principles and applications.* (Prentice Hall, 2001).
7. González, M. A. Force fields and molecular dynamics simulations. *Collect. SFN* **12**, 169–200 (2011).
8. Cheung, D. L., Anton, L., Allen, M. P. & Masters, A. J. Computer simulation of liquids and liquid crystals. *Computer Physics Communications* **179**, 61–65 (2008).
9. Frenkel, D. & Smit, B. Understanding Molecular Simulation. *Understanding Molecular Simulation* 638 (2002). doi:10.1016/B978-012267351-1/50006-7
10. Baptista, a M., Martel, P. J. & Soares, C. M. Simulation of electron-proton coupling with a Monte Carlo method: application to cytochrome c3 using continuum electrostatics. *Biophys. J.* **76**, 2978–2998 (1999).
11. Baptista, A. M., Teixeira, V. H. & Soares, C. M. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* **117**, 4184–4200 (2002).
12. Meng, Y. & Roitberg, A. E. Constant pH replica exchange molecular dynamics in

- biomolecules using a discrete protonation model. *J. Chem. Theory Comput.* **6**, 1401–1412 (2010).
13. Börjesson, U. & Hünenberger, P. H. Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines. *J. Chem. Phys.* **114**, 9706–9719 (2001).
 14. Mongan, J., Case, D. A. & McCammon, J. A. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **25**, 2038–2048 (2004).
 15. Metropolis, N. & Ulam, S. The Monte Carlo Method. *J. Am. Stat. Assoc.* **44**, 335–341 (1949).
 16. Swails, J. M., York, D. M. & Roitberg, A. E. Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: Implementation, testing, and validation. *J. Chem. Theory Comput.* **10**, 1341–1352 (2014).
 17. Sabri Dashti, D., Meng, Y. & Roitberg, A. E. pH-replica exchange molecular dynamics in proteins using a discrete protonation method. *J. Phys. Chem. B* **116**, 8805–8811 (2012).
 18. Khandogin, J. & Brooks, C. L. Toward the accurate first-principles prediction of ionization equilibria in proteins. *Biochemistry* **45**, 9363–9373 (2006).
 19. Swails, J. M. & Roitberg, A. E. Enhancing conformation and protonation state sampling of hen egg white lysozyme using pH replica exchange molecular dynamics. *J. Chem. Theory Comput.* **8**, 4393–4404 (2012).
 20. Itoh, S. G., Damjanović, A. & Brooks, B. R. PH replica-exchange method based on discrete protonation states. *Proteins Struct. Funct. Bioinforma.* **79**, 3420–3436 (2011).
 21. LaFerla, F. M., Green, K. N. & Oddo, S. Intracellular amyloid- β in Alzheimer's disease. *Nat. Rev. Neurosci.* **8**, 499–509 (2007).
 22. Bolduc, D. M., Montagna, D. R., Seghers, M. C., Wolfe, M. S. & Selkoe, D. J. The amyloid-beta forming tripeptide cleavage mechanism of γ -Secretase. *Elife* **5**, 1–

- 21 (2016).
23. Aguayo-Ortiz, R. & Dominguez, L. Generation of Amyloid- β Peptides by γ -Secretase. *Isr. J. Chem.* 1–13 (2016).
 24. Mucke, L. Neuroscience: Alzheimer's disease. *Nature* **461**, 895–7 (2009).
 25. Strisovsky, K. Why cells need intramembrane proteases – a mechanistic perspective. *FEBS J.* **283**, 1837–1845 (2015).
 26. Sun, L. *et al.* Structural basis of human γ -secretase assembly. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6003–6008 (2015).
 27. Wolfe, M. S. *et al.* Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and gamma-secretase activity. *Nature* **398**, 513–7 (1999).
 28. Alberts, B. *et al.* *Molecular Biology of the Cell. 4th Edition, New York* (2002).
 29. Sun, L., Li, X. & Shi, Y. Structural biology of intramembrane proteases: Mechanistic insights from rhomboid and S2P to γ -secretase. *Curr. Opin. Struct. Biol.* **37**, 97–107 (2016).
 30. Bai, X. *et al.* An atomic structure of human γ -secretase. *Nature* **525**, 212–7 (2015).
 31. Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* **4**, (2015).
 32. Coates, L., Erskine, P. T., Wood, S. P., Myles, D. A. A. & Cooper, J. B. A neutron laue diffraction study of endothiapepsin: Implications for the aspartic proteinase mechanism. *Biochemistry* **40**, 13149–13157 (2001).
 33. Coates, L. *et al.* The catalytic mechanism of an aspartic proteinase explored with neutron and X-ray diffraction. *J. Am. Chem. Soc.* **130**, 7235–7237 (2008).
 34. Piana, S. & Carloni, P. Conformational flexibility of the catalytic Asp dyad in HIV-1 protease: An ab initio study on the free enzyme. *Proteins Struct. Funct. Genet.*

- 39**, 26–36 (2000).
35. Piana, S., Sebastiani, D., Carloni, P. & Parrinello, M. Ab initio molecular dynamics-based assignment of the protonation state of pepstatin A/HIV-1 protease cleavage site. *J. Am. Chem. Soc.* **123**, 8730–8737 (2001).
 36. Friedman, R. & Caflich, A. The protonation state of the catalytic aspartates in plasmepsin II. *FEBS Lett.* **581**, 4120–4124 (2007).
 37. Singh, R., Barman, A. & Prabhakar, R. Computational Insights into Aspartyl Protease Activity of Presenilin 1 (PS1) Generating Alzheimer Amyloid # - Peptides (A # 40 and A # 42) Computational Insights into Aspartyl Protease Activity of Presenilin 1 (PS1) Generating Alzheimer Amyloid -Pept. **1**, 2990–2999 (2009).
 38. McGee, T. D., Edwards, J. & Roitberg, A. E. pH-REMD simulations indicate that the catalytic aspartates of HIV-1 protease exist primarily in a monoprotonated state. *J. Phys. Chem. B* **118**, 12577–12585 (2014).
 39. Ellis, C. R. & Shen, J. pH-Dependent Population Shift Regulates BACE1 Activity and Inhibition. *J. Am. Chem. Soc.* **137**, 9543–9546 (2015).
 40. Uversky, V. N. What does it mean to be natively unfolded? - Uversky - 2003 - European Journal of Biochemistry - Wiley Online Library. *Eur. J. Biochem.* **12**, 2–12 (2002).
 41. Uversky, V. N. A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Sci.* **22**, 693–724 (2013).
 42. Le Gall, T., Romero, P. R., Cortese, M. S., Uversky, V. N. & Dunker, A. K. Intrinsic disorder in the Protein Data Bank. *J Biomol Struct Dyn* **24**, 325–342 (2007).
 43. Levine, Z. A. & Shea, J. Simulations of disordered proteins and systems with conformational heterogeneity. *Curr. Opin. Struct. Biol.* **43**, 95–103 (2017).
 44. Lee, R. Van Der *et al.* Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **114**, 6589–6631 (2014).

45. Uversky, V., Gillespie, J. & Fink, A. Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427 (2000).
46. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Pnas* **110**, 13392–13397 (2013).
47. Csizmok, V., Follis, A. V., Kriwacki, R. W. & Forman-Kay, J. D. Dynamic Protein Interaction Networks and New Structural Paradigms in Signaling. *Chem. Rev.* **116**, 6424–6462 (2016).
48. Jordan, M. A. & Wilson, L. Microtubules as a target for anticancer drugs. *Nat. Rev. Cancer* **4**, 253–265 (2004).
49. Mandelkow, E. & Mandelkow, E. M. Microtubules and microtubule-associated proteins. *Curr. Opin. Cell Biol.* **7**, 72–81 (1995).
50. Grundke-Iqbal, I. *et al.* Microtubule-associated protein tau. A component of Alzheimer paired helical filaments. *J. Biol. Chem.* **261**, 6084–9 (1986).
51. Iqbal, K. *et al.* Defective brain microtubule assembly in Alzheimer's disease. *Lancet* **328**, 421–426 (1986).
52. Grundke-Iqbal, I. *et al.* Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 4913–7 (1986).
53. Wang, J.-Z., Xia, Y.-Y., Grundke-Iqbal, I. & Iqbal, K. Abnormal Hyperphosphorylation of Tau: Sites, Regulation, and Molecular Mechanism of Neurofibrillary Degeneration. *J. Alzheimer's Dis.* **33**, S123–S139 (2013).
54. Peterson, D. W., Zhou, H., Dahlquist, F. W. & Lew, J. A soluble oligomer of tau associated with fiber formation analyzed by NMR. *Biochemistry* **47**, 7393–7404 (2008).
55. Larini, L. *et al.* Initiation of assembly of tau(273-284) and its Δ K280 mutant: an experimental and computational study. *Phys. Chem. Chem. Phys.* **15**, 8916 (2013).

56. Levine, Z. A., Larini, L., Lapointe, N. E., Feinstein, S. C. & Shea, J. Regulation and aggregation of intrinsically disordered peptides. (2015).
57. D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. G., A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C., Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I., Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. S. & R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. X. and P. A. K. AMBER 2016. (2016).
58. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* **78**, 1950–1958 (2010).
59. Onufriev, A., Bashford, D. & Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins Struct. Funct. Genet.* **55**, 383–394 (2004).
60. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
61. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).
62. DeLano, W. L. The PyMOL Molecular Graphics System. *Schrödinger LLC wwwpymol.org Version 1.*, <http://www.pymol.org> (2002).
63. Dickson, C. J. *et al.* Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **10**, 865–879 (2014).
64. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).

65. Abraham, M., Hess, B., Spoel, D. van der & Lindahl, E. GROMACS User Manual version 5.0.7. *Www.Gromacs.Org* (2015).
66. McLendon, C. *et al.* Cell-free assays for gamma-secretase activity. *FASEB J.* **14**, 2383–6 (2000).
67. Frånberg, J. *et al.* Rat brain γ -secretase activity is highly influenced by detergents. *Biochemistry* **46**, 7647–7654 (2007).
68. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

APÉNDICE A. Marco teórico complementario

A.1. El algoritmo de Verlet

La solución aparentemente inmediata del problema de integración es usar una simple expansión de Taylor, como la siguiente:

$$\mathbf{r}_i(t_0 + \Delta t) = \mathbf{r}_i(t_0) + \frac{d\mathbf{r}_i(t_0)}{dt}\Delta t + \frac{1}{2}\frac{d^2\mathbf{r}_i(t_0)}{dt^2}\Delta t^2 + O(\Delta t^3) \quad (1)$$

Sin embargo, es un algoritmo inestable e inexacto. Una mejor solución fue propuesta por Verlet.¹ Si se suman las expansiones de Taylor para $+\Delta t$ y $-\Delta t$, se obtiene:

$$\mathbf{r}_i(t_0 + \Delta t) = \mathbf{r}_i(t_0 - \Delta t) + 2\mathbf{r}_i(t_0) + \mathbf{a}_i(t_0)\Delta t^2 + O(\Delta t^4) \quad (2)$$

Y las velocidades se pueden calcular de la siguiente manera:

$$\mathbf{v}_i(t_0) = \frac{1}{2\Delta t}[\mathbf{r}_i(t_0 + \Delta t) - \mathbf{r}_i(t_0 - \Delta t)] \quad (3)$$

El error intrínseco o de truncamiento local es $O(\Delta t^4)$.

