



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

## Maestría y Doctorado en Ciencias Bioquímicas

Cambios en el muestreo conformacional de la proteína de unión a lisina, arginina y Ornitina de *Salmonella typhimurium* (LAO-BP) asociados con mutaciones en el sitio 117

TESIS

QUE PARA OPTAR POR EL GRADO DE:  
Maestro en Ciencias

PRESENTA:

Diego Sebastián Granados Villanueva

TUTOR PRINCIPAL

Alejandro Sosa Peinado  
Facultad de Medicina, UNAM

MIEMBROS DEL COMITÉ TUTOR

Carmen Nina Pastor Colón  
Centro de Dinámica Celular, UAEM

Rogelio Rodríguez Sotres  
Facultad de Química, UNAM

Ciudad Universitaria, CDMX. Junio, 2019



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.

## Agradecimientos Institucionales

Este trabajo se realizó en el **Laboratorio de Físicoquímica e Ingeniería de Proteínas** , que se encuentra en el departamento de Bioquímica de la **Facultad de Medicina en la UNAM**. Estuve bajo la supervisión del Doctor Alejandro Sosa-Peinado. Durante la duración de mi proyecto, recibí los siguientes apoyos:

- Beca de Maestría del CONACyT (**CVU 660317**)
- Programa de Apoyo a los Estudios del Posgrado para la participación en los congresos de la **Protein Society 2016** (Baltimore, EEUU) y **2017** (Montreal, CA)

Adicionalmente, conté con recursos provenientes de los siguientes proyectos de investigación:

- PAPIIT: **IN222719**
- Proyecto de Supercómputo DGTIC : **114 2019-1**

## Agradecimientos personales

Como partidario de la brevedad, limitaré la enumeración de esta sección. Esto no quiere decir que me falten ganas de agradecer, sino que no quiero robarle el tiempo a los que lean este documento para buscar información que les resulte de utilidad.

En lo académico, quiero agradecer a mi tutor **Alejandro Sosa Peinado**, que me dio la libertad de explorar a mi antojo la forma en que LAO se mueve. En segundo lugar, a los miembros del comité tutor, la **Dra. Nina Pastor** y al **Doctor Rogelio Rodriguez** por sus comentarios en cada una de las evaluaciones que hicieron que este trabajo creciera en direcciones inesperadas. Finalmente, **a mis compañeros de grupo y del Laboratorio de Físicoquímica e Ingeniería de Proteínas**, particularmente a **Jesús Banda, Saira Maldonado y Jorge Marcos**.

En lo personal, quiero agradecer a los seres con los que convivo con mayor frecuencia: **Estefanía, mis papás y Lupe**.

# Índice

1	Introducción	4
1.1	Marco teórico	4
1.1.1	Las proteínas son polímeros que se mueven para funcionar	4
1.1.2	Las dinámicas moleculares nos permiten estudiar la dinámica de las proteínas	5
1.1.3	Las DMs involucran resolver numéricamente ecuaciones diferenciales y estimar la energía	6
1.1.4	Las proteínas periplásmicas de unión (PBPs) son un modelo de estudio para la dinámica	8
1.2	Antecedentes y motivación del trabajo	9
2	Hipótesis	13
3	Objetivos	14
3.1	Objetivo general	14
3.2	Objetivos específicos	14
4	Materiales y métodos	15
4.1	Construcción de los sistemas y descripción del esquema de simulación	15
4.1.1	Selección de mutantes adicionales: L117R y L117Q	15
4.1.2	Coordenadas iniciales para las simulaciones	17
4.1.3	Armado de cajas de simulación	17
4.1.4	Descripción del protocolo de simulación	17
4.2	Análisis de las simulaciones	18
4.2.1	Evaluación del impacto de las mutaciones en la estabilidad de la estructura en LAO	19
4.2.2	Identificación de residuos con movilidad diferencial a nivel de cadena principal	19
4.2.3	Huella dinámica de una PBP: plano apertura-torsión	20
4.2.4	Descripción inicial de las estructuras obtenidas de la dinámica	21
4.2.5	Agrupamiento para determinar conformaciones frecuentes en las simulaciones	22

4.2.6 Obtención de mapas de contactos.....	23
4.2.7 Caracterización de los grupos estructurales.....	23
4.2.8 Residuos de mayor conectividad y pares de interacción significativa.....	24
4.2.9 Redes de contactos de residuos para estudiar cambios globales en la conectividad.....	25
5 Resultados y discusión .....	27
5.1 Las mutaciones no modifican el patrón de puentes de hidrógeno de cadena principal.....	27
5.2 Las mutaciones modifican poco las distribuciones de estructura secundaria.....	28
5.3 Las huellas dinámicas nos dicen que las mutantes se mueven distinto.....	31
5.4 Los residuos que se mueven diferente en L117K se ubican en las bisagras.....	34
5.5 Agrupar por apertura y torsión revela las conformaciones significativas.....	37
5.6 Las mutaciones en el residuo 117 brindan nuevas posibilidades de interacción.....	44
5.7 Simular con una lisina neutral reduce el cierre espontáneo de L117K.....	49
5.8 Los efectos de cambiar el residuo 117 se propagan a zonas distantes de la estructura.....	52
5.9 La topología de la red de interacción del grupo 1 de L117K difiere de la referencia.....	59
6 Conclusiones .....	73
6.1 Evaluación del impacto en la estructura de las distintas mutaciones.....	73
6.2 Impacto de las mutaciones puntuales en la dinámica.....	74
6.2 Localización de residuos con movilidad diferencial.....	74
6.3 Agrupamiento y conectividades de conformaciones relevantes.....	74
6.5 Impactos en la conectividad interna.....	75
7 Perspectivas .....	77

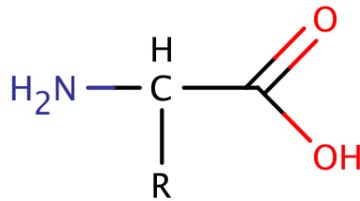
# 1 Introducción

En esta sección proporcionamos la información necesaria a nuestros lectores para que comprendan el contexto de nuestro trabajo. En la primera parte explicamos conceptos breves sobre las proteínas en general y sobre las proteínas periplásmicas de unión a las cuales pertenece LAO, nuestro modelo de estudio. Después describimos las investigaciones previas de nuestro grupo que fungen como los antecedentes de nuestro trabajo para entender de dónde provienen las preguntas que respondimos con nuestros resultados

## 1.1 Marco teórico

### 1.1.1 *Las proteínas son polímeros que se mueven para funcionar*

Las proteínas son polímeros de una colección de 20 moléculas quirales llamadas L-aminoácidos, que tienen una estructura general que mostramos en la figura 1, donde un carbono central (carbono alfa) tiene unidos un grupo amino, un grupo carboxilo y una cadena lateral que suele representarse con la letra R, que determina la identidad del aminoácido en cuestión. Cada proteína tiene una combinación única de estos bloques a la que le llamamos secuencia. La secuencia de una proteína le otorga una conformación característica (estructura nativa) y determina de forma directa su función. Para comprender las fuerzas que mantienen la estructura nativa en orden, solemos diseccionar la conformación de las proteínas a distintos grados de complejidad que llamamos niveles estructurales. El primer nivel estructural (estructura primaria) hace referencia a los vínculos covalentes que forman la cadena hidrocarbonada (*backbone* en inglés), es decir, su secuencia. La estructura secundaria hace referencia a cómo se arreglan de forma espontánea cúmulos de aminoácidos en cercanía directa, adoptando geometrías extendidas (hojas beta) o helicoidales (hélices alfa), vueltas o zonas sin estructura regular como las asas, dependiendo de los patrones de puentes de hidrógeno intermoleculares que establecen, o de la ausencia de éstos. La estructura terciaria se relaciona con cómo se disponen en el espacio los elementos de estructura secundaria y por ende, nos dice la configuración tridimensional de una proteína. Para las proteínas que se componen de más de una cadena polipeptídica, existe un nivel estructural adicional llamado estructura cuaternaria que describe la forma en que se disponen las diferentes subunidades de una proteína multimérica en el espacio [1].



**Figura 1.** Esquema de un L-aminoácido. En el centro se encuentra el carbono alfa, a la izquierda el grupo amino, a la derecha el grupo carboxilo y abajo el grupo R, que es distinta para cada aminoácido.

Como polímeros que son, las proteínas son entes dinámicos que se reconfiguran durante su existencia. Cuando recién son sintetizadas o mientras este proceso ocurre, las proteínas experimentan una reconfiguración inicial conocida como plegamiento, en la cual con o sin ayuda de otras moléculas, adoptan su conformación nativa funcional que corresponde con un mínimo energético. Aunque para 2019 seguimos sin tener claras todas las reglas que hacen que una proteína encuentre su conformación nativa, el trabajo del grupo Anfinsen a finales de la década de 1950 [2] nos mostró que toda la información necesaria para que eso ocurra está codificada en la secuencia. Una vez que se plegaron, las proteínas no se quedan como ladrillos en la conformación nativa sino que coexisten en una colección de conformaciones (ensamble) prácticamente idénticas en términos energéticos, llamadas subestados conformacionales, que coexisten en el equilibrio y entre los cuales las proteínas oscilan para llevar a cabo sus funciones [3]. La cantidad de subestados que componen el paisaje funcional, la frecuencia de las transiciones entre éstos y cómo cambian entre proteínas, constituyen su personalidad dinámica. La dinámica de las proteínas también está codificada en la secuencia y describirla nos permite entender la forma en que desempeñan sus tareas en las células [4].

### *1.1.2 Las dinámicas moleculares nos permiten estudiar la dinámica de las proteínas.*

Las transiciones conformacionales asociadas al comportamiento dinámico de las proteínas se dan a distintas escalas de tiempo de forma concurrente. Estas distintas escalas de tiempo presentan distintos retos de descripción y existen diversas técnicas experimentales para describirlas, cuyos detalles han sido abordados de forma minuciosa en revisiones pertinentes [5]. Las técnicas para caracterizar la dinámica tienen limitantes en su resolución temporal y espacial. Esto lo podemos solventar utilizando metodologías de simulación, particularmente, dinámicas moleculares (DMs).

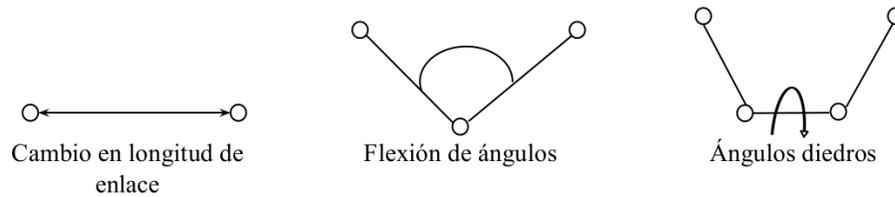
La DM es una técnica de simulación de la física molecular en donde se obtiene una trayectoria de la evolución temporal de las coordenadas asociadas a un sistema de múltiples partículas mediante la resolución numérica de las ecuaciones de movimiento de Newton. Esta técnica se formuló en 1959 por Alder y Wainwright [6] usándose por primera vez para proteínas en 1977, cuando McCammon y colaboradores [7] simularon por 2 picosegundos el inhibidor de tripsina pancreática bovina (BPTI). Desde ese momento las capacidades de cómputo se han desarrollado de forma veloz de manera que para el año 2009 se generó una trayectoria de simulación de 1.031 milisegundos de la misma proteína, que equivale a 550 veces la longitud de aquella simulación pionera [8]. Este incremento en el cómputo, sumado al refinamiento de la capacidad descriptiva de los modelos asociado a este tipo de simulaciones ha devenido en que actualmente las DMs se hayan convertido en un caballo de batalla para los bioquímicos, quienes las han usado para describir problemas biológicos diversos como el plegamiento, los cambios conformacionales y el reconocimiento de proteínas por sus ligandos y pares de interacción [9].