Las características principales del algoritmo de Verlet son: a) es reversible en el tiempo; b) es simpléctico (conserva el volumen del espacio de fase); c) es de bajo orden en el tiempo, por lo tanto, permite tiempos largos; d) requiere sólo una evaluación de la fuerza por paso; e) es fácil de programar.^{2,3}

A.2. Algoritmo de Metropolis

En el método clásico de Metropolis⁴, las configuraciones son generadas de un estado previo utilizando una probabilidad de transición que depende de la diferencia de energía entre los estados final e inicial. La secuencia de estados producidos sigue un orden temporal, pero al tiempo en este caso se le refiere como “tiempo de Monte Carlo” y es no determinista. Para modelos de relajación, como el modelo de Ising, el comportamiento dependiente del tiempo se describe por la ecuación (4),

$$\frac{\partial P_n(t)}{\partial t} = - \sum_{n \neq m} [P_n(t)W_{n \rightarrow m} - P_m(t)W_{m \rightarrow n}] \quad (4)$$

en la que $P_n(t)$ es la probabilidad del sistema de estar en el estado n al tiempo t , y $W_{n \rightarrow m}$ es la probabilidad de transición de $n \rightarrow m$. En el equilibrio $\partial P_n(t)/\partial t = 0$ y los dos términos en el segundo miembro de la ecuación son iguales (ec. (5)).

$$P_n(t)W_{n \rightarrow m} = P_m(t)W_{m \rightarrow n} \quad (5)$$

Si el estado n se produce del estado m , la probabilidad relativa es la razón de las probabilidades individuales y el denominador de la función de probabilidad se cancela; como resultado, solo la diferencia de energía entre los dos estados es necesaria:

$$\Delta E = E_n - E_m \quad (6)$$

Cualquier probabilidad de transición que satisfaga la ecuación (5) es aceptable. El primer criterio de aceptación que fue usado en Física Estadística es el criterio de Metropolis⁴:

$$\begin{aligned} W_{m \rightarrow n} &= \tau_0^{-1} e^{-\frac{\Delta E}{k_B T}}, & \Delta E > 0 \\ W_{m \rightarrow n} &= \tau_0^{-1}, & \Delta E < 0 \end{aligned} \quad (7)$$

donde τ_0 es el tiempo requerido para hacer un intento de transición de estado, a menudo es igual a la unidad por lo que comúnmente se suprime en las ecuaciones.⁵

Algoritmo de Metropolis

- Se elige un estado inicial m con energía E_m
- Se genera un estado n mediante un mecanismo de perturbación
- Se calcula ΔE ,
 - Si $\Delta E < 0$, se acepta el cambio,
 - Si no, se genera un número aleatorio r , tal que $0 < r < 1$
 - Si $r < \exp\left(-\frac{\Delta E}{k_B T}\right)$, se acepta el cambio con la probabilidad $W_{m \rightarrow n}$

A.3. Monte Carlo

El método de Monte Carlo fue desarrollado por von Neumann, Ulam y Metropolis al final de la Segunda Guerra Mundial para estudiar la difusión de neutrones en material fisionable. El nombre "Monte Carlo", fue elegido por el uso extensivo de números aleatorios en el cálculo, fue acuñado por Metropolis en 1947 y usado en el título de un artículo describiendo los primeros trabajos en Los Alamos. ^{6,7}

En una simulación de Monte Carlo se intenta seguir la "dependencia temporal" de un modelo para el cual el cambio, o crecimiento, no procede de una manera rigurosamente predefinida, sino más bien de una manera estocástica que depende de una secuencia de números aleatorios que se genera durante la simulación. Con una segunda secuencia diferente de números aleatorios, la simulación no dará resultados idénticos, pero dará valores promedio similares a los obtenidos de la primera secuencia.

Una gran variedad de diferentes fenómenos físicos puede explorarse utilizando métodos de Monte Carlo. Se pueden considerar modelos que pueden ser discretizados de forma natural o por aproximación. El movimiento de los átomos individuales se puede examinar directamente; por ejemplo, puede ser modelado el salto aleatorio de átomos a sitios vecinos en una aleación metálica binaria (AB) donde interesa la cinética de interdifusión o desunión. También se puede modelar por el método de Monte Carlo la agregación limitada por difusión de partículas coloidales, ya que su movimiento está bien descrito por el movimiento Browniano clásico y aleatorio.

Las grandes colecciones de partículas clásicas interactuantes son directamente susceptibles de simulación de Monte Carlo. También el comportamiento de la interacción de partículas cuantizadas se estudia por este método, ya sea transformando el sistema en un modelo pseudoclásico o considerando directamente las propiedades de permutación.⁵

En contraste con una simulación de Dinámica Molecular, en una simulación de Monte Carlo no hay contribución de momento; el cálculo de cuantificables termodinámicos se lleva a cabo con base termodinámica estadística.⁸

El equilibrio mecánico estadístico se basa en la idea de una función de partición que contiene toda la información esencial sobre el sistema considerado. La forma general de la función de partición para un sistema clásico es,

$$Z = \sum_{\text{todos los estados}} e^{-\mathcal{H}/k_B T} \quad (8)$$

donde \mathcal{H} es el Hamiltoniano para el sistema, T es la temperatura y k_B la constante de Boltzmann. La suma es sobre todos los estados del sistema, que depende del tamaño del sistema y el número de grados de libertad para cada partícula. Para sistemas que consisten de pocas partículas interactuantes, la función de partición se puede escribir exactamente y por lo tanto las propiedades del sistema pueden ser calculadas en forma cerrada. Hay pocos casos en los que es posible extraer resultados exactos para sistemas muy grandes de partículas interactuantes, pero en general la función de partición no puede ser evaluada en forma exacta.

La probabilidad de ocupación de un estado en particular del sistema es también determinada por la función de partición. Así, la probabilidad de que el sistema esté en un estado μ está dada por la ecuación (9),

$$P_\mu = e^{-\mathcal{H}(\mu)/k_B T} / Z \quad (9)$$

donde $\mathcal{H}(\mu)$ es el Hamiltoniano cuando el sistema está en el μ -ésimo estado. Esta probabilidad usualmente no es exactamente conocida por el denominador (Z); sin embargo, esta dificultad se puede evitar generando una cadena de estados de Markov, es decir, generando cada nuevo estado directamente del anterior. ⁵

A.4. Simulaciones de Dinámica Molecular en disolvente implícito generalizado de Born

Además del modelo propuesto por Baptista *et al.*, Walczak y Antosiewicz⁹ y Bürgi *et al.*¹⁰ también desarrollaron modelos involucrando estados de protonación discretos. Cada uno de estos usa el muestreo de Monte Carlo para seleccionar estados de protonación basados en diferencias de energía calculadas entre los posibles estados de protonación. El método de Baptista *et al.* y el de Walczak y Antosiewicz utilizan métodos de Poisson Boltzmann (PB) para calcular las energías de protonación, el primero empleando la dinámica newtoniana del disolvente explícito y el último, dinámica de constante dieléctrica de Langevin entre los pasos de Monte Carlo. Sin embargo, un inconveniente de estos enfoques es que los potenciales utilizados para la dinámica no son coherentes con los utilizados para la elección de los estados de protonación en los pasos de Monte Carlo. Bürgi *et al.* evitaron el problema de consistencia calculando las energías de transición utilizando la integración termodinámica (*TI*) bajo las mismas condiciones de disolvente explícito utilizadas para la dinámica. Sin embargo, los cálculos de *TI* perturban la trayectoria dinámica incluso cuando se rechaza el paso de Monte Carlo. Además, los cálculos de PB y *TI* son computacionalmente caros, limitando el número de pasos de Monte Carlo que se pueden evaluar.¹¹

Debido a los inconvenientes mencionados, Mongan *et al.* propusieron un modelo que utiliza solvatación implícita generalizada de Born (*GB*).¹¹ La misma electrostática *GB* se utiliza para calcular energías de transición de estados de protonación y para la dinámica, por lo que los potenciales son consistentes. Además, el cálculo de energías de transición usando *GB* es rápido, dado que no hay necesidad de equilibrio del disolvente. Mediante la comparación entre los valores experimentales de pKas de la lisozima de clara de huevo de gallina y los predichos por su modelo, demostraron que éste muestrea con precisión los estados de protonación y proporciona una base más realista físicamente para estudiar la dinámica de sistemas con grupos titulables.

Definición del sistema y ensamble

Para simplificar las ecuaciones, se asume que en el caso de un soluto no protonable el sistema puede ser representado por el ensamble canónico; en el que el sistema tiene un número de partículas N y un volumen V constantes, y está inmerso en un baño térmico con temperatura T constante. Por lo tanto, una diferencia de los potenciales químicos estándar del soluto se aproximará como la diferencia de las energías libres de Helmholtz.

La inclusión de la protonación se hace abriendo el sistema no protonable a los protones titulantes, lo que conduce a un conjunto semi gran canónico (SGC); se restringe la localización de los protones transitorios a los sitios protonables. El estado de protonación de un soluto con B sitios de unión para protones se representará como un vector $\mathbf{n} = (n_1, n_2, \dots, n_B)$, donde $n_s = 0$ o 1 indica que el sitio s está vacío u ocupado, respectivamente. Las coordenadas y momentos generalizados del soluto estarán representados por q y p , y los del disolvente por \bar{q} y \bar{p} . Por simplicidad, los grados de libertad de cada protón tituable se mantienen en el estado desprotonado, que se considera tiene una posición de protón "vacía". Por lo tanto, q y p se supone que son independientes de \mathbf{n} .

La función de distribución del sistema simulado está dada por

$$\rho(q, p, \bar{q}, \bar{p}, \mathbf{n}) = \frac{\exp[\beta\mu n - \beta H(q, p, \bar{q}, \bar{p}, \mathbf{n})]}{\sum_{\mathbf{n}'} \int \exp[\beta\mu n' - \beta H(q', p', \bar{q}', \bar{p}', \mathbf{n}')] dq' dp' d\bar{q}' d\bar{p}'}, \quad (10)$$

donde $\beta = 1/(k_B T)$, μ es el potencial químico de los protones, $n = \sum_s n_s$ es el total de sitios ocupados, y H es el Hamiltoniano del sistema. Excepto por la ausencia de los términos semiclásicos que corrigen la indistinguibilidad de las partículas y la discretización del espacio de fase, el denominador de esta ecuación es igual a la función de partición SGC del sistema. El Hamiltoniano H consiste de una parte clásica H_c y una parte cuántica H_q ; H_q explica la formación y ruptura de los enlaces que implican los protones titulantes, mientras que los términos de enlace y no enlace restantes están contenidos en H_c . Como es común en las simulaciones de DM clásicas, se supone que la parte cuántica no depende de las coordenadas y momentos, $H_q = H_q(\mathbf{n})$. Debido al tratamiento de

los protones "vacíos" descritos anteriormente, la parte clásica depende de la protonación sólo a través de la energía potencial U

$$H_c(q, p, \bar{q}, \bar{p}, \mathbf{n}) = U(q, \bar{q}, \mathbf{n}) + K(p, \bar{p}) \quad (11)$$

La independencia total de la energía cinética K sobre \mathbf{n} puede lograrse conservando la masa de protones para la posición del protón vacío, que sólo afecta al cálculo de las propiedades temporales.

Muestreo condicional

Cuando se considera que el soluto tiene un estado de protonación fijo, el sistema sigue la distribución canónica

$$\rho(q, p, \bar{q}, \bar{p} | \mathbf{n}) = \frac{\exp[\beta\mu n - \beta H(q, p, \bar{q}, \bar{p}, \mathbf{n})]}{\sum_{\mathbf{n}'} \int \exp[-\beta H_c(q', p', \bar{q}', \bar{p}', \mathbf{n}')] dq' dp' d\bar{q}' d\bar{p}'}, \quad (12)$$

donde el término H_q ha sido cancelado. Cuando el soluto se considera fijo en el espacio, la distribución de estados de protonación a un pH dado se puede escribir

$$\rho(\mathbf{n} | q) = \frac{\exp[-n \ln(10) pH - \beta \Delta G^\circ(q, \mathbf{n})]}{\sum_{\mathbf{n}'} \exp[-n' \ln(10) pH - \beta \Delta G^\circ(q, \mathbf{n}')] } \quad (13)$$

donde $\Delta G^\circ(q, \mathbf{n})$ es la energía libre estándar de la reacción de protonación $0 \rightarrow \mathbf{n}$ con q fija; $\Delta G^\circ(q, \mathbf{n})$ y $\rho(\mathbf{n} | q)$ dependen sólo de la conformación del soluto. Estas energías libres se pueden calcular usando métodos *CE*, y el muestreo llevarse a cabo con un método de Monte Carlo. En este tipo de métodos, la respuesta dieléctrica del disolvente está implícitamente incluida por lo que, cada movimiento de MC corresponde a una transición de un estado \mathbf{n} con disolvente relajado a otro estado \mathbf{n}' con disolvente relajado.

El estado de protonación del soluto cambia aleatoriamente (o periódicamente), después de lo cual se ejecuta un segmento de DM sobre el disolvente, manteniendo las coordenadas del soluto fijas en el estado q . Esto corresponde, por definición, a un movimiento de prueba de Monte Carlo híbrido (*HMC*), donde el cambio estocástico del momento ha sido reemplazado por el cambio estocástico de \mathbf{n} . Si se elige un tiempo de relajación del disolvente adecuado, cada movimiento *HMC* toma el sistema de un estado

\mathbf{n} con disolvente relajado a otro estado \mathbf{n}' con disolvente relajado. También, como ha sido mencionado en el párrafo anterior, la estabilidad relativa de estos dos estados se puede estimar usando el enfoque basado en CE , incluso antes de que se produzca la relajación del disolvente. Obviamente, esta relajación debe hacerse después, aunque ya no sea necesaria la energía del estado relajado final. El procedimiento propuesto es: (i) seleccionar un nuevo estado \mathbf{n} desde un movimiento CE/MC o una simulación completa; y (ii) hacer una relajación de DM del disolvente, fijando al soluto en el estado q .

Este método combina los métodos de muestreo condicional anteriores con el fin de obtener un muestreo conjunto de coordenadas, momentos y estados de protonación. Se ejecuta una simulación de DM a temperatura constante, mientras que en tiempos aleatorios o periódicos el estado de protonación del sistema se cambia usando un movimiento de HMC estimado por CE . El procedimiento resulta en un muestreo apropiado siguiendo la Ec. (10); para la evaluación de estados de protonación a tiempos aleatorios o periódicos se muestrea la distribución SGC exacta $\rho(q, p, \bar{q}, \bar{p}, \mathbf{n})$ y no una distribución de campo medio, por lo que no hay pérdida de los efectos de la correlación sitio-sitio.¹²

A.5. Estudios sobre la β -secretasa

La β -secretasa (BACE-1) es una aspartil proteasa asociada a la membrana, cuyo sitio activo se encuentra en el ectodominio. Su estudio es de gran interés terapéutico debido a su participación en la primera escisión de la *APP*. La estructura de su ectodominio fue resuelta por primera vez mediante difracción de rayos X en el 2000,¹³ desde entonces se han obtenido un gran número de estructuras, por lo que además de ser un objeto de estudio experimental, se han realizado una gran cantidad y variedad de estudios computacionales para la identificación del estado de protonación de su díada catalítica, así como para la elucidación de su mecanismo de reconocimiento de sustrato y de escisión.

Entre los primeros estudios computacionales que se hicieron para la determinación de los estados de protonación de los aspárticos catalíticos de la BACE-1, Park y Lee¹⁴ hicieron dos simulaciones de DM independientes de la enzima unida al inhibidor OM99-2. Partieron de la premisa de que los dos aspárticos operan cooperativamente con una carga neta -1, por lo que en cada una de las simulaciones consideraron protonado uno de los aspárticos; de este modo determinaron que el estado más favorable es en el que Asp32 está protonado.

Rajamani y Reynolds¹⁵ llevaron a cabo el primer estudio de modelado mecánico cuántico de estados de protonación y localización de protones para el par catalítico de una aspartil proteasa en el que el sitio activo de la proteína fue incluido explícitamente. Determinaron que para la BACE-1 libre se favorece el estado didesprotonado (Asp32⁻, Asp228⁻), mientras que cuando está unida a OM99-2 se favorece un estado monoprotonado (Asp32⁻, Asp228). Este estudio lo llevaron a cabo empleando un método cuántico semiempírico e incluyendo solvatación PB.

Polgár y Keserü¹⁶ calcularon el pKa de los aspárticos catalíticos de la BACE-1 para demostrar la importancia del estado de protonación para el cribado virtual. Partieron de dos estructuras cristalizadas a pH 6.5 y 7.4 (cristalizada con OM99-2) y utilizaron un solucionador de la ecuación de PB. Obtuvieron que a pH 6.5 y 7.4 sólo Asp228 está ionizado (Asp32, Asp228⁻) y, por comparación entre las curvas de titulación de las

estructuras cristalizadas con y sin ligando, argumentaron que el estado de protonación puede cambiar por la unión del ligando.

Yu *et al.*¹⁷ llevaron a cabo la asignación de estados de protonación de los aspárticos catalíticos de la BACE-1 por refinamiento de la estructura de rayos X del complejo de BACE-1 con OM99-2, usando QM/MM. En su estudio, los átomos del sitio activo fueron modelados por mecánica cuántica y los otros átomos por mecánica molecular. Consideraron 8 configuraciones posibles de la díada aspártica y llevaron a cabo el refinamiento para cada una de ellas. Los resultados de sus cálculos sugieren la configuración monoprotonada (Asp32, Asp228⁻), en concordancia con Park y Lee y Polgár y Keserü. En la configuración propuesta, como en la mayoría de los estudios llevados a cabo, el protón está orientado por la formación de puente de hidrógeno carboxilo-carboxilo.

Además del estudio computacional, José L. Domínguez *et al.*¹⁸ determinaron experimentalmente el efecto del pH sobre las afinidades de un conjunto de inhibidores a la región del ectodominio de la BACE-1. Llevaron a cabo esta determinación mediante un ensayo con un biosensor de resonancia de plasmón superficial, con el cual calcularon las curvas de titulación para un conjunto de complejos para entender las interacciones moleculares. Uno de sus resultados más llamativos es que para los inhibidores peptídicos con isómeros de hidroxietileno (HE) predijeron que la díada Asp será diprotonada a pH bajo, en contraposición a todos los estudios previos que apuntaron a un estado monoprotonado. En un estudio posterior,¹⁹ usando MM/MD muestran que el estado diprotonado reproduce el *ranking* de unión de un conjunto de inhibidores mejor que el estado monoprotonado propuesto por estudios previos, un resultado que proporciona soporte adicional a la asignación de carga de la BACE-1 cuando se une a inhibidores análogos del estado de transición.

Barman y Prabhakar llevaron a cabo un estudio computacional basado en la combinación de Dinámica Molecular y DFT²⁰ para estudiar las interacciones de la APP silvestre y de su variante sueca con la BACE-1, así como su mecanismo de escisión. En este estudio se encontraron 2 aguas formando una red de puentes de hidrógenos con

los aspárticos catalíticos, como había sido propuesto en un análisis de estructuras cristalográficas de aspartil proteasas²¹. Posteriormente, hicieron un estudio de acoplamiento molecular²² en el que 8 configuraciones (diferentes estados de protonación y localización de protones) de la díada Asp de la BACE-1 fueron acoplados con 8 inhibidores diferentes; se encontró que 5 de los inhibidores acoplados favorecen dos diferentes estados monoprotonados y los otros 3 un estado didesprotonado.

Cafilisch y Mishra²³ estudiaron mediante DM la dinámica conformacional y energías libres de superficies de la BACE-1 en tres pasos del mecanismo catalítico. Gueto-Tettay *et al.*²⁴ publicaron un estudio más reciente, utilizando también Dinámica Molecular, en el que llevaron a cabo una serie de simulaciones de la BACE-1 acoplada con inhibidores análogos al estado de transición; a partir de sus resultados proponen que el estado didesprotonado es el más preferible en presencia de estos inhibidores, contradiciendo las propuestas anteriores de estados monoprotonados y diprotonados.

Lo más cercano a un consenso de los múltiples estudios del efecto de los inhibidores en el estado de protonación de la díada Asp de la BACE-1 es que la naturaleza química del inhibidor juega un papel importante en la determinación del estado de protonación.

A.6. Herramientas de análisis

RMSD

La desviación cuadrática media (*RMSD*) es la medida cuantitativa más comúnmente utilizada de la similitud entre dos coordenadas atómicas superpuestas.²⁵ Los valores de *RMSD* se presentan en Å y se calculan por

$$RMSD = \sqrt{\frac{1}{M} \sum_{i=1}^n m_i |r_i(t_j) - r_i^{ref}|^2} \quad (14)$$

Donde el promedio se realiza sobre los n pares de átomos equivalentes, $r_i(t_j)$ es la posición del átomo i al tiempo t_j y r_i^{ref} es la posición de referencia, m_i la masa del átomo i y M la masa total. Cuando el *RMSD* no es ponderado por la masa, $m_i = 1$ y $M = n$. El *RMSD* se puede calcular para cualquier tipo y subconjunto de átomos; por ejemplo, los C α de la proteína completa o de un subconjunto específico, todos los átomos pesados de un subconjunto específico de residuos o todos los átomos pesados en un ligando de molécula pequeña.²⁵

RMSF

La fluctuación cuadrática media (*RMSF*) es una medida de la dinámica de cada átomo o residuo en la estructura de la proteína. Se calcula del segundo momento de la fluctuación de la posición atómica a lo largo de la trayectoria²⁶ usando la fórmula:

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t_j=1}^T |r_i(t_j) - r_i^{ref}|^2} \quad (15)$$

Donde T es el tiempo sobre el que se quiere promediar.

Monitoreo de la evolución de los estados de protonación

Este monitoreo se puede llevar a cabo utilizando series temporales de la fracción protonada (o desprotonada) calculada por promedios acumulativos, corrientes o por “trozos”.²⁷

Promedios acumulativos

Un promedio acumulativo es una serie temporal cuyos valores en el tiempo t se calculan de acuerdo con

$$\langle A \rangle_t = \frac{\int_0^t A(t) dt}{t} \quad (16)$$

De manera que representa el valor promedio desde el tiempo 0 hasta t . Como $\langle A \rangle_t$ cambia rápidamente cuando t es pequeño, y muy lentamente a medida que t se hace grande. Esta serie temporal puede dar la impresión de que una propiedad se está convirtiendo en un valor particular cuando esta fluctúa mucho.

Promedios corrientes

Un promedio corriente es una serie temporal cuyos valores en el tiempo t se calculan de acuerdo con

$$\langle A \rangle_t = \frac{\int_{t-\sigma}^{t+\sigma} A(t) dt}{2\sigma} \quad (17)$$

Que representa el valor promedio de $t - \sigma$ a $t + \sigma$, 2σ se conoce como el tamaño de ventana. La ventaja de un promedio corriente sobre un promedio acumulativo es que la forma de la curva a grandes valores de t no depende de los valores cercanos a $t=0$.

Promedios por "trozos"

Un promedio corriente es una serie temporal en la que la trayectoria está segmentada en trozos separados de tamaño específico de 2σ pasos de tiempo. Estos promedios son un caso especial de promedios corrientes en los que los tiempos a los que se calcula $\langle A \rangle_t$ son los puntos centrales de los "trozos" de tiempo de tamaño de 2σ pasos de tiempo, por lo que cada punto se asigna sólo a un trozo y no hay superposición.

Un promedio corriente es una serie temporal en la que la trayectoria está segmentada en trozos separados de tamaño específico de 2σ pasos de tiempo. El valor promedio es calculado de acuerdo con la ecuación (17).

Radio de giro

El radio de giro (R_g) es un parámetro que se utiliza como un indicador de la compacidad de una estructura y se calcula en dos pasos. Primero, se determinan las coordenadas del centro de masa R_c , con la siguiente ecuación:²⁸

$$\sum m_i(r_i - R_c) = 0 \quad (18)$$

Donde m_i es la masa del i -ésimo átomo y r_i sus coordenadas. Si los átomos son tomados como puntos en un espacio 3D, el radio de giro se obtiene como

$$R_g = \sqrt{\frac{\sum m_i(r_i - R_c)^2}{\sum m_i}} \quad (19)$$

Clustering

Clustering es el proceso de particionar un conjunto de objetos en subconjuntos (denominados *clusters*) de manera que cada subconjunto contiene objetos similares y los objetos en diferentes subconjuntos son diferentes.²⁹ En simulación, este análisis permite identificar conformaciones similares a lo largo de la trayectoria. Existen muchos métodos de *clustering* dentro de los cuales uno de los más utilizados es el método de Daura *et al.*³⁰. Este método se basa en el *RMSD* entre todas las conformaciones muestreadas durante la simulación, se elige un *RMSD* de corte (*cutoff*) para determinar los miembros del *cluster*, se toma la estructura con la mayor cantidad de vecinos y se forma un *cluster*; el análisis reinicia con las estructuras restantes.

APÉNDICE B. Series temporales de las fracciones protonadas y curvas de titulación de los sistemas del R2/wt.

B.1. Monómero con solvatación GB

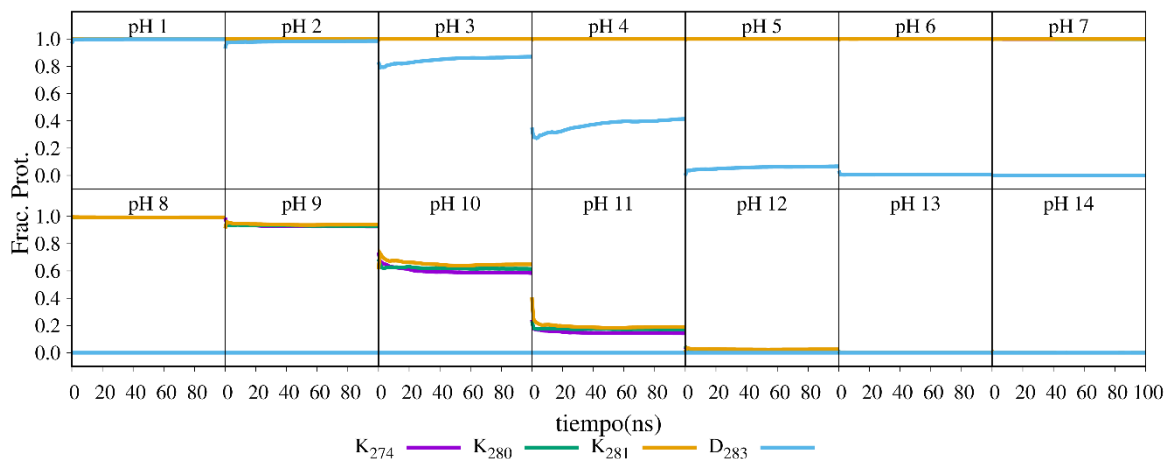


Figura B1. Series temporales de las fracciones de protonación calculadas por promedios acumulativos.