### *1.1.3 La DMs involucran resolver numéricamente ecuaciones diferenciales y estimar la energía.*

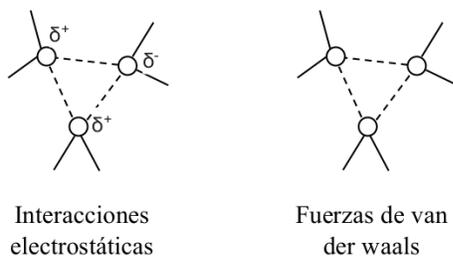
Para ejecutar las DMs es necesario resolver la segunda derivada de la posición de las partículas del sistema respecto al tiempo mediante algoritmos de integración en un tiempo discreto, es decir, seccionado en pequeños fragmentos. En cada uno de estos lapsos se estiman las contribuciones energéticas del sistema usando funciones empíricas llamadas campos de fuerza, que diseccionan las diferentes contribuciones energéticas del sistema (figura 2) para calcularlas utilizando formas funcionales que se aproximan de forma aceptable. Vistos de una forma burda, los campos de fuerza se estructuran como en la ecuación 1, sumando la energía asociada a las interacciones que se dan entre los átomos que se encuentran enlazados ( $E_{\text{bonded}}$ ) y las relacionadas con las interacciones que se establecen entre átomos no enlazados ( $E_{\text{non bonded}}$ ). En la figura 3 mostramos ejemplos de formas funcionales usadas para estimar las distintas contribuciones energéticas presentes en un sistema sometido a dinámicas moleculares. El acortamiento de enlaces y los ángulos de flexión entre residuos se representan usando potenciales armónicos. Los ángulos diedros se representan utilizando expansiones de senos y cosenos. Las interacciones electrostáticas se modelan usando la ley de Coulomb y las interacciones electrostáticas se obtienen implementando el potencial de Lennard-Jones 12-6 [10].

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{nonbonded}} \quad (1)$$

**Interacciones entre átomos enlazados ( $E_{bonded}$ )**



**Interacciones entre átomos no enlazados ( $E_{non\ bonded}$ )**



**Figura 2.** Disección de las interacciones de un campo de fuerza

**Interacciones entre átomos enlazados ( $E_{bonded}$ )**

$$\sum_{\text{enlaces}} \frac{k_i}{2} (l_i - l_{i,0})^2 \quad \sum_{\text{ángulos}} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 \quad \sum_{\text{torsiones}} \frac{V_n}{2} (1 + \cos(\eta\omega - \gamma))$$

Cambio en longitud de enlace      Flexión de ángulos      Ángulos diedros

**Interacciones entre átomos no enlazados ( $E_{non\ bonded}$ )**

$$\sum_{i=1}^N \sum_{j=1}^N 4\epsilon_{ij} \left( \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right)$$

Interacciones electrostáticas      Fuerzas de van der waals

**Figura 3.** Ejemplos de formas funcionales usadas en los campos de fuerza

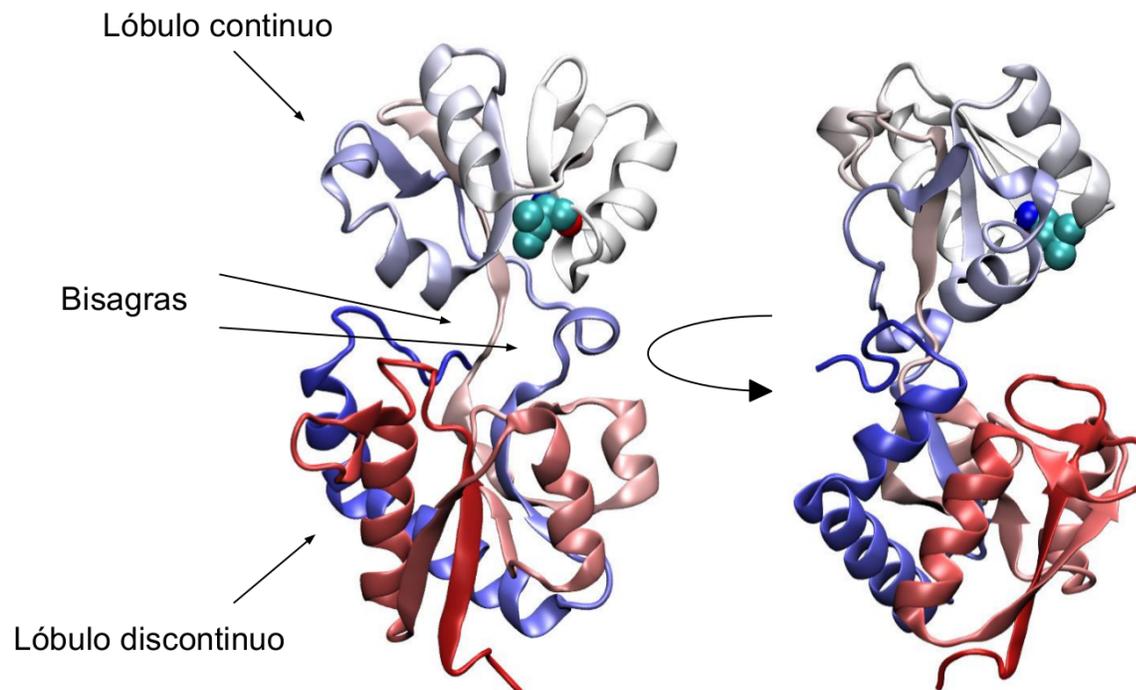
#### 1.1.4 Las proteínas periplásmicas de unión (PBPs) son un modelo de estudio para la dinámica

En este trabajo, estudiamos la dinámica de diversas mutantes de la proteína de unión a Lisina, Arginina y Ornitina (LAO-BP). LAO-BP es una proteína que pertenece a la superfamilia de las proteínas periplásmicas de unión (PBP). Estas proteínas funcionan como elementos de reconocimiento en los sistemas de dos componentes bacterianos participando en procesos celulares como la quimiotaxis y la ingesta de nutrientes. Sin importar la diversidad de sus blancos, todas las PBPs comparten un plegamiento que consiste en dos dominios globulares donde una hoja beta está rodeada de hélices alfa (plegamiento de Rossmann), los cuales están unidos por una zona intermedia de plegamiento variable denominada bisagra. Dentro de las PBPs existen dos subclases que difieren en cómo distribuyen la secuencia dentro del polipéptido: las PBP de la clase I presentan seis hebras en su hoja beta (ordenadas en un patrón 213456) mientras que las de clase II tienen 5 (ordenadas en un patrón 21354) [11].

Además de compartir plegamiento, las PBPs tienen un mecanismo dinámico común para interactuar con sus ligandos que involucra el acercamiento de sus lóbulos a manera de cierre, lo que genera una interfaz de recepción para el ligando compuesta por residuos que solían estar expuestos en la superficie en la conformación abierta. Este mecanismo, al que se le ha encontrado analogía con la forma en la cual las plantas carnívoras devoran a sus presas (*Venus Flytrap Mechanism*) ha permitido que estas proteínas además de ser blancos interesantes para la ingeniería de proteínas, también sean modelos para entender los vínculos que se establecen entre la dinámica y la función.

Retomando la proteína de este trabajo, LAO-BP es una PBP de la clase I compuesta por 238 residuos que reconoce L-aminoácidos con cadena lateral cargada positivamente (figura 4). Une con alta afinidad a Lisina, Arginina y Ornitina (15, 14 y 30 nM), con una afinidad que es dos órdenes de magnitud menor a Histidina (500 nM). Las secciones que le componen se distribuyen en un lóbulo discontinuo compuesto por los residuos 1 a 88 y 195 a 238, uno continuo que abarca los residuos 93 a 185 y dos segmentos polipeptídicos que funcionan como bisagras (residuos 89 a 92 y 186 a 194) [12]. Por la manera en la cual lo visualizamos en este trabajo, el lóbulo continuo también lo nombraremos como lóbulo superior y el discontinuo como lóbulo inferior. Tomando como referencia las estructuras cristalográficas correspondientes, la transición de conformación abierta a cerrada involucra una rotación de cuerpo rígido de 38.2° [13]. Por redes cinéticas construidas usando datos de dinámica molecular, sabemos que el mecanismo de interacción de LAO con sus ligandos involucra diversas rutas paralelas de exploración conformacional que convergen en un subestado conocido como el complejo de encuentro, cuyas estructuras asociadas están parcialmente cerradas e interactúan de forma

débil con su ligando. Al alcanzar el complejo de encuentro, ocurre una transición rápida hacia la conformación cerrada de manera que el mecanismo dominante de la unión es el ajuste inducido. En la red cinética no se observaron transiciones hacia el estado cerrado en ausencia de los ligandos, lo que significa que es necesario que estos estén presentes para que LAO pueda cerrarse [14].