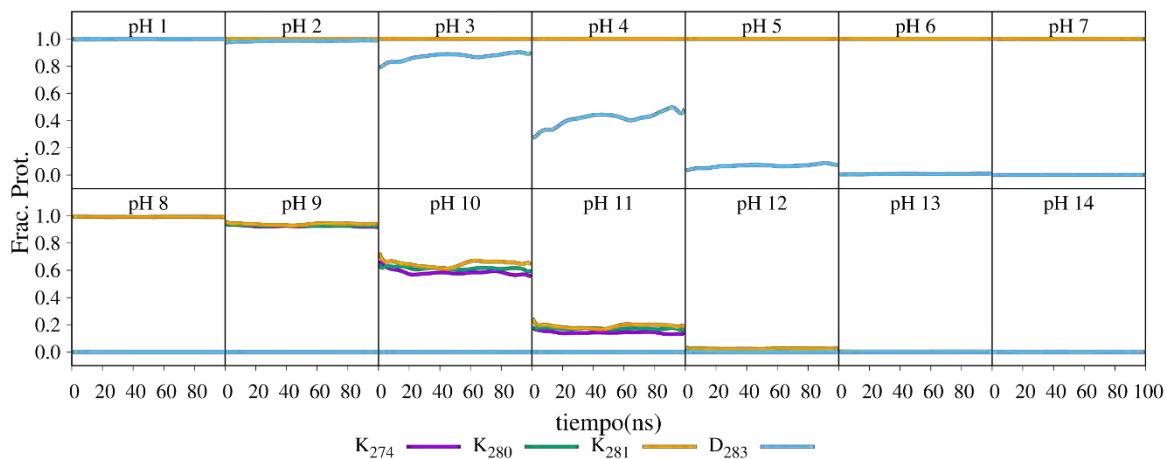


Figura B2. Series temporales de las fracciones de protonación calculadas por promedios por trozos de 1 ns.

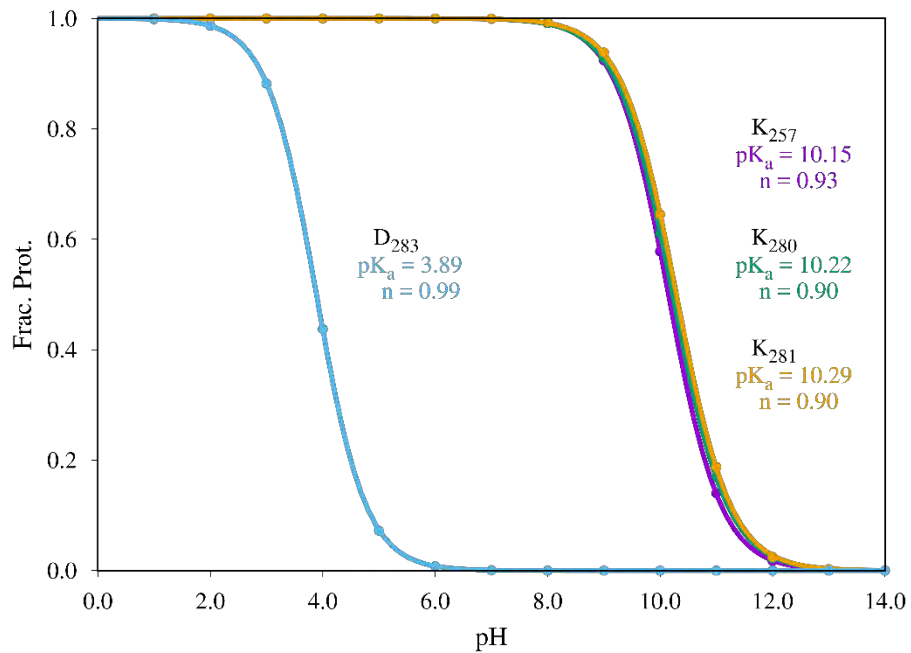


Figura B3. Curvas de titulación de los aminoácidos titulables ajustadas a la ecuación de Hill.

B.2. Monómero con agua explícita

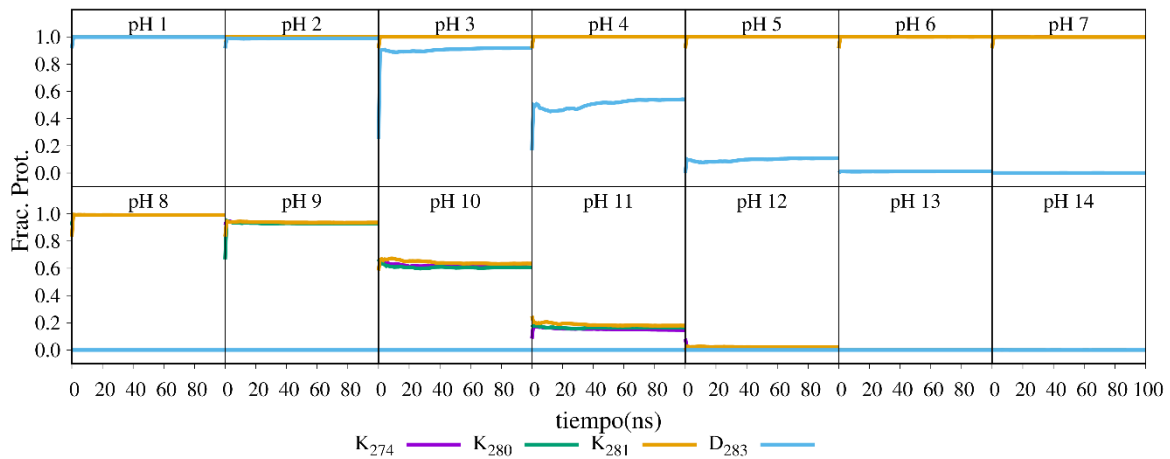


Figura B4. Series temporales de las fracciones de protonación calculadas por promedios acumulativos.

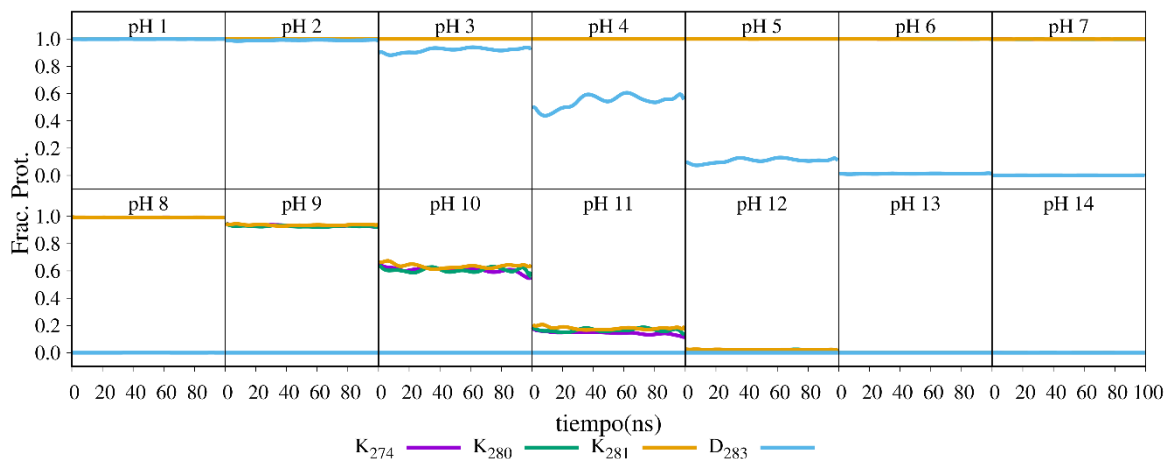


Figura B5. Series temporales de las fracciones de protonación calculadas por promedios por trozos de 1 ns.

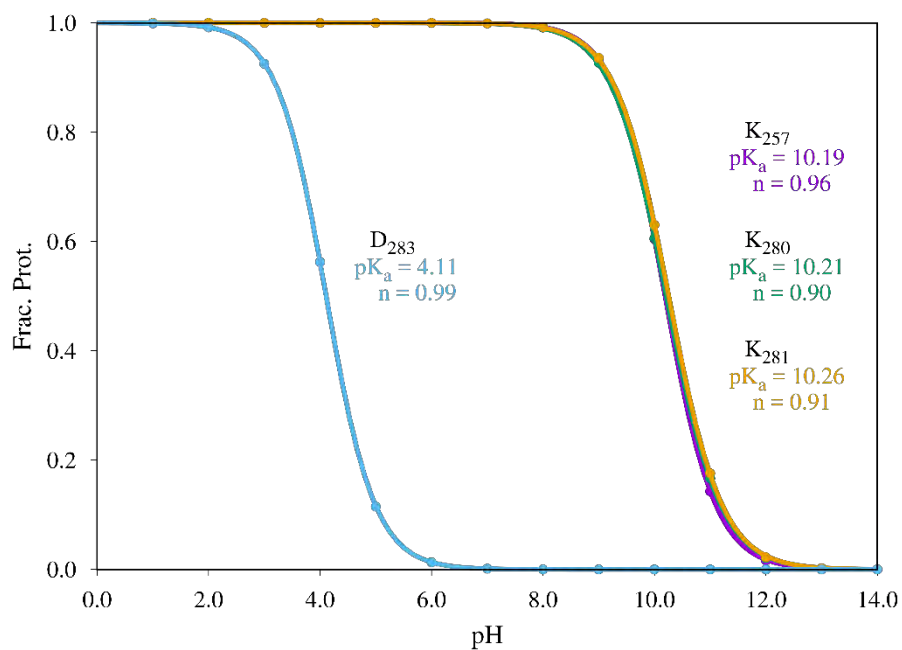


Figura B6. Curvas de titulación de los aminoácidos titulables ajustadas a la ecuación de Hill.

B.3. Dímero con solvatación GB

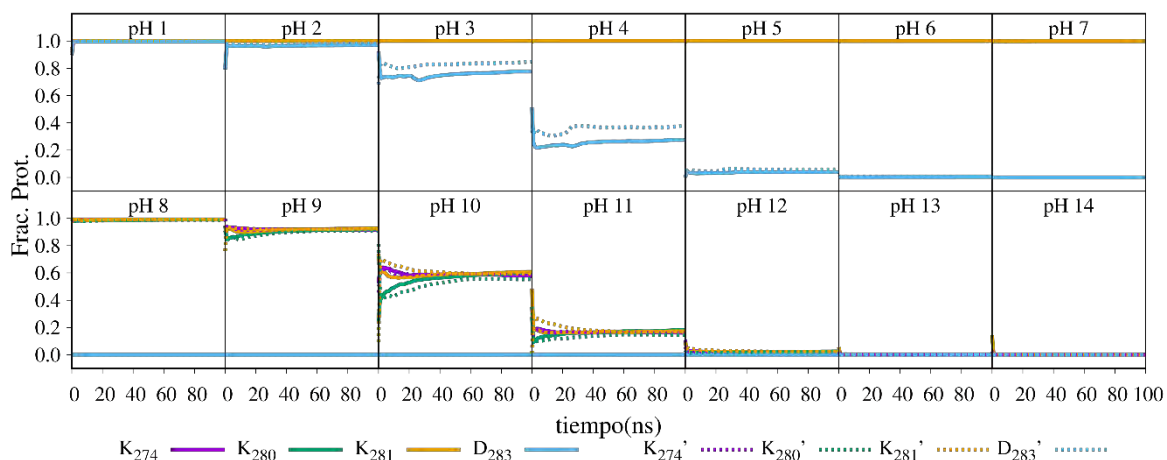


Figura B7. Series temporales de las fracciones de protonación calculadas por promedios acumulativos.

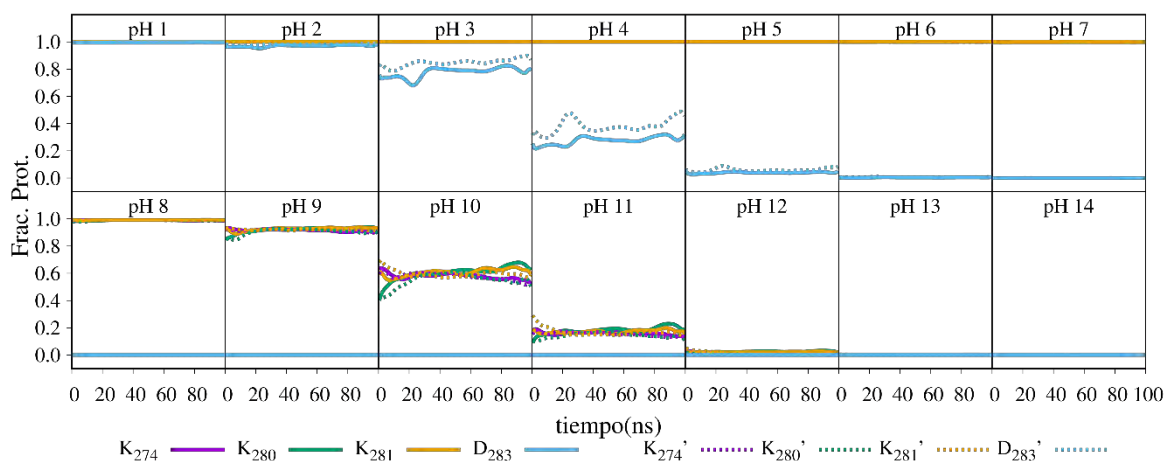


Figura B8. Series temporales de las fracciones de protonación calculadas por promedios por trozos de 1 ns.