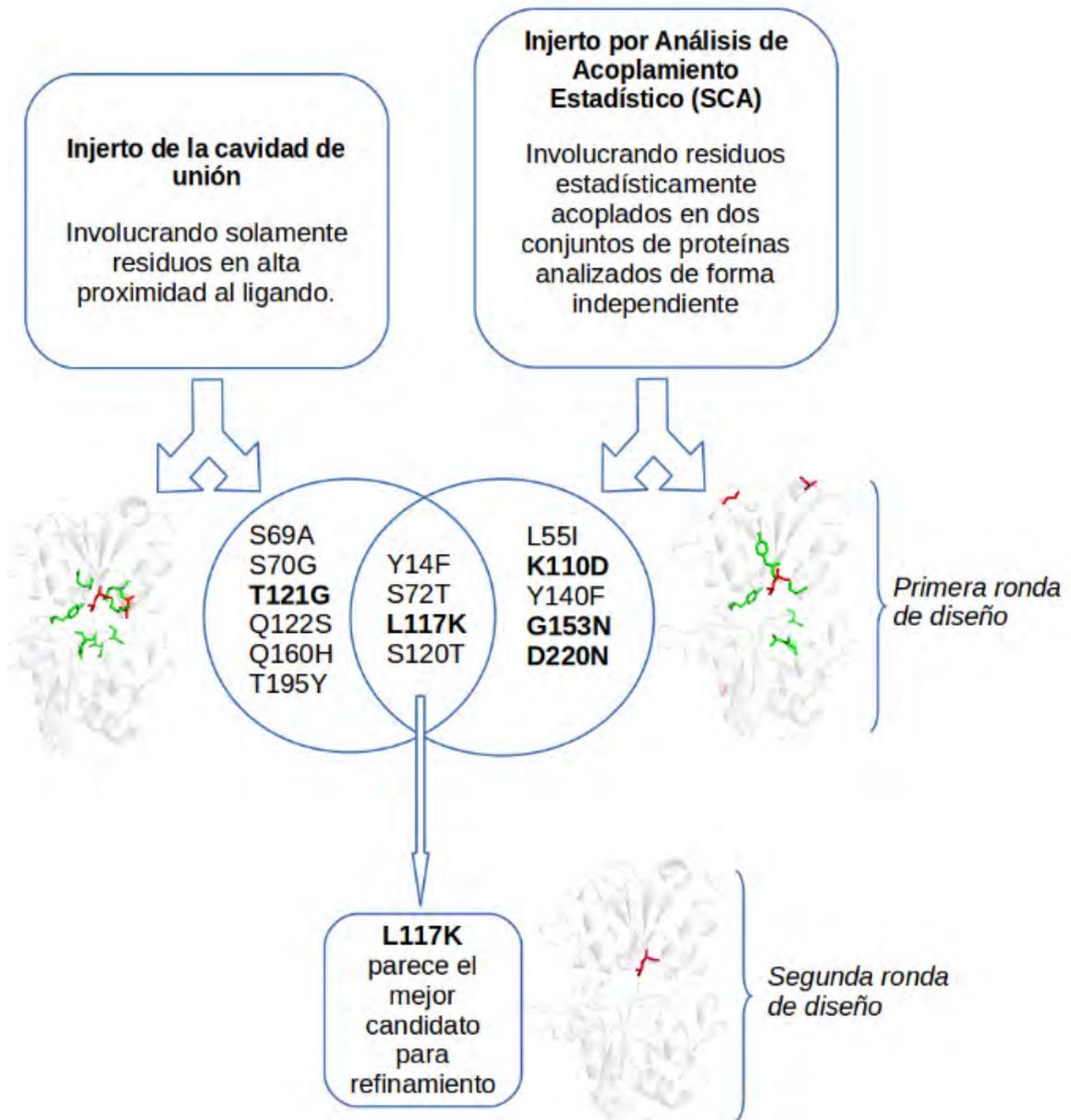


**Figura 4.** Distintas regiones de la estructura tridimensional de LAO-BP en su conformación abierta. En el panel de la derecha vemos una rotación de la misma estructura. Ambas visualizaciones se encuentran coloreadas por el índice de la secuencia, en color rojo el extremo amino y en azul el extremo carboxilo. En esferas, se resalta la localización del residuo L117.

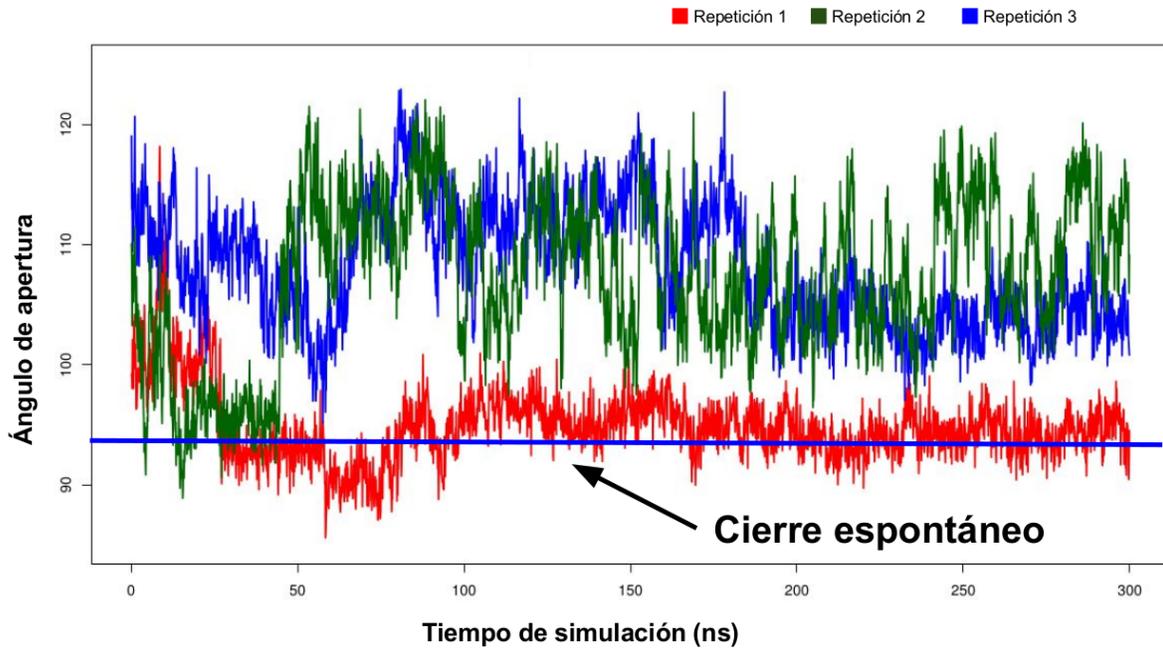
## 1.2 Antecedentes y motivación del trabajo

En el laboratorio de Físicoquímica e Ingeniería de proteínas (LFQIP) del Departamento de Bioquímica de la Facultad de Medicina de la UNAM, los grupos de Alejandro Sosa y Alejandro Fernández hemos estudiado diversas PBP's para entender su plegamiento y su función con la intención de desarrollar variantes que puedan reconocer ligandos distintos a los que unen naturalmente. Este trabajo se enmarca de forma directa en el trabajo de doctorado de Jesús Banda, que cambió la funcionalidad de LAO-BP para que pudiera reconocer glutamina, un residuo que tiene una cadena lateral neutral y que la versión silvestre de esta proteína es incapaz de reconocer. Banda diseñó una colección de mutantes usando dos enfoques: en el primero de ellos, sustituyó los residuos asociados al sitio de unión de LAO-BP por los que están en el sitio de reconocimiento a ligando de la proteína de unión a glutamina de *Escherichia coli*

(QBP); dicha estrategia se conoce como *grafting* o injerto del sitio de unión. En el segundo enfoque generó un universo de mutaciones susceptibles a cambiar la especificidad de LAO mediante la comparación de los resultados del análisis de acoplamiento estadístico (SCA) de alineamientos múltiples de secuencias de conjuntos de proteínas generados con el plegamiento de LAO-BP y de QBP. Este enfoque permitió que existiera una lista de cambios posibles para generar la nueva afinidad por glutamina de LAO-BP. Al estudiar las mutaciones propuestas por ambos enfoques, se encontró un conjunto de 4 mutaciones compartidas (figura 5) de las cuales solo una representa una mutación no conservativa, la de la introducción de una lisina en la posición 117 que en la secuencia silvestre tiene una leucina (L117K) [15]. Al inicio de mi trabajo de maestría, dentro de las proteínas que simulé preliminarmente estuvo L117K. De las tres simulaciones iniciales que llevé a cabo, partiendo del estado abierto y en ausencia de ligandos, en una de ellas la proteína visitaba valores atípicos en el ángulo de apertura (figura 6), propiedad que calculamos al obtener el ángulo que se presenta entre los centros de masa del lóbulo inferior, bisagra y lóbulos superior. Sabemos por trabajo previo del laboratorio que mencioné anteriormente, que la visita de valores de apertura reducida en ausencia del ligando son inexistentes en la versión silvestre de LAO, por lo que consideramos relevante comprender si este evento es frecuente y si es exclusivo de esta mutación. Indagar en la forma que una mutación que altera la afinidad por ligandos modifica la dinámica es de vital importancia para indagar en los orígenes de los cambios de funcionamiento que se dan en la mutante L117K. Ampliando un poco los alcances, trabajos como el que llevamos a cabo tienen posibilidades de enriquecer los algoritmos de diseño racional mostrándonos cómo es que modificando su personalidad dinámica, las proteínas adquieren nuevas funciones, ya que hay evidencia de que el rediseño exitoso de proteínas involucra un proceso inicial de expansión de la diversidad conformacional para posteriormente seleccionar con mutaciones adicionales aquellas conformaciones que son más propensas a desempeñar de forma exitosa la nueva función [16].



**Figura 5.** Estrategia de diseño usada para la generación de L117K, mutación que se decidió ubicando la única mutación no conservativa compartida por ambas estrategias de diseño. Figura tomada de [15].



**Figura 6.** Simulaciones de control partiendo de la conformación abierta de L117K sin ligandos donde se observa cierre espontáneo.

## 2 Hipótesis

La introducción de la mutación L117K en LAO-BP modifica las conexiones intermoleculares que establece esta proteína en su versión silvestre, lo que incrementa la diversidad conformacional que visita la conformación abierta en ausencia de ligandos, dándole oportunidad de desarrollar nuevas afinidades.

# 3 Objetivos

## 3.1 Objetivo general

Determinar, usando simulaciones por Dinámica Molecular, la manera en que cambia la diversidad conformacional del estado abierto de LAO al introducir la mutación L117K.

## 3.2 Objetivos específicos

- Efectuar simulaciones por Dinámica Molecular de la versión silvestre de LAO y de L117K para entender los cambios en la exploración conformacional que se presentaron respecto a la silvestre.
- Efectuar simulaciones de mutantes adicionales, L117Q y L117R, para entender si las diferencias observadas respecto a la silvestre son exclusivas de la mutación L117K o se originan por simplemente cambiar el residuo silvestre en ese sitio bajo ciertos criterios.
- Identificar las conformaciones más importantes en la exploración conformacional del estado abierto de las proteínas simuladas usando algoritmos de agrupamiento.
- Generar mapas de contactos para identificar las interacciones intermoleculares más importantes en las proteínas simuladas y compararlas con las que establece la versión silvestre.
- Generar redes de contactos de residuos para estudiar cómo se modifica la globalidad de las conexiones intermoleculares con las mutaciones que introdujimos e identificar los residuos críticos para que ocurran los cambios observados.