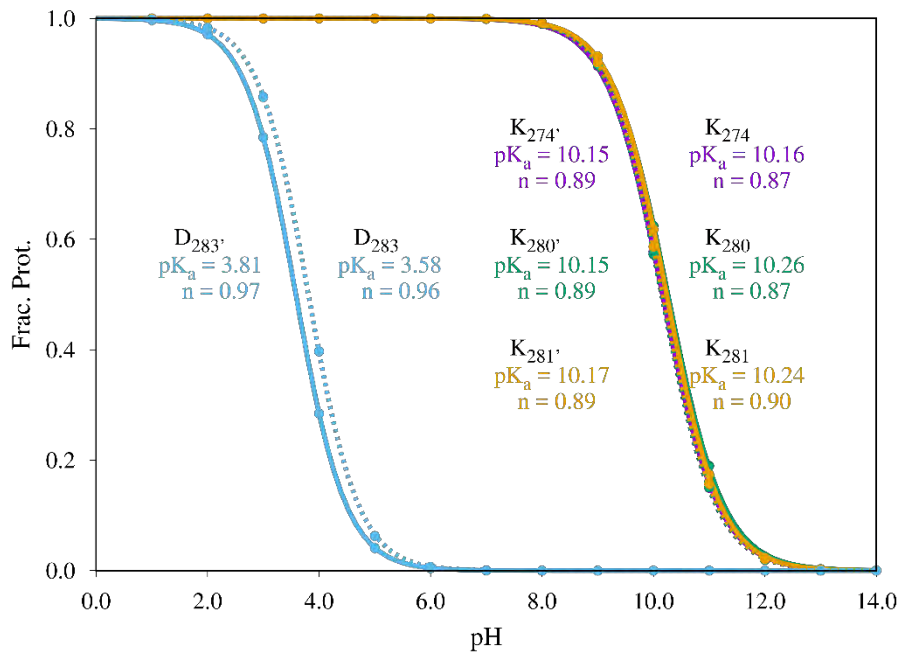


Figura B9. Curvas de titulación de los aminoácidos titulables ajustadas a la ecuación de Hill.

B.4. Dímero con agua explícita

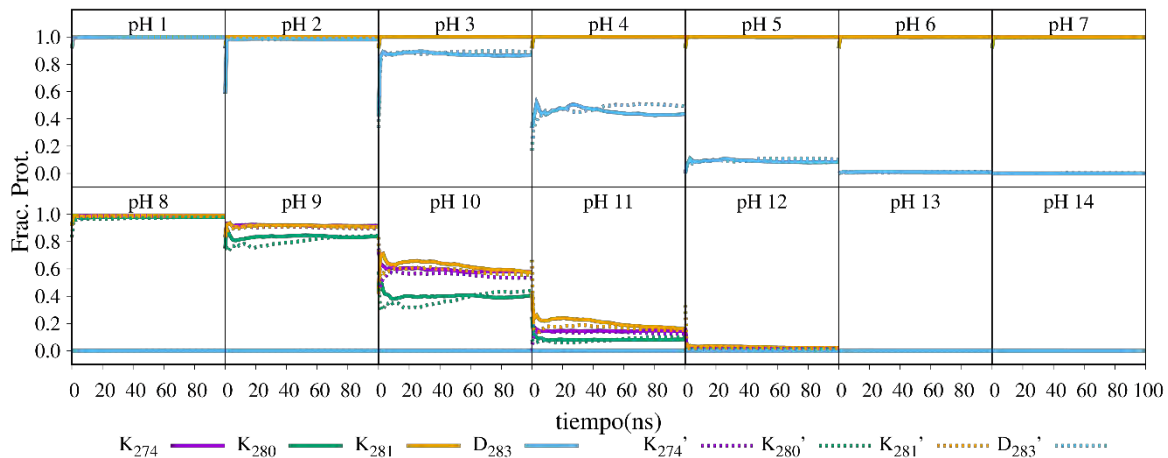


Figura B10. Series temporales de las fracciones de protonación calculadas por promedios acumulativos.

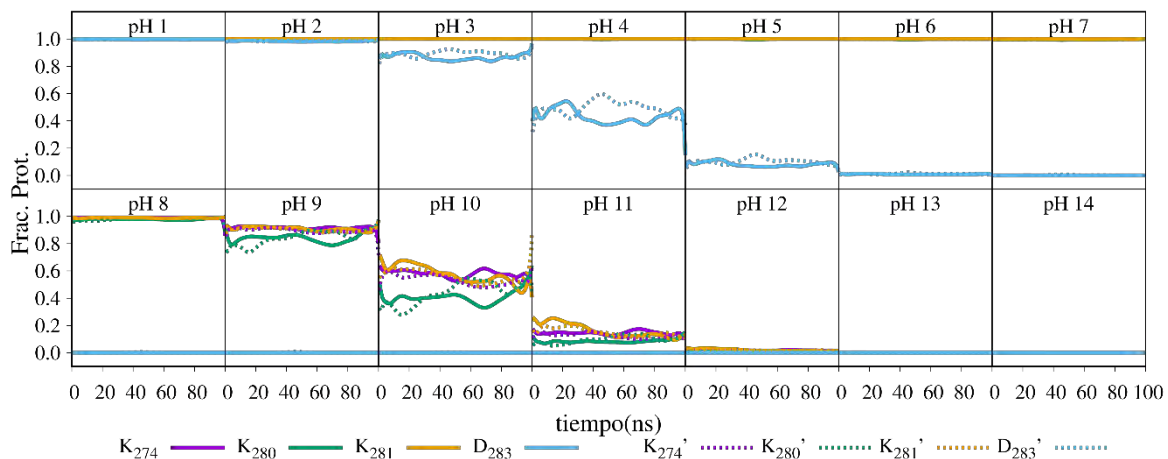


Figura B11. Series temporales de las fracciones de protonación calculadas por promedios por trozos de 1 ns.

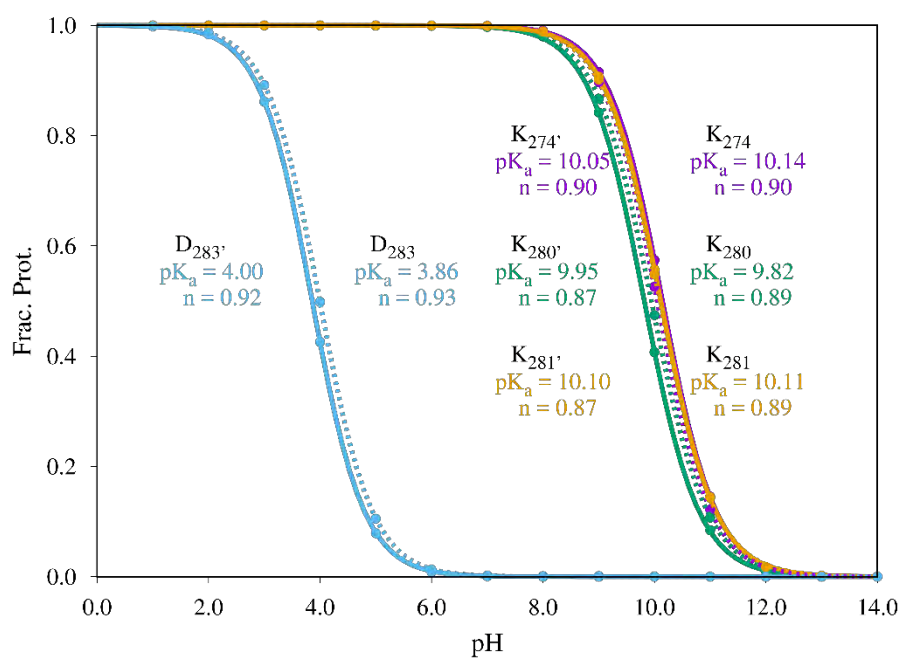


Figura B12. Curvas de titulación de los aminoácidos titulables ajustadas a la ecuación de Hill.

APÉNDICE C. Caracterización estructural del monómero R2/wt con solvatación GB.

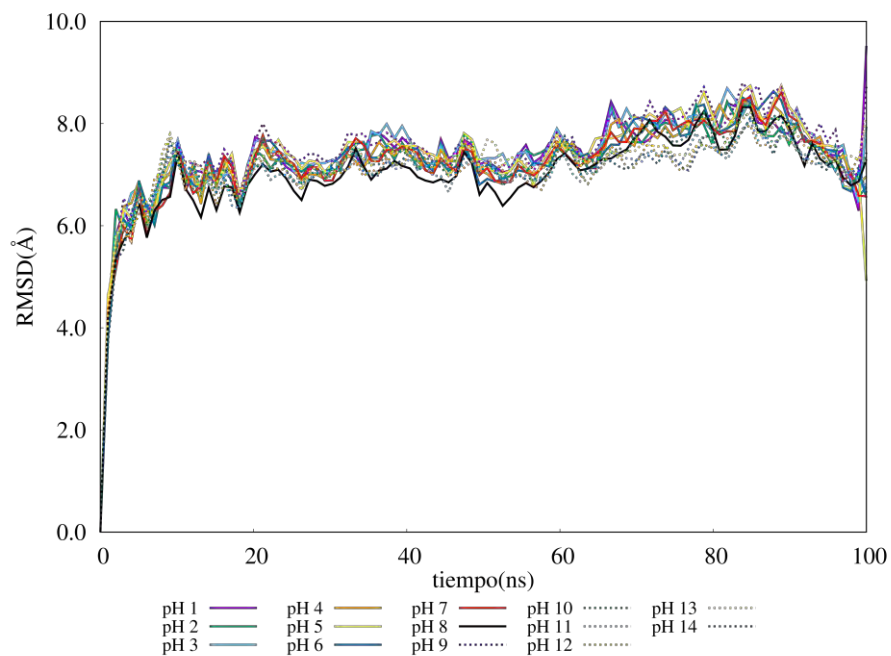


Figura C1. Desviación cuadrática media del *backbone* (N, C α , C) del R2/wt a lo largo de la trayectoria a cada pH.

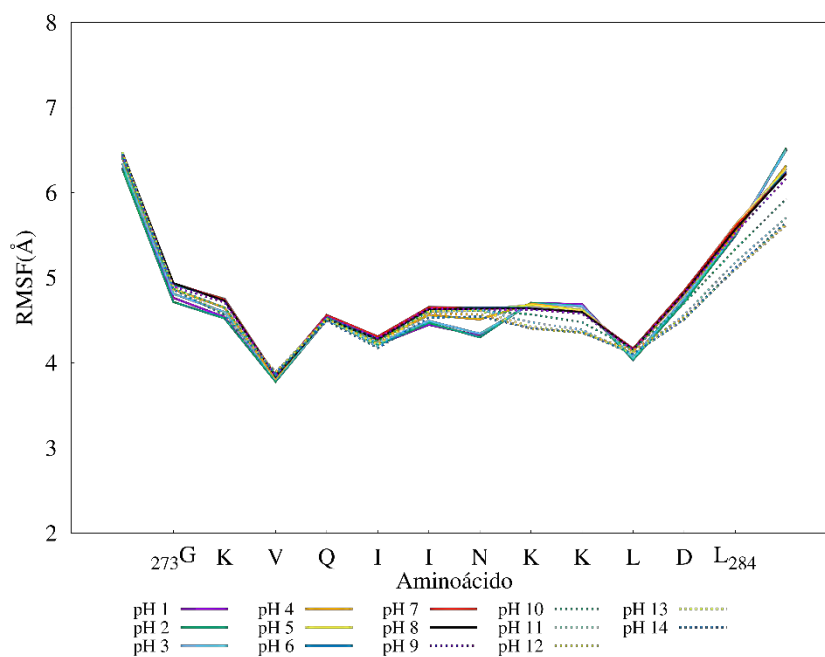


Figura C2. Fluctuación cuadrática media por aminoácido del R2/w.

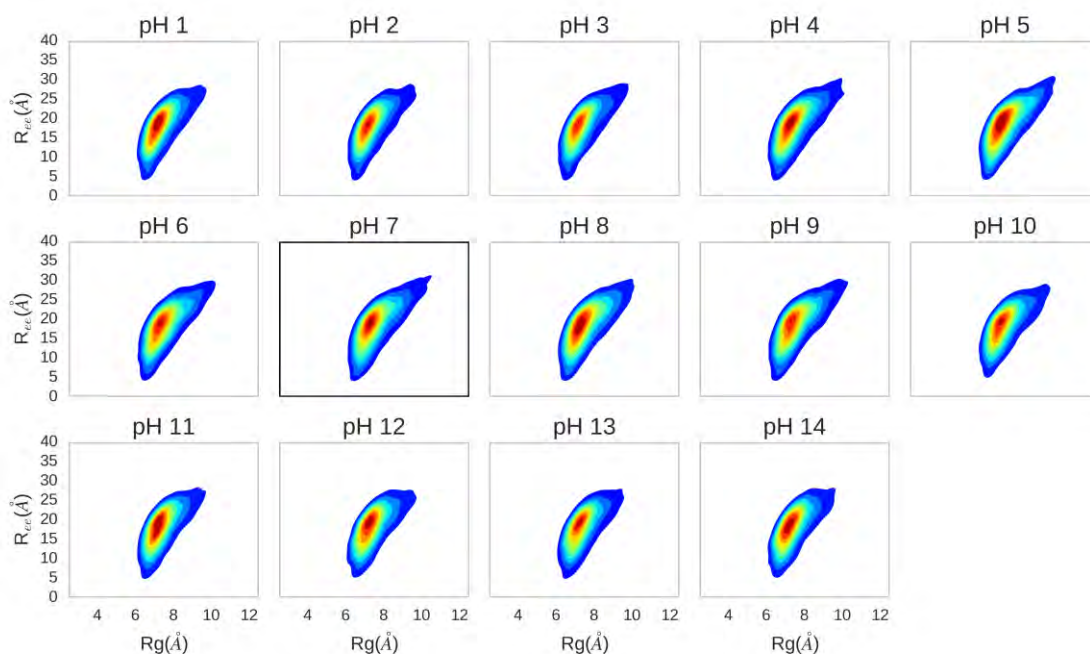


Figura C3. Densidades de población de conformaciones del péptido. R_{ee} es una medida de la distancia de extremo a extremo del péptido (N-terminal a C-terminal) y R_g es el radio de giro sobre el centro de masa. Las densidades altamente pobladas se indican en rojo, mientras que la población disminuye hacia azul.

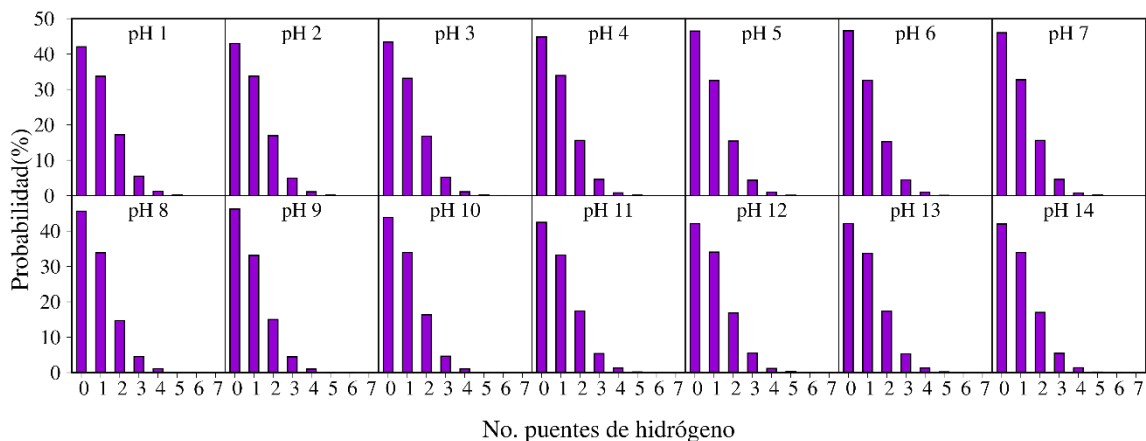


Figura C4. Probabilidades de formación de puentes de hidrógeno intramoleculares (C=O a N-H) sobre toda la trayectoria.

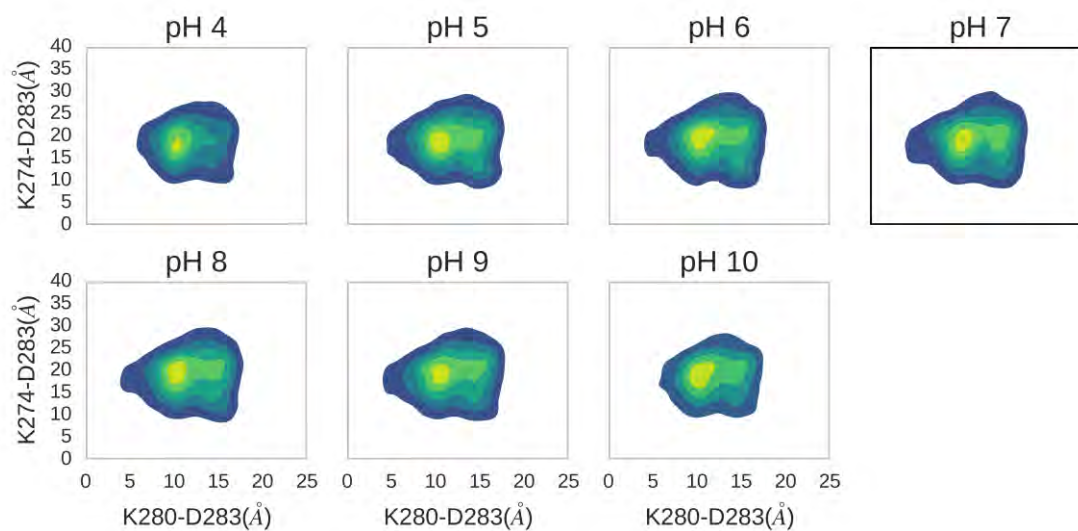


Figura C5. Densidades de población de puentes salinos monoméricos, éstos son representados por la proximidad relativa entre el aspártico y las lisinas. La densidad disminuye de amarillo a verde oscuro.

APÉNDICE D. Caracterización estructural del dímero de R2/wt con solvatación GB.

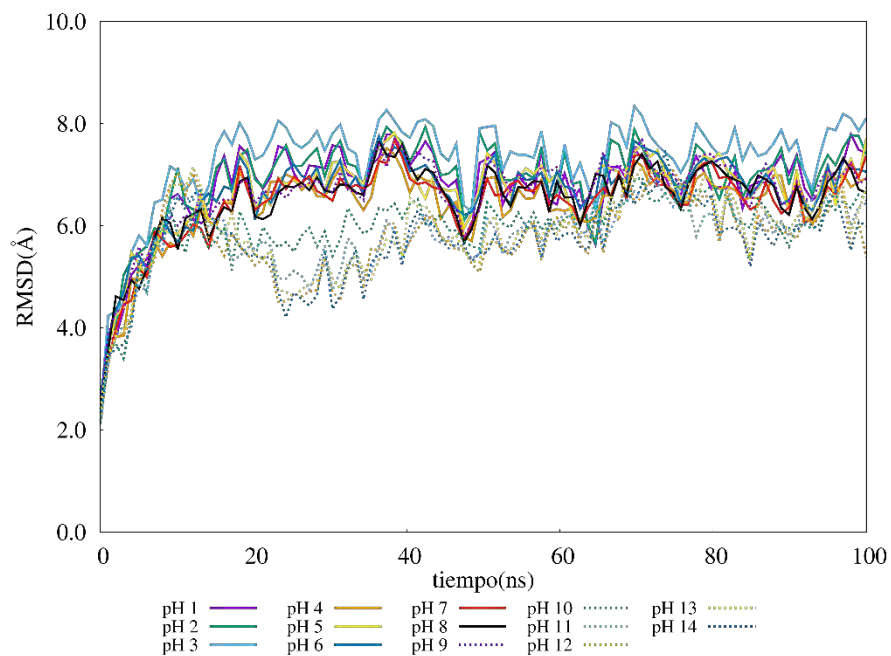


Figura D1. Desviación cuadrática media del *backbone* (N, C α , C) del dímero de R2/wt a lo largo de la trayectoria a cada pH.

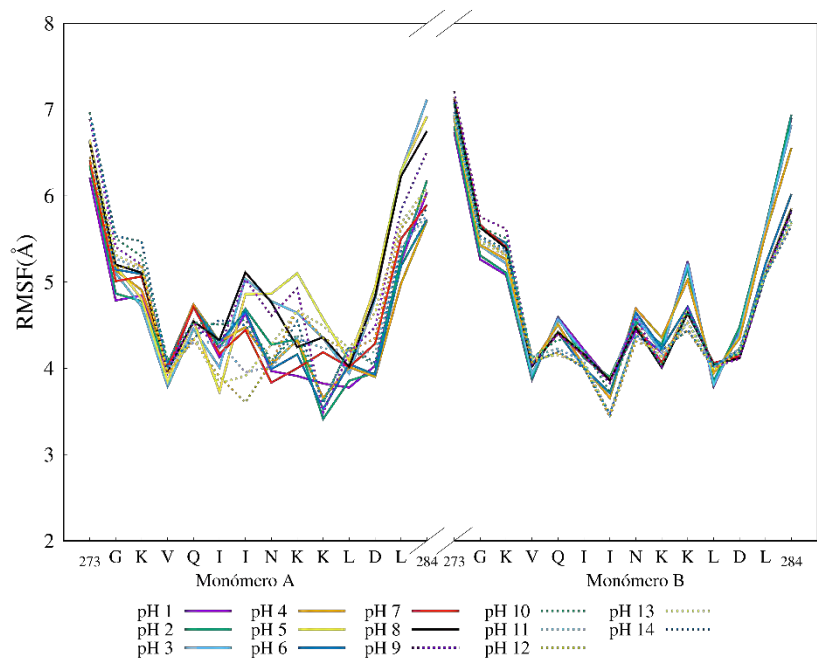


Figura D2. Fluctuación cuadrática media por aminoácido del dímero de R2/wt.

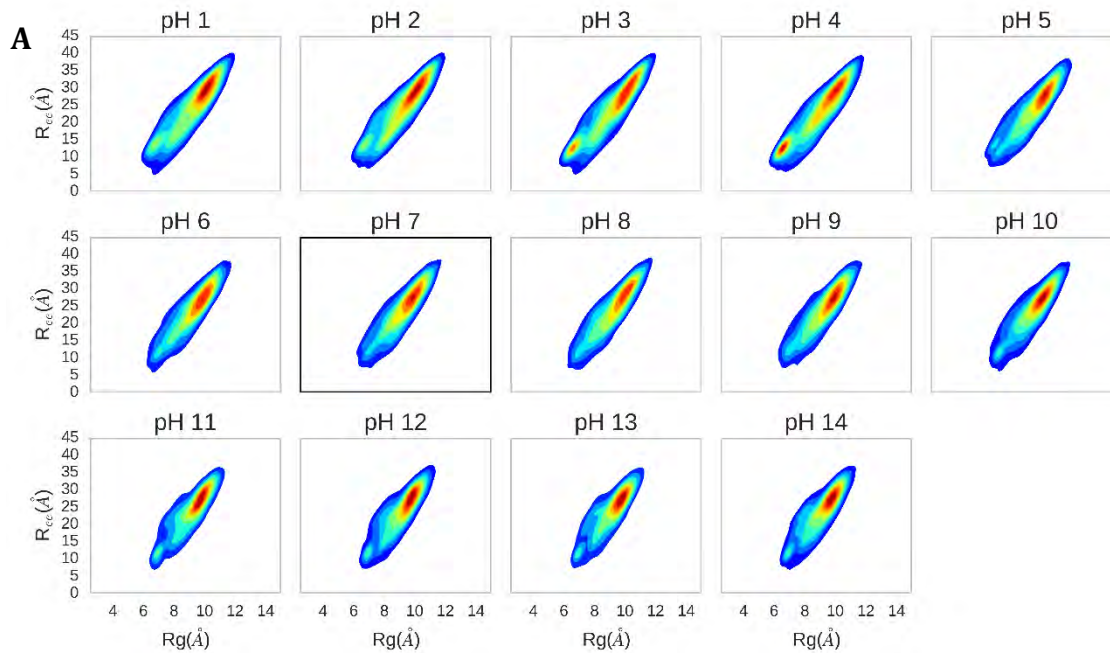


Figura D3. Densidades de población de conformaciones del monómero A. R_{ee} es una medida de la distancia de extremo a extremo del péptido y R_g es el radio de giro sobre el centro de masa.