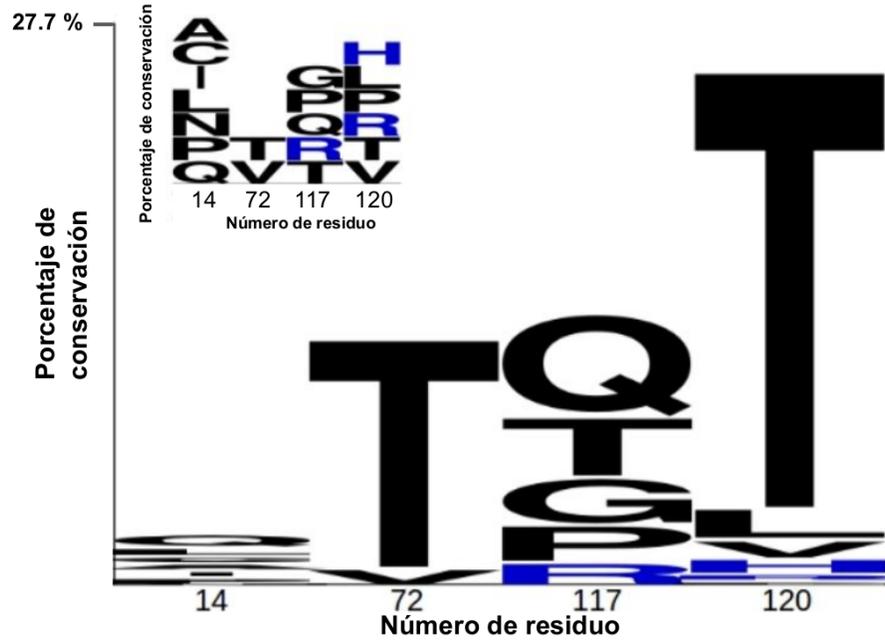
# 4 Materiales y métodos

Por fines de claridad, dividimos la sección en dos apartados. En la primera abordamos todo lo relacionado con el establecimiento de los sistemas de simulación y describimos el esquema que seguimos para simular. En la segunda sección hacemos una descripción minuciosa de las rutinas de análisis de las simulaciones. Para fluidez del texto, describimos los modelos matemáticos y el código asociados a las rutinas de manera sencilla, y enfatizando en los conceptos, dejando para los lectores más curiosos las referencias que permiten acceder a los detalles que no cubrimos aquí

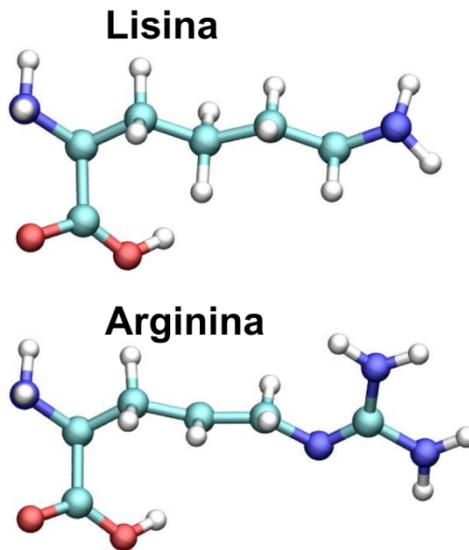
## **4.1 Construcción de los sistemas y descripción del esquema de simulación**

### *4.1.1 Selección de mutantes adicionales: L117R y L117Q*

Exploramos el espacio de secuencias para la posición 117 con mutantes adicionales para tener un punto de comparación respecto a L117K y entender si las nuevas propiedades dinámicas que encontramos en ella son producto exclusivo de esta mutación, o si con solo sustituir por otro aminoácido con propiedades fisicoquímicas equivalentes a la lisina silvestre observamos efectos similares. Las dos mutantes adicionales que reportamos en este trabajo son L117R y L117Q. La selección de L117R se justifica por la identidad de carga y distribución atómica de su cadena lateral, ya que consiste en un grupo R extendido con carga positiva, que es parecido al de la lisina (figura 8). En el caso de L117Q, la respuesta es la conservación por sitio. Construimos colecciones de secuencias de proteínas con la arquitectura PBP, cuyos ligandos son arginina y glutamina y están presentes en la base de datos UNIREF, filtrada con un valor de redundancia del noventa por ciento (UniRef90). Cuando hicimos alineamientos múltiples para las secuencias que cumplieron con las características previamente descritas (1387 para las que unen arginina y 839 para las que tiene como ligando glutamina) encontramos que el residuo más conservado para la posición 117 en ambos casos es glutamina (figura 7). Por lo tanto, en L117Q exploramos la viabilidad de la elección consenso al cambiar el residuo nativo.



**Figura 7.** Porcentaje de conservación para diversos sitios en proteínas de la arquitectura PBP que unen arginina (figura principal) y glutamina (inserción en la esquina superior izquierda). En ambos universos de proteínas, glutamina es el residuo más conservado para la posición 117.



**Figura 8.** Comparación de lisina y arginina. Ambos aminoácidos poseen grupos R extendidos con carga positiva, que se pueden observar en el lado izquierdo de ambas estructuras.

#### 4.1.2 Coordenadas iniciales para las simulaciones

Obtuvimos las coordenadas iniciales de las mutantes (L117K, L117R y L117Q) en estado abierto con el servidor de modelado tridimensional I-TASSER [17]. Para ello, proporcionamos la secuencia correspondiente y establecimos que el servidor utilizara como molde las coordenadas cristalográficas de LAO, en estado abierto y sin ligandos (PDB ID: 2LAO). En todas las mutantes, el proceso de modelado arrojó un solo cluster estructural con un C-Score superior a 1.5, lo que habla de la fiabilidad en los modelos propuestos. En el caso de las simulaciones de LAO en estado abierto sin ligando y L117K cerrada con glutamina, usamos las coordenadas de las estructuras cristalográficas correspondientes a las cuales les removimos moléculas de agua y co-factores de cristalización.

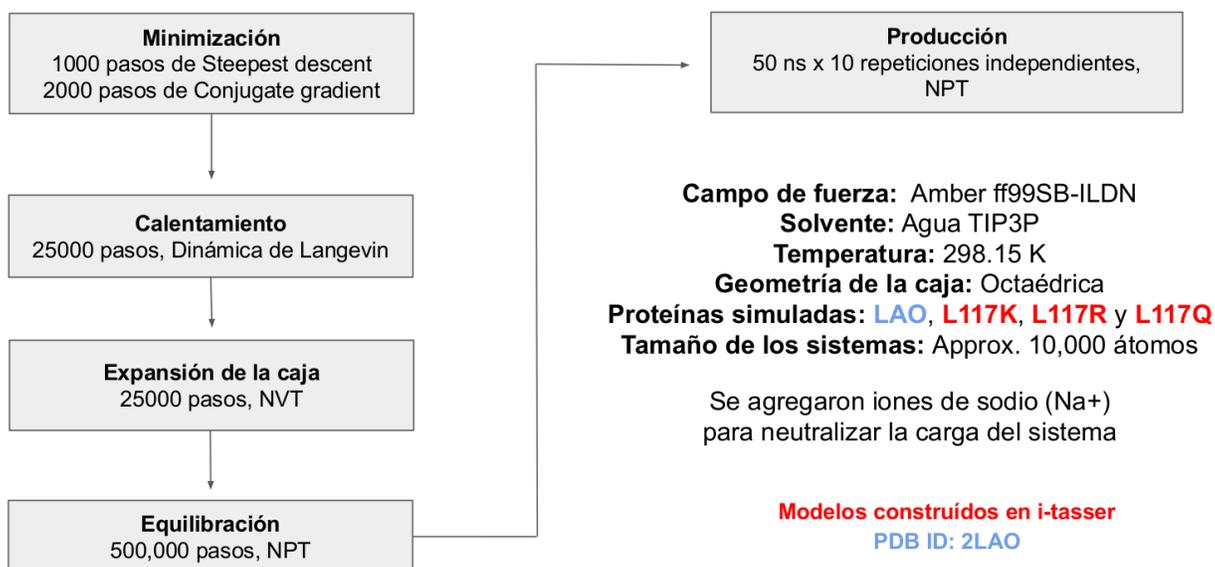
#### 4.1.3 Armado de cajas de simulación

Antes de ser simuladas, colocamos las proteínas en cajas de simulación utilizando la suite de modelado molecular Amber [18]. Como primer paso, corregimos problemas de numeración en los archivos PDB correspondientes utilizando la aplicación *pdb4amber*. Posteriormente, utilizamos *tLeap* para neutralizar, solvatar y generar los archivos de topología necesarios para llevar a cabo las simulaciones. Neutralizamos la carga de los sistemas utilizando iones de Sodio con la función *addions*. Solvatamos los sistemas en cajas octaédricas de agua TIP3P [20] con una longitud de arista de al menos 12 Å tomando como referencia cualquier átomo perteneciente a las proteínas, asegurando que la longitud mínima se respetara usando la función *iso*. El campo de fuerza que utilizamos fue ff99SB-ILDN [20] debido a su buena reproducción de propiedades experimentales de proteínas [21].

#### 4.1.4 Descripción del protocolo de simulación

Llevamos a cabo las simulaciones usando el motor *pmemd.cuda* de Amber14. Un esquema del protocolo de simulación que seguimos está en la figura 9. Como primer paso, corregimos problemas estéricos presentes en las configuraciones iniciales de las proteínas simuladas mediante un proceso de minimización donde 1000 pasos del algoritmo *steepest descent* seguidos de 2000 pasos de *conjugate gradient*. Ajustamos la densidad de las cajas de simulación con 25,000 pasos de Dinámica NVT, seguidos de 25,000 pasos de una dinámica de Langevin donde elevamos progresivamente la temperatura de la simulación hasta 298.15 K. En estas dos etapas restringimos el movimiento de los átomos correspondientes a la parte proteica de los sistemas. Eliminamos las restricciones de movimiento y equilibramos el sistema durante 500,000 pasos en una dinámica NVT. Para los pasos de producción usamos el ensamble NPT.

Para cada sistema, se generaron 10 simulaciones de producción independientes con una longitud individual de 50 ns. Para fijar las posiciones de átomos de hidrógeno utilizamos el algoritmo *settle* [22] que nos permitió extender el paso de integración a 2 fs en las etapas de producción y equilibración. El control de temperatura se llevó a cabo utilizando el termostato de Langevin con una frecuencia de colisión ( $\gamma_{ln}$ ) de  $2.0 \text{ ps}^{-1}$ . Para el control de presión se utilizó barostato de Berendsen. Consideramos una distancia de 8 Å para el cálculo explícito de las interacciones electrostáticas y las fuerzas de Van der Waals. El resto de las interacciones no enlazadas fueron calculadas de forma implícita utilizando la implementación de *Particle Mesh Ewald* (PME).



**Figura 9.** Esquema del protocolo de simulación.

#### 4.2 Análisis de las simulaciones

Obtuvimos los datos que mencionamos a continuación, a menos que digamos algo distinto, utilizando rutinas implementadas en *cpptraj* de Ambertools14 [23] y *mdanalysis* de Python [24]. El procesamiento adicional se llevó a cabo con *scripts* construidos en los lenguajes de programación bash, Python y R. Las representaciones tridimensionales de las estructuras de proteínas las llevamos a cabo en los visualizadores PyMOL [25] y VMD [26]. Los diagramas de Venn los generamos usando Venny 2.1 de BioinfoGP del CNB [27].