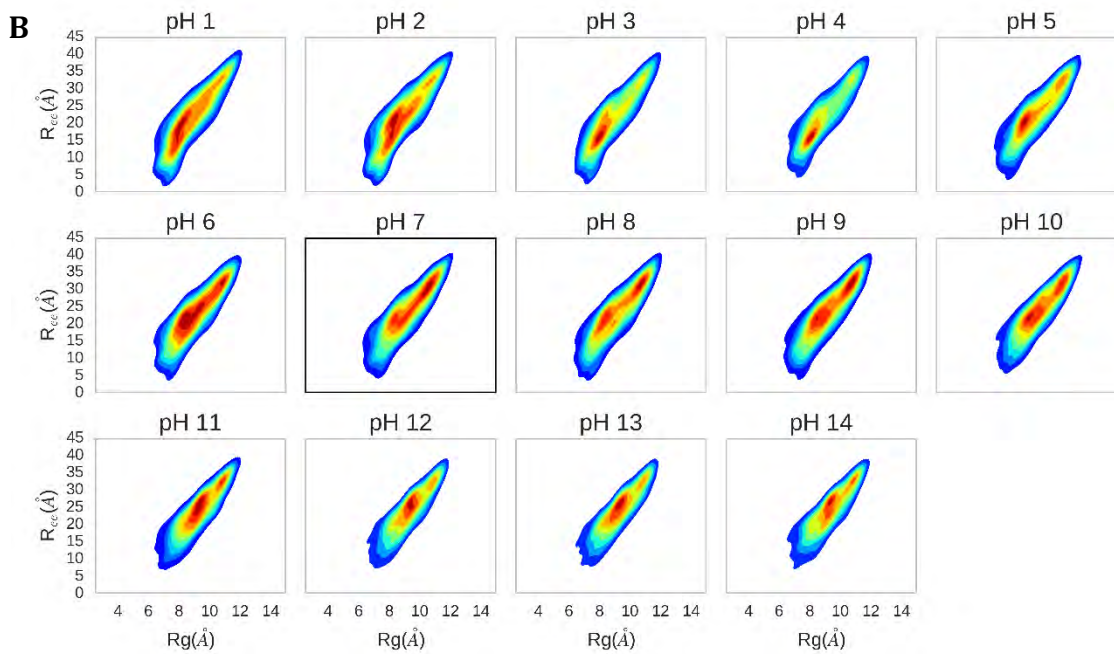


Figura D4. Densidades de población de conformaciones del monómero B.

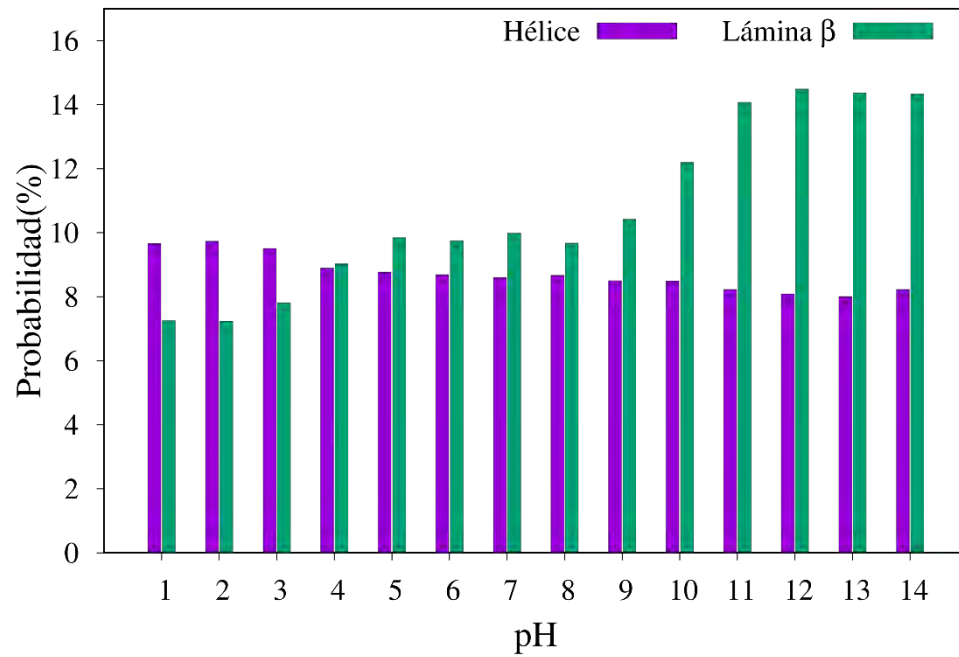
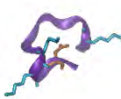































Figura D5. Probabilidad de formación de hélices y láminas β en el dímero de R2/wt. El análisis de estructura secundaria se llevó a cabo utilizando DSSP. En hélice están contenidas α , 3-10 y π -hélices y en lámina β , paralelas y antiparalelas.

APÉNDICE E. Estructuras más representativas del monómero y dímero de R2/wt simulados con agua explícita

pH 1			pH 8		
					
5.02%	4.80%	1.81%	1.44%	1.07%	1.02%
pH 2			pH 9		
					
4.85%	4.66%	1.88%	1.41%	1.35%	0.93%
pH 3			pH 10		
					
4.56%	3.83%	1.83%	2.02%	1.07%	1.04%
pH 4			pH 11		
					
2.40%	2.23%	1.13%	2.85%	1.21%	1.02%
pH 5			pH 12		
					
1.31%	1.06%	0.92%	3.49%	1.51%	1.09%

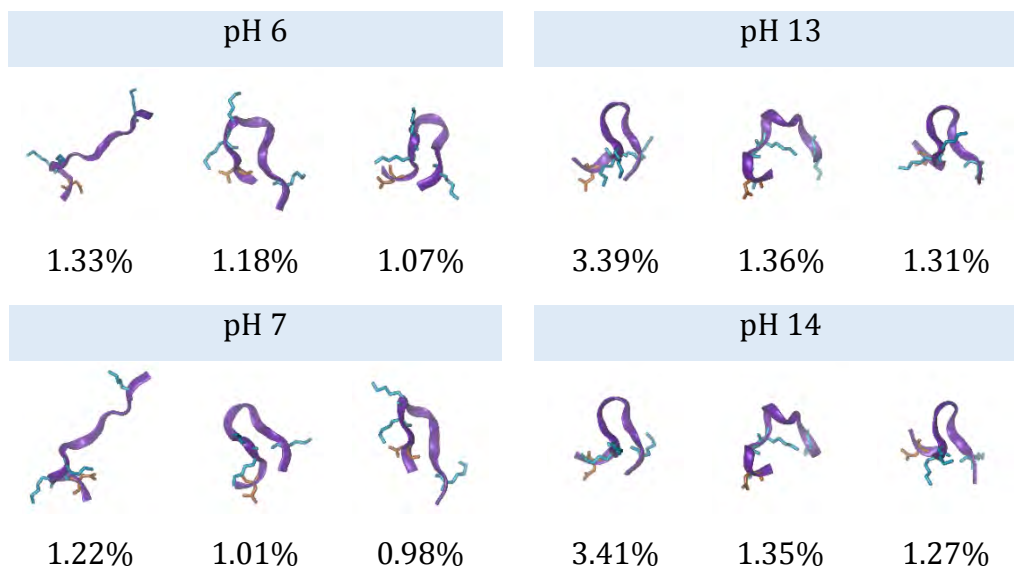
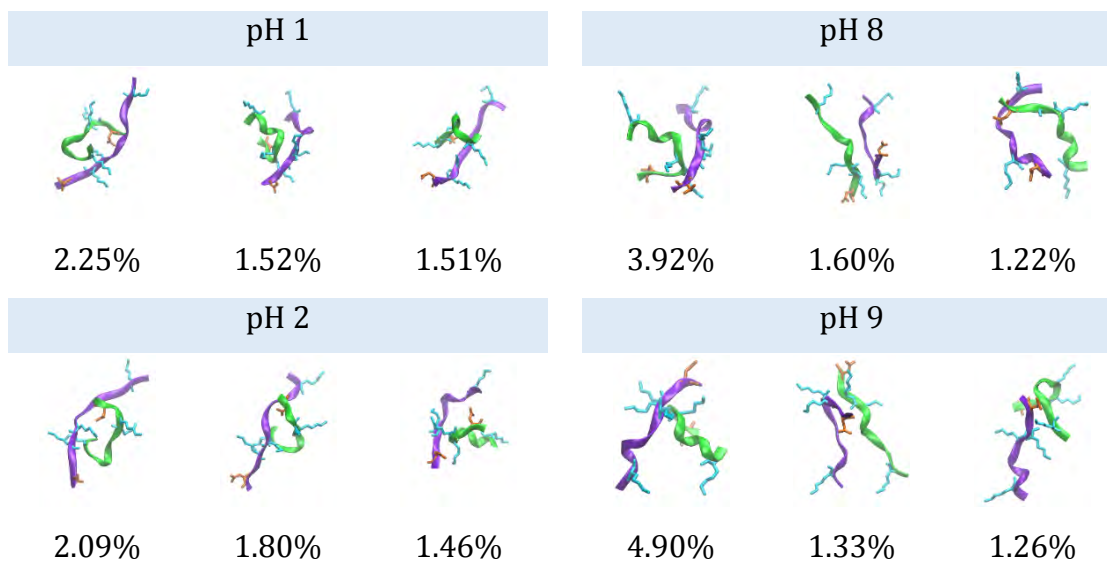


Figura E1. Estructuras más representativas del monómero simulado con agua explícita a cada pH. Las cadenas laterales del aspártico y las lisinas son representadas de color naranja y cian respectivamente.



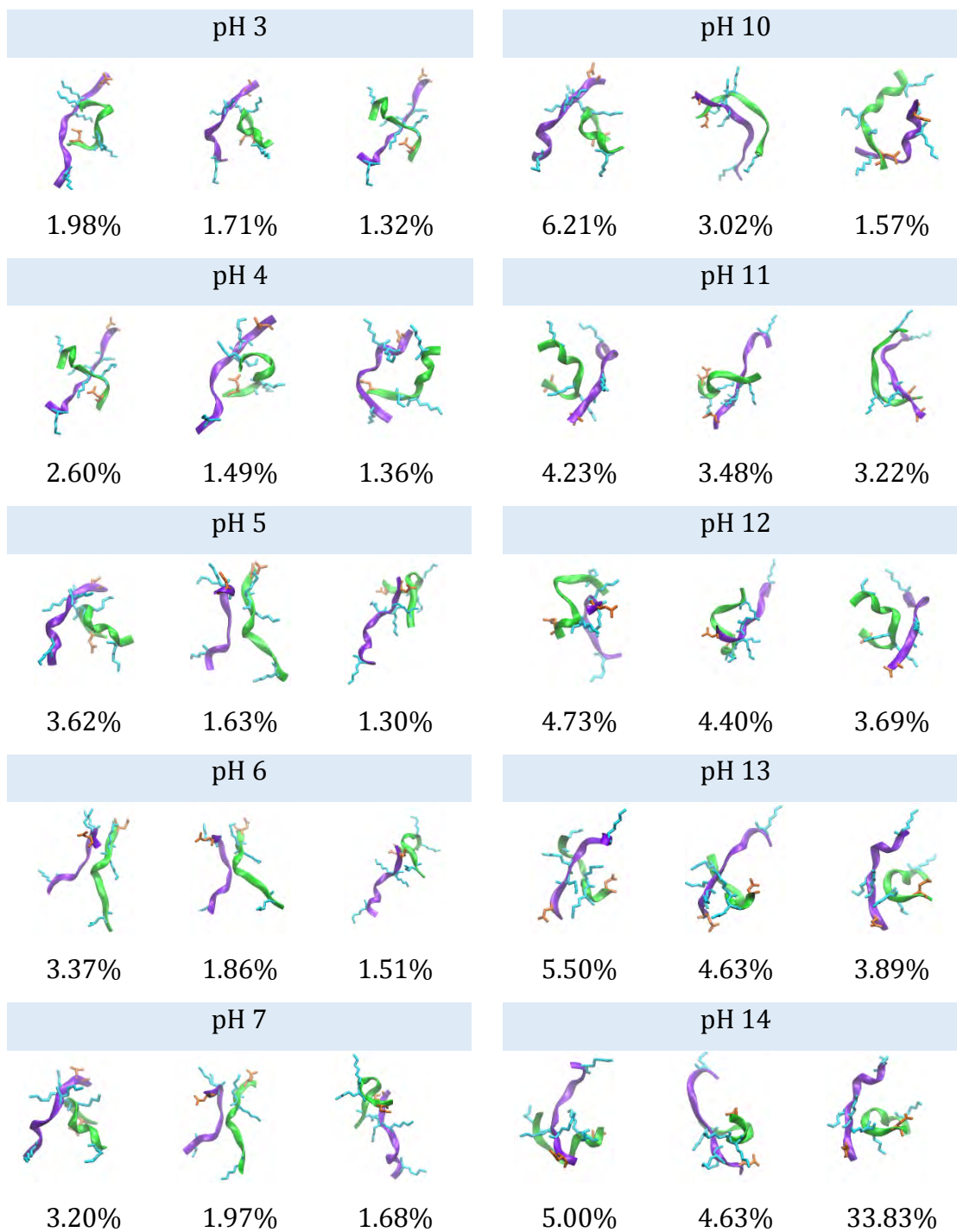


Figura 1. Estructuras más representativas del dímero simulado con agua explícita a cada pH.

APÉNDICE F. Contactos intermoleculares entre las regiones PHF6* de los monómeros A y B de R2/wt

F.1. Dímerización con solvatación GB

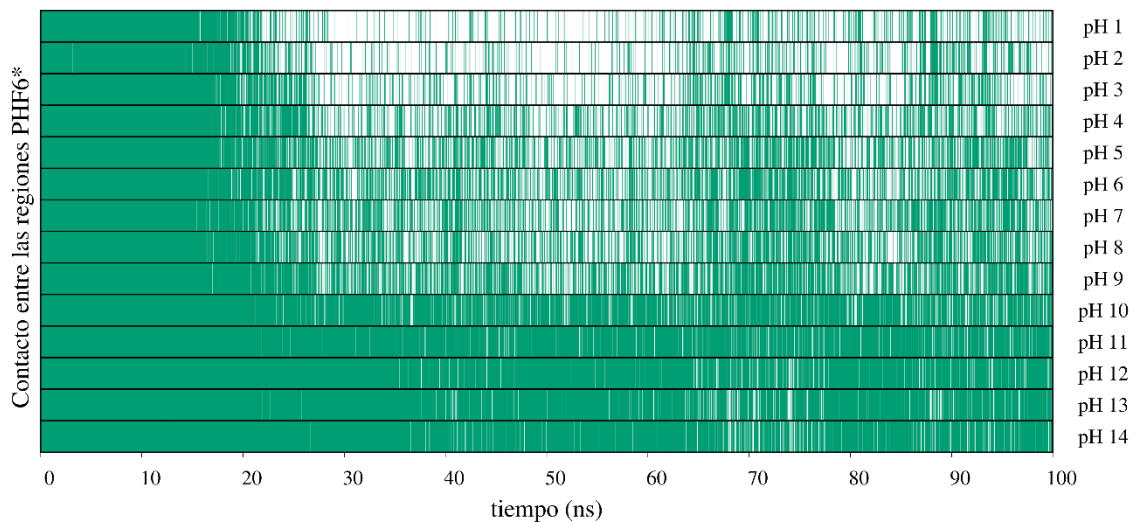


Figura F1. Contacto entre las regiones PHF6* de los monómeros A y B a cada pH. La dimerización está favorecida por el pH.

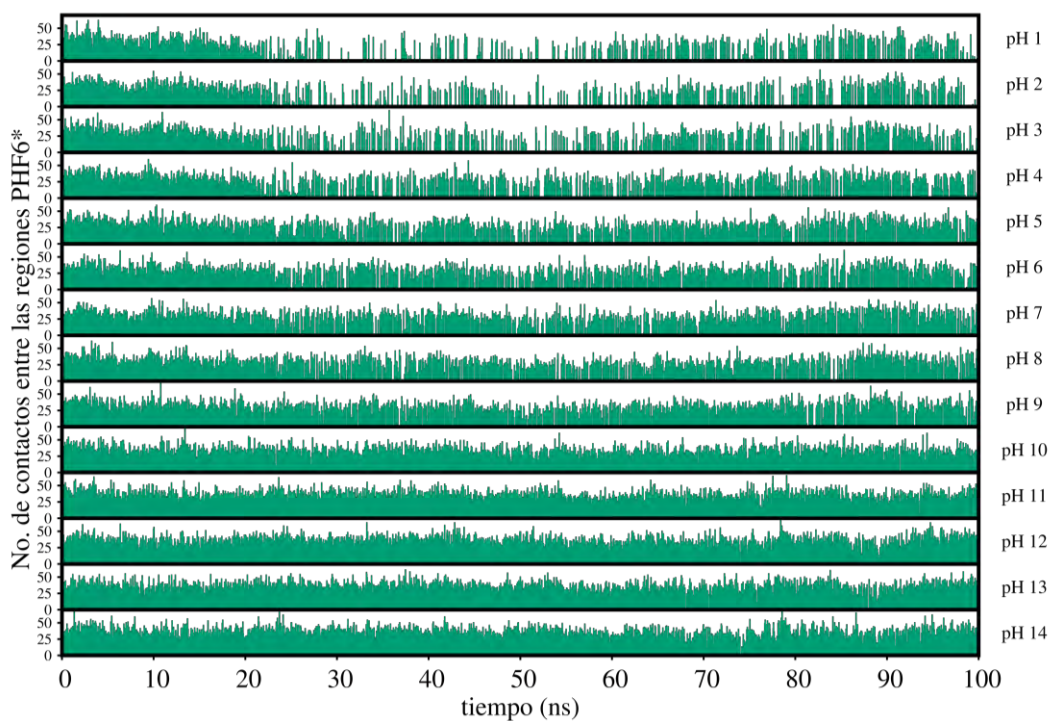


Figura F2. Número de contactos entre las regiones PHF6* de los monómeros A y B a cada pH.

F.2. Dímerización con solvatación explícita

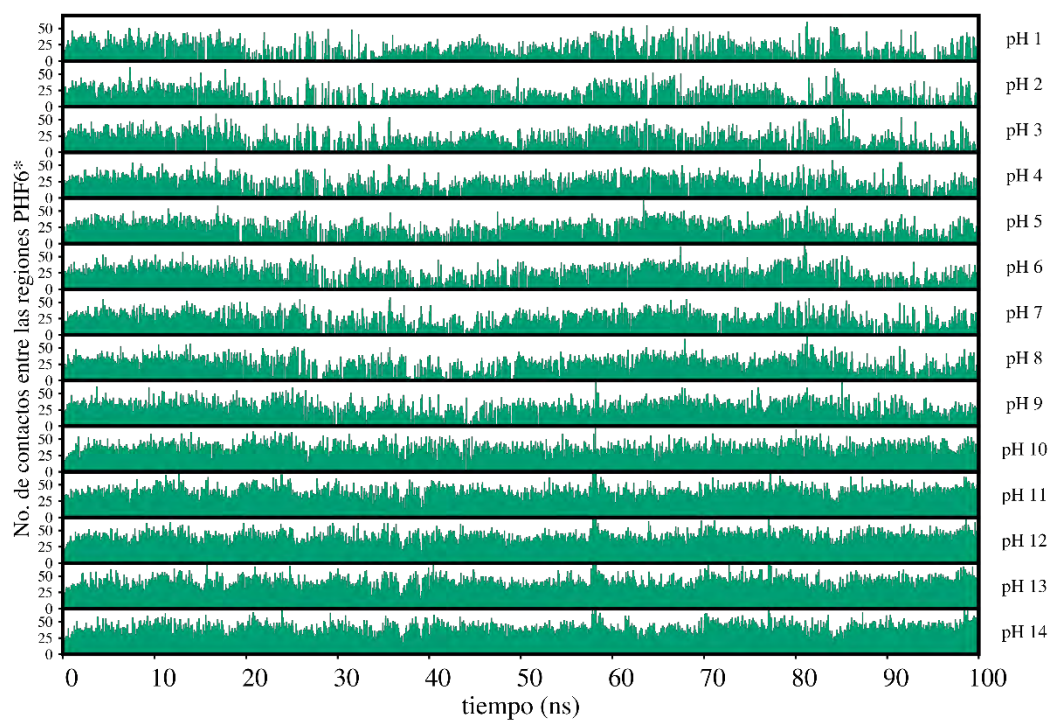


Figura F3. Número de contactos entre las regiones PHF6* de los monómeros A y B a cada pH.

Referencias

1. Verlet, L. Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **159**, 98–103 (1967).
2. Allen, M. Introduction to molecular dynamics simulation. *Comput. Soft Matter From Synth. Polym. to Proteins* **23**, 1–28 (2004).
3. González, M. A. Force fields and molecular dynamics simulations. *Collect. SFN* **12**, 169–200 (2011).
4. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).
5. Landau, D. P. & Binder, K. *A guide to Monte-Carlo simulations in statistical physics.* (2009).
6. Metropolis, N. & Ulam, S. The Monte Carlo Method. *J. Am. Stat. Assoc.* **44**, 335–341 (1949).
7. Cheung, D. L., Anton, L., Allen, M. P. & Masters, A. J. Computer simulation of liquids and liquid crystals. *Computer Physics Communications* **179**, 61–65 (2008).
8. Leach, A. R. *Molecular modelling: principles and applications.* (Prentice Hall, 2001).
9. Walczak, A. M. & Antosiewicz, J. M. Langevin dynamics of proteins at constant pH. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **66**, 051911/1-051911/8 (2002).
10. Bürgi, R., Kollman, P. A. & Van Gunsteren, W. F. Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation. *Proteins Struct. Funct. Genet.* **47**, 469–480 (2002).
11. Mongan, J., Case, D. A. & McCammon, J. A. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **25**, 2038–2048 (2004).

12. Baptista, A. M., Teixeira, V. H. & Soares, C. M. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* **117**, 4184–4200 (2002).
13. Hong, L. Structure of the Protease Domain of Memapsin 2 (beta -Secretase) Complexed with Inhibitor. *Science (80-.).* **290**, 150–153 (2000).
14. Park, H. & Lee, S. Determination of the Active Site Protonation State of β -Secretase from Molecular Dynamics Simulation and Docking Experiment: Implications for Structure-Based Inhibitor Design. *J. Am. Chem. Soc.* **125**, 16416–16422 (2003).
15. Rajamani, R. & Reynolds, C. H. Modeling the protonation states of the catalytic aspartates in β -secretase. *J. Med. Chem.* **47**, 5159–5166 (2004).
16. Polgár, T. & Keserü, G. M. Virtual screening for γ -secretase (BACE1) inhibitors reveals the importance of protonation states at Asp32 and Asp228. *J. Med. Chem.* **48**, 3749–3755 (2005).
17. Yu, N. *et al.* Assigning the protonation states of the key aspartates in β -secretase using QM/MM X-ray structure refinement. *J. Chem. Theory Comput.* **2**, 1057–1069 (2006).
18. Domínguez, J. L. *et al.* Effect of the protonation state of the titratable residues on the inhibitor affinity to BACE-1. *Biochemistry* **49**, 7255–7263 (2010).
19. Sussman, F. *et al.* On a possible neutral charge state for the catalytic dyad in β -secretase when bound to hydroxyethylene transition state analogue inhibitors. *J. Med. Chem.* **54**, 3081–3085 (2011).
20. Barman, A., Schürer, S. & Prabhakar, R. Computational modeling of substrate specificity and catalysis of the β -secretase (BACE1) enzyme. *Biochemistry* **50**, 4337–4349 (2011).
21. Andreeva, N. S. & Rumsh, L. D. Analysis of crystal structures of aspartic proteinases: On the role of amino acid residues adjacent to the catalytic site of pepsin-like enzymes. *Protein Sci.* **10**, 2439–2450 (2001).
22. Barman, A. & Prabhakar, R. Protonation states of the catalytic dyad of β -secretase

- (BACE1) in the presence of chemically diverse inhibitors: A molecular docking study. *J. Chem. Inf. Model.* **52**, 1275–1287 (2012).
23. Mishra, S. & Caflisch, A. Dynamics in the active site of β -secretase: A network analysis of atomistic simulations. *Biochemistry* **50**, 9328–9339 (2011).
 24. Gueto-Tettay, C., Pestana-Nobles, R. & Drosos-Ramirez, J. C. Determination of the protonation state for the catalytic dyad in γ -secretase when bound to hydroxyethylamine transition state analogue inhibitors: A molecular dynamics simulation study. *J. Mol. Graph. Model.* **66**, 155–167 (2016).
 25. Kufareva, I. & Abagyan, R. Methods of protein structure comparison. *Methods Mol. Biol.* **857**, 231–57 (2012).
 26. Zvelindovsky, A. V. *Nanostructured soft matter: experiments, theory, simulation and perspectives*. (Springer, 2007).
 27. D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. G., A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C., Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. S. & R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. X. and P. A. K. AMBER 2016. (2016).
 28. Lobanov, M. Y., Bogatyreva, N. S. & Galzitskaya, O. V. Radius of gyration as an indicator of protein structure compactness. *Molecular Biology* **42**, 623–628 (2008).
 29. Ackermann, M. R., Blömer, J., Kuntze, D. & Sohler, C. Analysis of agglomerative clustering. in *Algorithmica* **69**, 184–215 (2014).
 30. Daura, X. *et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed* **38**, 38: 236-240 (1999).