#### 4.2.1 Evaluación del impacto de las mutaciones en la estabilidad de la estructura en LAO

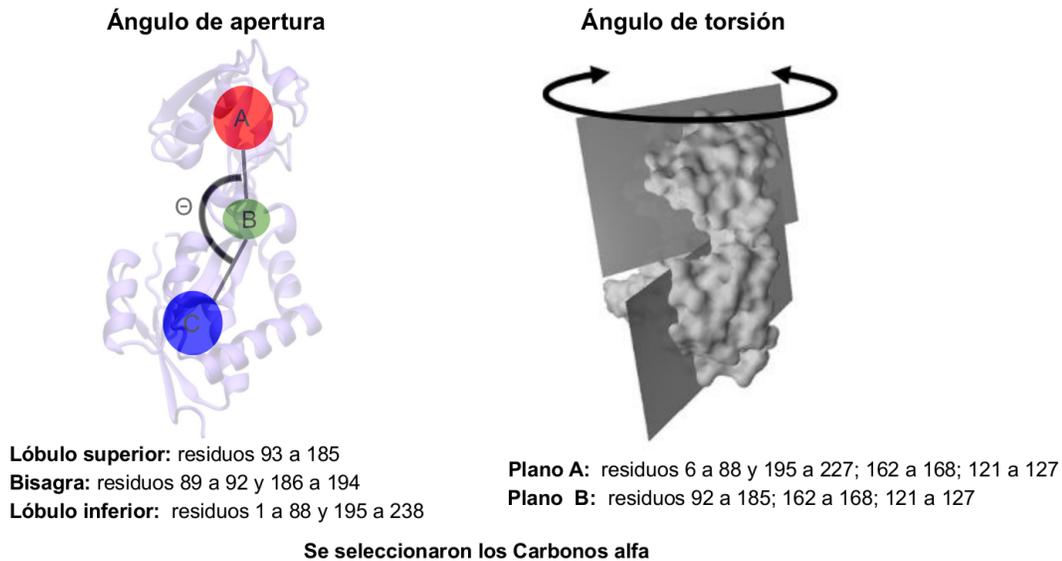
Calculamos las distribuciones de número de puentes de hidrógeno de cadena principal y el porcentaje de residuos en conformación alfa o beta para evaluar si las mutaciones que introdujimos tuvieron impactos en la topología de LAO. Los puentes de hidrógeno los calculamos usando la función *hbond* de *cpptraj* seleccionando exclusivamente donadores y aceptores asociados a la cadena principal. Una vez que obtuvimos el número de puentes de hidrógeno graficamos los histogramas de esta propiedad para cada una de las mutantes en contra del de LAO, que funge como referencia silvestre. Obtuvimos el porcentaje de intersección entre el histograma de la mutante y de la referencia usando el paquete *Histogram Tools* [28] de R. Para las gráficas de porcentaje de estructura secundaria, generamos los datos usando el programa *stride* [29]. Tomamos esos datos y extrajimos el porcentaje de residuos en conformación alfa o beta usando rutinas en R, y estimamos sus respectivas distribuciones usando la función *density*. Posteriormente obtuvimos la intersección de las distribuciones de cada una de las mutantes y de la referencia silvestre usando la librería *overlapping* de R [30].

#### 4.2.2 Identificación de residuos con movilidad diferencial a nivel de cadena principal

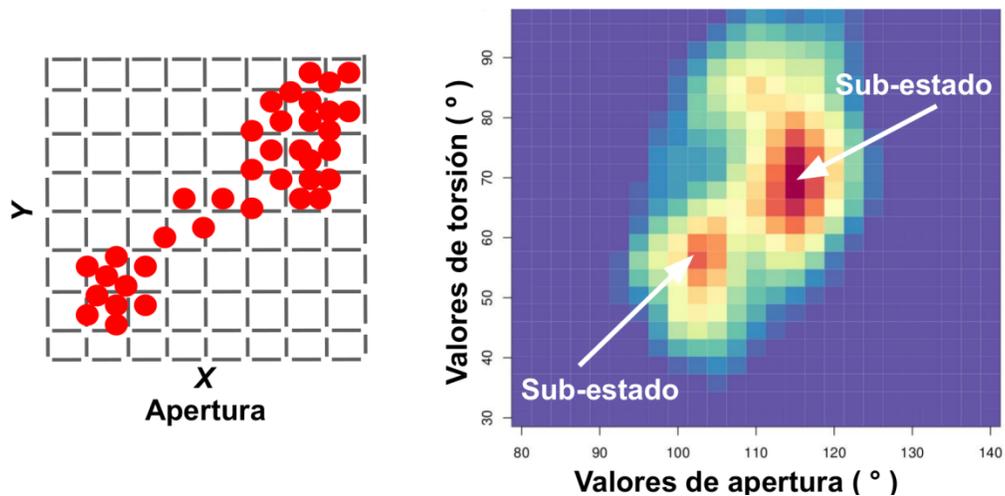
Para identificar los residuos que presentan movilidad diferencial a nivel de cadena principal comparamos los gráficos de Ramachandran de los residuos de las mutantes respecto a los de la referencia silvestre. Para esto, obtuvimos los valores de los ángulos Phi y Psi usando las rutina *dihedral* de *cpptraj* para cada uno de los residuos que componen las proteínas que simulamos. Posteriormente, comparamos los residuos de las mutantes respecto a su contraparte silvestre usando la divergencia de Kullback-Leibler (DKL) [31] para identificar aquellos cuyas distribuciones difirieron en mayor grado y por lo tanto presentan diferencias en la movilidad. La DKL es una medida que nos dice la cantidad de información de una variable aleatoria que puede ser explicada por una distribución aproximada. Este valor oscila de 0 a 1, donde el valor es creciente en sentido contrario a la cantidad de información descrita por la aproximación. Para llevar a cabo la comparación, estimamos distribuciones bidimensionales constituidas por los ángulos  $\phi$  y  $\psi$  usando la función KDE de R y calculamos la DKL usando la función *dkl* de la librería *FNN* de R [32]. Tomamos como variable aleatoria de referencia la distribución obtenida para la versión silvestre de LAO, la cual se comparó con las distribuciones obtenidas para las mutantes. Seleccionamos los residuos que presentaron una dispersión más grande para cada una de las mutantes y visualizamos la manera en que éstos se distribuyeron en la estructura utilizando las funciones de selección de residuos en VMD.

#### 4.2.3 Huella dinámica de una PBP: plano apertura-torsión

Describimos las conformaciones que LAO visitó durante las simulaciones usando los ángulos de apertura y torsión. Estas propiedades en conjunto representan el 60% de la varianza conformacional en la dinámica de esta PBP, por lo que el plano formado por los ángulos de apertura y torsión constituye una huella dinámica para las proteínas de esta arquitectura. El ángulo de apertura se definió como el existente entre los centros de masa de las tres secciones que componen a LAO (lóbulo inferior, bisagra y lóbulo superior) mientras que el de torsión es el ángulo dihedral que existe entre los dos planos que están definidos en la figura 10. Para representar la forma en que se distribuyeron las conformaciones en el plano apertura-torsión durante las simulaciones construimos histogramas bidimensionales (*heatmaps*), como el que se muestra en el panel derecho de la figura 11. En ellos, dividimos una malla de graficación en intervalos regulares y asociamos los valores de apertura con la coordenada X y los de torsión con la coordenada Y. Cada *frame* de la simulación se representa como un punto en esta malla y la densidad de puntos en las distintas zonas de la superficie de graficación se representa con un código de colores, que en nuestro caso oscila de violeta en las zonas con baja densidad de puntos a rojo, que es el color que tienen las zonas de mayor densidad. Las zonas de alta densidad corresponden con conformaciones altamente visitadas durante las simulaciones cuyas estructuras pueden ser identificadas mediante algoritmos de agrupamiento (*clustering*) para caracterizar sus propiedades estructurales. Concatenamos las repeticiones de las simulaciones para cada proteína simulada y extrajimos los valores de apertura y torsión usando la rutinas *angle* y *dihedral* de *cpptraj*, respectivamente. Para generar el heatmap, utilizamos la función *KDE2D* de la librería *MASS* de R [33]. Para que la escala de colores fuera igual en todas las gráficas que generamos, definimos una escala de colores asociada a la densidad compartida, usando como referencia la que se genera para LAO, que presenta los puntos de mayor densidad de todas las proteínas que simulamos.



**Figura 10.** Definiciones del ángulo de apertura y el ángulo de torsión. Las definiciones fueron tomadas de [34].



**Figura 11.** Ejemplo de un heatmap para representar la exploración conformacional de LAO. En el esquema del panel izquierdo los puntos rojos representan conformaciones individuales exploradas durante la simulación. En la gráfica de la derecha, la densidad de puntos en la malla se enuncia mediante el código de colores. En este ejemplo se observan dos sub-estados, que son señalados en el panel derecho y corresponden a las zonas con una concentración alta de puntos presentes en el esquema del lado izquierdo.

#### 4.2.4 Descripción inicial de las estructuras obtenidas de la dinámica

Para comenzar a describir las estructuras que componen las trayectorias de las proteínas que simulamos, obtuvimos diagramas de caja y bigote de las distribuciones de apertura y torsión. Los diagramas de caja y bigote son representaciones que nos muestran la manera en la cual se distribuyen los datos de una variable a través de la representación de sus cuartiles. En ellos,

se dibuja una caja que tiene como límites inferior y superior el primer ( $Q_1$ ) y tercer cuartil ( $Q_3$ ), respectivamente. Dentro de esta caja hay una línea a la mitad que representa el valor de la mediana, que corresponde al segundo cuartil ( $Q_2$ ). Las dos líneas que salen de la caja delimitan el mínimo y el máximo de los valores de la distribución. La línea hacia abajo es el límite inferior que se calcula restando el valor de  $Q_1$  menos 1.5 veces el valor de la distancia entre  $Q_3$  y  $Q_1$  que se conoce como el rango intercuartil (IQR). El límite superior se obtiene sumando 1.5 veces el valor de IQR al valor de  $Q_3$ . Graficamos los diagramas de caja y bigote y calculamos el IQR usando la función *boxplot* en R.

#### 4.2.5 Agrupamiento para determinar conformaciones frecuentes en las simulaciones

Encontramos las conformaciones más frecuentes durante las simulaciones usando el algoritmo de agrupamiento (clustering) HDBSCAN [35], tomando como datos de clasificación los valores conjuntos de apertura y torsión de las trayectorias. Elegimos este algoritmo porque agrupa mediante la densidad de puntos y necesita de la definición de pocos parámetros, haciéndolo un algoritmo mucho más agnóstico que otros frecuentemente usados para el agrupamiento de trayectorias de dinámica molecular. HDBSCAN transforma la distancia de los puntos de una manera en la que se pueden diferenciar de forma robusta los puntos informativos de los que solo aportan ruido, para posteriormente construir una jerarquía, la cual es posteriormente condensada para poder extraer de ella clusters estables.

Para este trabajo utilizamos la implementación del algoritmo en Python [36]. Para efectuar la clasificación concatenamos todas las repeticiones para cada uno de los sistemas, generando una gran trayectoria de la cual tomamos estructuras cada 10 frames, lo que nos dio un total de 5000. Definimos como único parámetro el tamaño mínimo de cluster, que establece el umbral mínimo de estructuras congregadas necesarias para formar un grupo y lo establecimos en 50, de manera que para formar un cluster es necesario que al menos el 1% de estructuras estén en él. La medida de distancia que usamos fue la distancia de Haversine (ecuación 1), por lo cual transformamos previamente los valores de apertura y torsión en radianes. Esta medida se usa en navegación, para encontrar la distancia que existe entre dos puntos en una esfera dadas su longitud y latitud. Dado que los valores de apertura y torsión pueden ser entendidos como coordenadas en un plano, consideramos pertinente usar esta ecuación ya que así nuestra clasificación no solo toma en cuenta la densidad, sino la localización en el espacio dentro del plano apertura-torsión que define el espacio conformacional.

#### 4.2.6 Obtención de mapas de contactos

Los mapas de contacto representan visualmente cómo se relacionan entre sí los átomos que constituyen una macromolécula y en el caso de las proteínas, condensan en una representación simple información tanto de la estructura secundaria como de la topología tridimensional [37]. Constan de una matriz en la que los ejes tienen una longitud idéntica al número de átomos, donde cada entrada representa la relación binaria entre un par de átomos y el valor que tiene asociada refleja la interacción o (ausencia de ésta) entre ellos. En el caso más sencillo, la interacción se define de manera booleana: si el valor de la distancia entre una pareja de átomos cualquiera ( $r_{ij}$ ) es menor a un radio de corte ( $r_{cut}$ ), se considera que existe una relación y la entrada correspondiente en la matriz adquiere el valor de 1. En el caso contrario, se considera que la relación no existe y la entrada que adquiere el valor es 0. Para las proteínas, los mapas de contacto se construyen normalmente usando las distancias entre los carbonos  $\alpha$  y el radio de corte que suele oscilar entre 8 y 10 Å.

Como primer paso, tomamos las estructuras asociadas a cada uno de los grupos estructurales de las mutantes y generamos trayectorias de longitudes distintas para cada una de ellas usando rutinas implementadas en la librería de Python *MDAnalysis* [38]. Habiendo hecho esto, construimos los mapas de contactos usando la suite CONAN [39], que posee rutinas para construir y analizar mapas de contactos. CONAN genera matrices de distancia interatómicas usando la herramienta *mdmat* de GROMACS, a partir de las cuales genera un mapa de contactos promedio junto con una serie de gráficas de propiedades asociadas a éste. Adicionalmente, puede generar películas que reflejan la evolución de los contactos a lo largo de la trayectoria que construye, usando gráficas generadas en GNUplot con *mencoder*. En los mapas que reportamos en este trabajo, seleccionamos todos los átomos de proteína y consideramos que una interacción entre átomos existía si la distancia entre ellos era menor de 10 Å, pero mayor de 5 Å en al menos la mitad de los frames. Este criterio nos permite filtrar las interacciones poco estables de nuestros análisis posteriores.

#### 4.2.7 Caracterización de los grupos estructurales

Extrajimos las conformaciones pertenecientes a los distintos grupos estructurales usando *scripts* contruidos en Python con funciones de *MDAnalysis*. Construimos trayectorias para cada grupo estructural de cada una de las mutantes para usarlos en pasos posteriores. Obtuvimos las distribuciones de apertura y torsión de cada uno de los grupos estructurales y las representamos como gráficas de violín usando la función *violinplot* de la librería *seaborn* de Python. Los diagramas de violín son representaciones parecidas al diagrama de caja y bigotes,

que nos muestran los cuartiles de la distribución con el beneficio adicional de que nos permiten observar la forma de las distribuciones. Dividimos las conformaciones que ubicamos gracias al agrupamiento en tres categorías, basadas en los valores promedio de la distribución de ángulo de aperturas de las estructuras que les componen:

- **Conformaciones abiertas:** clusters que contienen estructuras cuyos valores promedio de apertura fueron superiores a  $110^\circ$
- **Conformaciones parcialmente cerradas:** clusters compuestos por estructuras que tuvieron un valor promedio de distribución de apertura menor que  $110^\circ$ , pero igual o mayor que  $100^\circ$
- **Conformaciones cerradas:** clusters cuyos valores promedio de la distribución de apertura fueron menores que  $100^\circ$

Usamos la función *barplot* de R para representar la forma en que se distribuyen los *frames* de las distintas repeticiones en los grupos estructurales.

#### 4.2.8 Residuos de mayor conectividad y pares de interacción significativa

Con los mapas de contacto promedio obtuvimos los residuos de mayor conectividad para cada uno de los grupos estructurales de las proteínas simuladas. Para esto construimos rutinas de análisis y visualización usando la PyMOL API de Python, en donde generamos listas de los residuos que exhiben un número de interacciones superiores al 90% de los residuos. Para encontrar los residuos de alta conectividad que eran exclusivos de cada cluster de las mutantes, establecimos como referencia de las conectividades internas silvestres el mapa de contacto del cluster mayoritario (número 1) de LAO. Encontramos los residuos de alta conectividad exclusivos para cada uno de los clusters de las mutantes obteniendo el diagrama de Venn de los conjuntos de la referencia y el cluster en cuestión. Para los residuos de alta conectividad que ubicamos como diferentes respecto a la referencia calculamos sus interacciones no enlazadas usando la función *lie* de Ambertools y seleccionamos los residuos que presentaron energías de interacción no enlazada menores o iguales a  $-2.0$  kcal/mol y les llamamos pares de interacción significativa. Construimos visualizaciones de la localización de los residuos de alta conectividad y sus pares de interacción significativa mediante rutinas de la PyMOL API en Python.

#### 4.2.9 Redes de contactos de residuos para estudiar cambios globales en la conectividad

Las propiedades funcionales de las proteínas no son fácilmente inferibles a partir de las de los aminoácidos que las componen, lo que quiere decir que son objetos complejos. Un formalismo que se ha utilizado para estudiar a los objetos complejos son las redes. En ellas, un sistema es representado en forma de un grafo en el que los componentes que lo integran son los nodos y las relaciones entre ellos son las aristas. La representación en forma de redes es poderosa porque una vez construidas permiten utilizar una serie de herramientas matemáticas bien establecidas de la teoría de grafos para obtener información valiosa de la globalidad de los sistemas que las originaron, que pueden ser tan diversos como la estructura del transporte público de una ciudad o la manera en la que interactúan todas las proteínas de *Saccharomyces cerevisiae*. Es decir, podemos olvidarnos de las complejidades asociadas a la naturaleza del sistema estudiado durante un rato para simplemente lidiar con entes matemáticos. En las redes de estructura de proteínas, normalmente se conocen como redes de contactos de residuos (RINs, por sus siglas en inglés), los aminoácidos se constituyen como los nodos y las aristas son las interacciones en ellos, las cuales se expresan como matrices de adyacencia que pueden ser construidas a partir de la información de un mapa de contactos o bien a partir de las matrices de movimientos correlacionados [41].

Para este trabajo construimos las matrices de adyacencia a partir de la información contenida en los mapas de contactos. Para ello, tomamos las interacciones que están presentes en al menos 80% de las estructuras de cada uno de los grupos estructurales. Importamos esta información con la librería *networkx* de Python [41] para construir redes no direccionadas. Exportamos las redes generadas usando las funcionalidades de esta librería y las visualizamos utilizando el programa yEd [42], dibujándolas con un diseño jerárquico. Las propiedades que reportamos de las redes las obtuvimos usando las recetas implementadas en *networkx* cuyas definiciones formales pueden encontrarse en la documentación correspondiente. Como propiedades generales obtuvimos el grado promedio, el coeficiente del clustering promedio, longitud característica, la eficiencia global y local promedio así como la distribución de grado. La gráfica de la distribución de grado la generamos usando las librerías de graficación *seaborn* y *matplotlib*. Para la identificación de los residuos críticos, obtuvimos tres tipos de centralidad: centralidad de grado, cercanía (*closeness centrality*) e intermediación (*betweenness centrality*). La primera ubica los nodos con mayor número de conexiones. Las otras dos centralidades se relacionan con las conexiones que se establecen entre los residuos. Obtuvimos los valores de centralidad usando las funciones pertinentes en *networkx* y representamos sus distribuciones usando gráficas de barra en R, donde también construimos

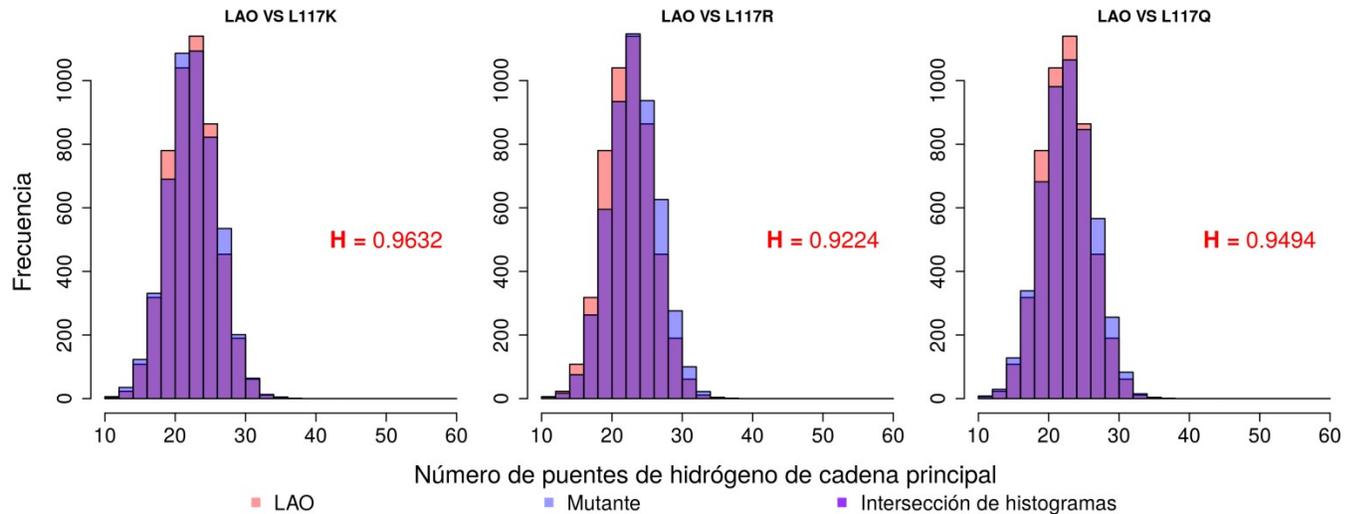
rutinas para identificar residuos críticos (hubs) y residuos con mayor variabilidad en los valores de cada una de las centralidades respecto a la referencia silvestre, es decir, el cluster 1 de LAO. Los residuos que identificamos como críticos son los que tienen un valor mayor que la media en al menos 1.8 desviaciones estándar y los que identificamos como de alta variabilidad fueron aquellos que están arriba o abajo del valor de referencia en al menos 1.8 desviaciones estándar.

# 5 Resultados y discusión

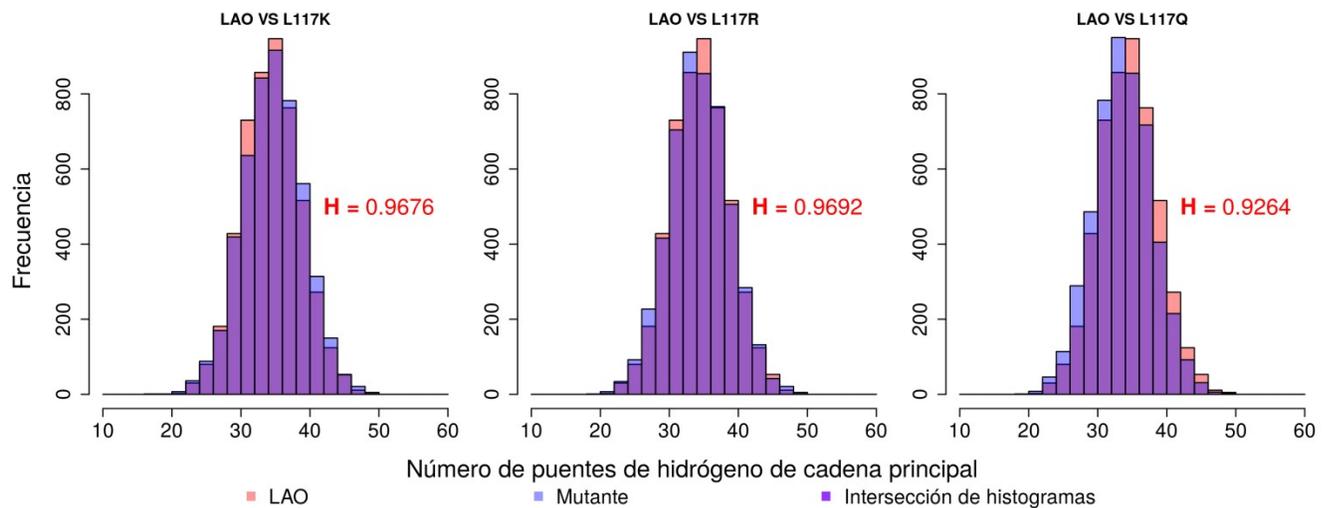
Comenzamos la sección determinando si las mutaciones puntuales que introducimos afectan la topología de LAO. Posteriormente, estudiamos los efectos que tuvieron estas mutaciones puntuales en la dinámica, encontrando que la mutante L117K tiene capacidades de cierre espontáneo atípicas. Ubicamos conformaciones relevantes usando un algoritmo de agrupamiento y después caracterizamos las diferencias de la conectividad de esta mutante respecto a la referencia silvestre usando la información de mapas de contactos y medidas asociadas a la teoría de redes.

## **5.1 Las mutaciones no modifican el patrón de puentes de hidrógeno de cadena principal**

La primera pregunta que contestamos con las trayectorias de dinámica molecular es si las mutaciones en la posición 117 generaron perturbaciones estructurales en LAO. Esto es relevante porque queremos estar seguros de que cualquier diferencia que observamos en la dinámica entre proteínas simuladas se debe a las reconfiguraciones asociadas a la introducción de un residuo no nativo y no a la pérdida de estructura. Aunque para L117K contamos con evidencia experimental que nos asegura que esto no sucede [15], no es el caso para L117Q y L117R. Como primer acercamiento obtuvimos los histogramas de la cantidad de puentes de hidrógeno totales en la cadena principal para cada una de sus mutantes y los comparamos con los de la versión nativa. La comparación está desglosada por lóbulo en las figuras 12 y 13. En color morado mostramos las zonas de intersección entre los histogramas de LAO (color rojo) y la mutante con la que se compara (color azul) mientras que el número de color rojo en el lado derecho de cada panel muestra la fracción de intersección entre ambos histogramas ( $H$ ). Ninguna de las mutantes en ninguno de sus lóbulos exhibe una intersección menor a 0.9 con el histograma de la nativa, nuestro valor de referencia. Esto quiere decir que las distribuciones de esta propiedad son lo suficientemente parecidas para que descartemos reconfiguraciones de los lóbulos a nivel de cadena principal.



**Figura 12.** Comparación de los histogramas de puente de hidrógeno de cadena principal para el lóbulo superior



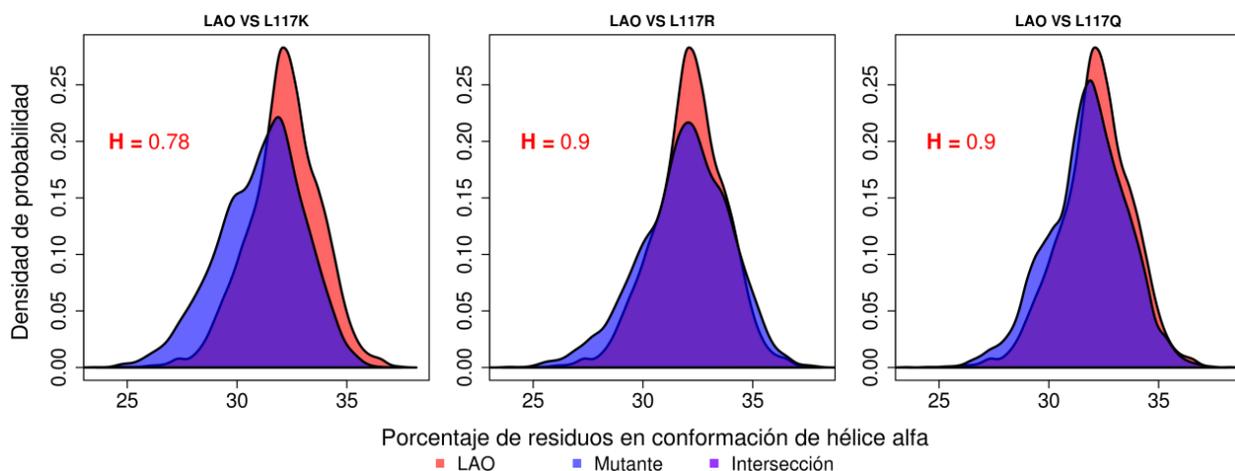
**Figura 13.** Comparación de los histogramas de puente de hidrógeno de cadena principal para el lóbulo inferior

## 5.2 Las mutaciones modifican poco las distribuciones de estructura secundaria

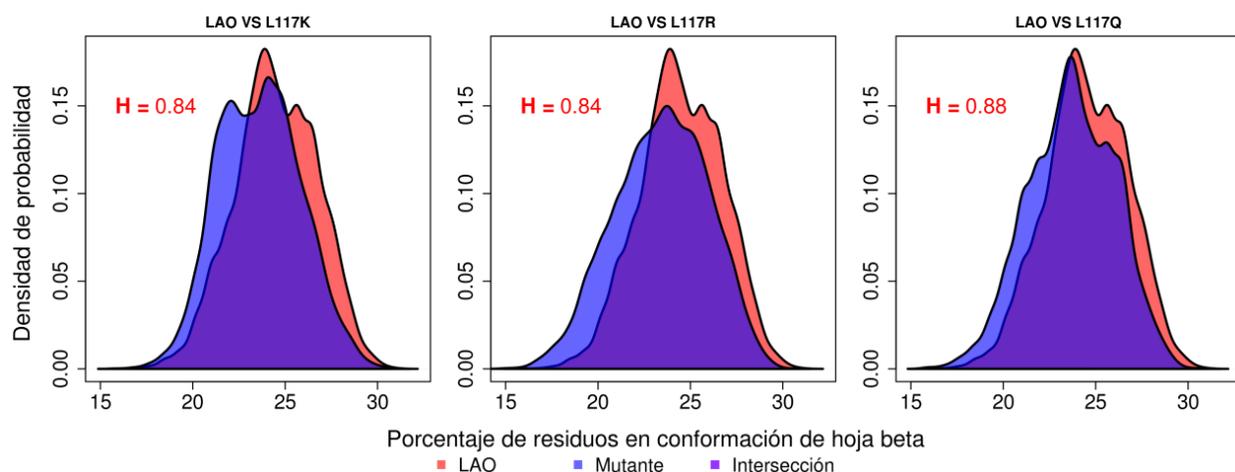
Posteriormente, evaluamos si las mutaciones que introdujimos causaron algún impacto a nivel de estructura secundaria ya que este nivel estructural, además de reacomodos a nivel de cadena principal, involucra reconfiguraciones de las cadenas laterales de los aminoácidos. Para estudiar esto dentro de las simulaciones, calculamos las distribuciones del porcentaje de residuos en conformación alfa o beta para cada una de las mutantes. Si bien este enfoque no es suficientemente fino para detectar cambios sutiles en la estructura secundaria, nos otorga

una idea global de ellos y esto nos permite saber en cuáles casos vale la pena hacer un análisis más detallado. Mostramos los resultados del análisis en las figuras 14 y 15 usando el mismo código de colores que elegimos para la comparación de histogramas de puentes de hidrógeno. Al igual que en la ronda de figuras anterior, el valor que acompaña a cada panel en estas figuras es la fracción de intersección de las distribuciones.

El análisis nos muestra que existen cambios no muy drásticos en las distribuciones de las mutantes a lo largo de las simulaciones si tomamos como referencia a la versión nativa de LAO. Cuando comparamos considerando las proteínas completas, existe un ligero corrimiento hacia la izquierda de las distribuciones en ambas figuras para todas las mutantes, lo que indica que hay una disminución en el porcentaje de residuos que se encuentran tanto en conformación alfa como beta. Los valores de intersección de las distribuciones señalan que L117Q es la mutante que menor disminución presenta. Por otro lado, L117K es la que exhibe diferencias mayores en las distribución, especialmente en el porcentaje de residuos en conformación alfa ( $H=0.79$ ). En el caso del porcentaje de residuos en conformación beta, el valor de intersección se encuentra en el mismo rango que el de las otras mutantes ( $H=0.84$ , lo que es idéntico al valor de L117K y ligeramente menor al de L117Q) pero la forma que tiene la distribución es distinta, lo que sugiere que la cantidad de residuos que visitan la conformación beta difiere de la referencia. Finalmente, L117R exhibe un comportamiento ambivalente, ya que en el porcentaje de residuos en conformación alfa tiene el mismo valor que L117Q y para el caso de los residuos en beta, como se mencionó arriba, presenta el mismo valor que L117K. Una explicación plausible para las diferencias observadas entre mutantes para este rubro es la naturaleza química de las cadenas laterales de los residuos asociados a las mutaciones. En L117Q, a diferencia de los otros dos casos, la cadena lateral

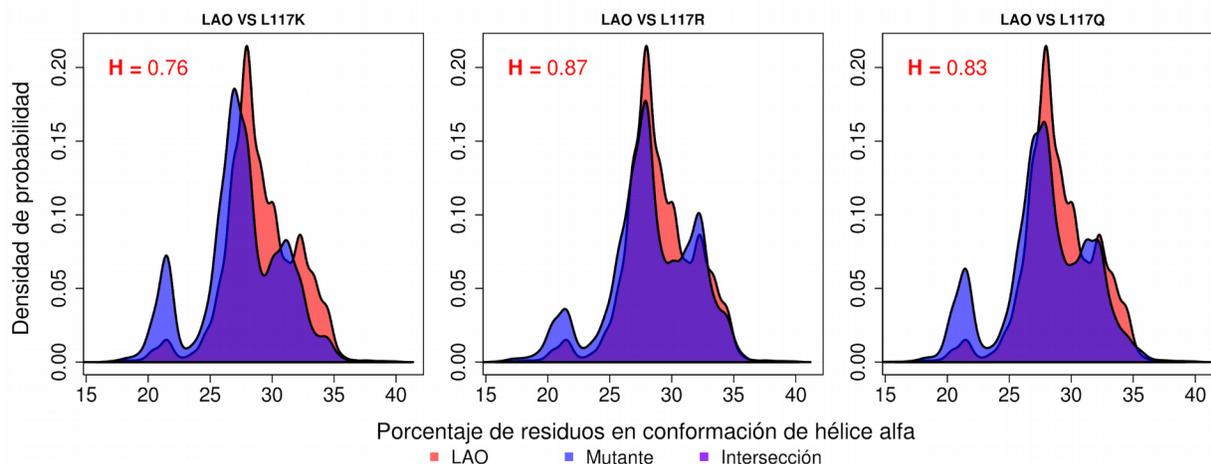


**Figura 14.** Comparación de las distribuciones del porcentaje de residuos en conformación alfa tomando en cuenta toda la estructura

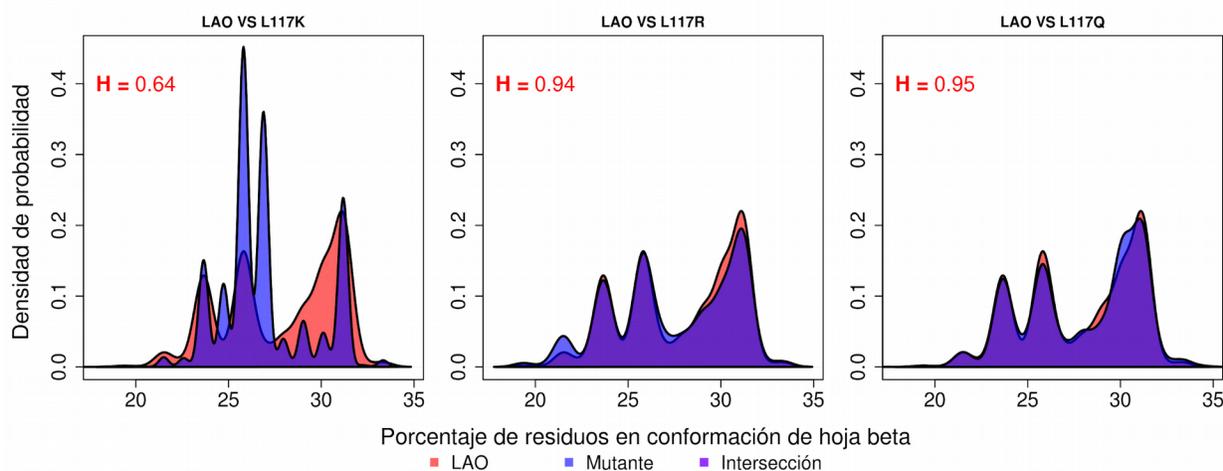


**Figura 15.** Comparación de las distribuciones del porcentaje de residuos en conformación beta tomando en cuenta toda la estructura

del residuo que introducimos no presenta carga real. Recibir un residuo cargado donde no solía haberlo resulta disruptivo para las interacciones que el residuo 117 establece con su periferia. Indagamos la pertinencia de esta hipótesis obteniendo las distribuciones exclusivamente para el lóbulo superior (donde se localiza la mutación) que se ven en las figuras 16 y 17. L117K es la mutante que presenta una disminución de mayor proporción en el porcentaje de residuos en cualquiera de las conformaciones. Este efecto es mucho más drástico para los residuos en hoja beta donde la forma de la distribución es completamente distinta (el valor H es de 0.64), lo que indica que la introducción de una lisina en la posición 117 de LAO tiene un impacto considerable en la manera en la que se forman hojas beta en ese lóbulo. No obstante, esto no se traduce en una gran pérdida de residuos estructurados, como puede verse en la figura 18, donde comparamos la distribución del porcentaje de residuos sin asignación de esta mutante con la versión nativa de LAO. El valor de intersección ( $H=0.85$ ) es alto, lo que nos lleva a pensar que si bien introducir una lisina en la posición 117 no pasa desapercibido por LAO, las perturbaciones estructurales asociadas no son de consideración.



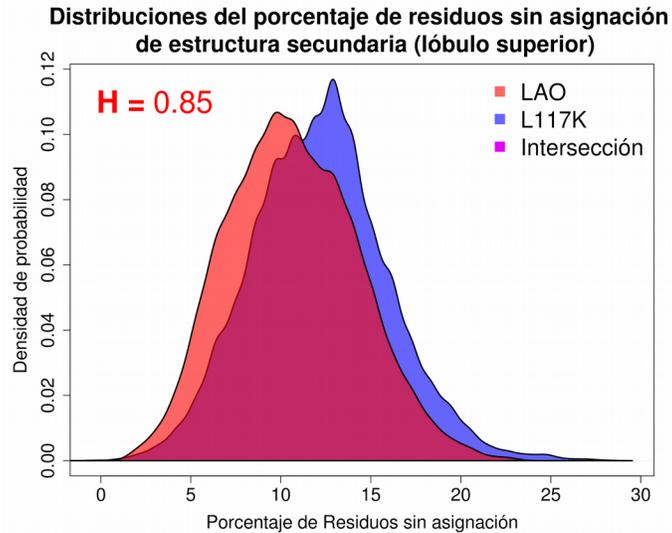
**Figura 16.** Comparación de las distribuciones del porcentaje de residuos en conformación alfa para el lóbulo superior



**Figura 17.** Comparación de las distribuciones del porcentaje de residuos en conformación beta para el lóbulo superior.

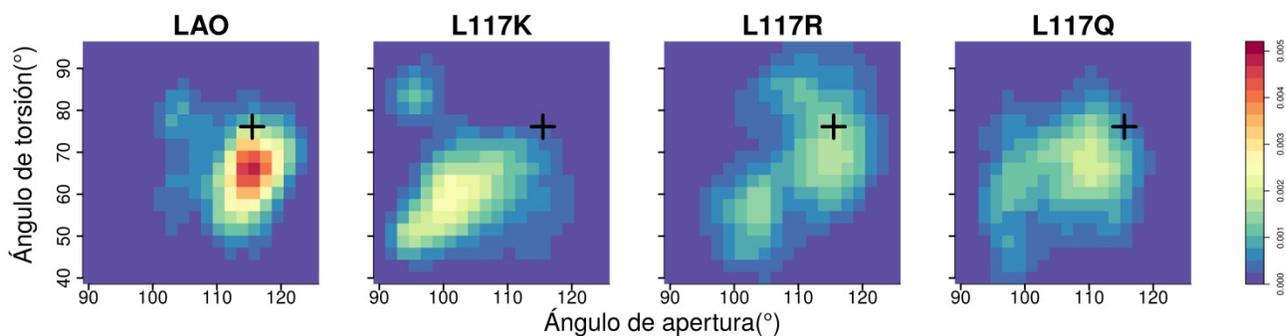
### 5.3 Las huellas dinámicas nos dicen que las mutantes se mueven distinto

Una vez que confirmamos que las mutaciones no indujeron pérdida estructural, nos concentramos en las diferencias dinámicas que existen entre las proteínas que simulamos. En la figura 19 mostramos los histogramas bidimensionales de apertura y torsión para cada una de ellas. Como mencionamos en la sección de metodologías, estas gráficas constituyen una especie de huella dinámica para las PBP y al compararlas podemos encontrar diferencias en las conformaciones que cada una visita.

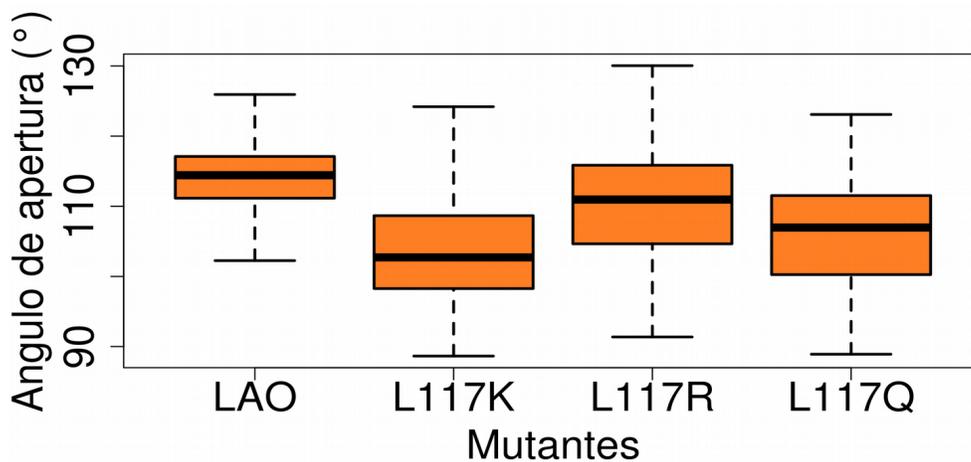


**Figura 18.** Comparación de las distribuciones del porcentaje de residuos sin asignación de estructura secundaria entre LAO y L117K.

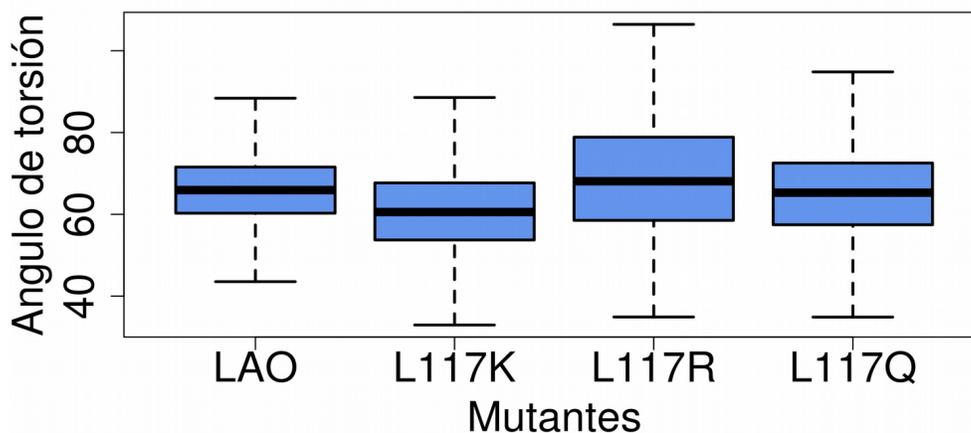
Sorpresivamente, todas las mutaciones puntuales que introdujimos modifican la exploración conformacional de la versión silvestre de LAO en nuestra ventana de muestreo. Las mutantes difieren de la silvestre tanto en la forma que tiene su exploración conformacional como en la manera en la cual la densidad de puntos se reparte en las zonas exploradas. En el panel correspondiente a LAO podemos ver que los puntos en el histograma se encuentran concentrados en una zona pequeña alrededor de una región de alta densidad cuyo centro es el punto  $\langle 115^\circ, 67^\circ \rangle$ , muy próximo a los valores de la estructura cristalográfica. Esta zona de puntos es explorada por las mutantes pero con valores de densidad menor, lo que indica que las conformaciones asociadas a este macroestado no son mayoritarias. Adicionalmente, las mutantes exploran conformaciones con valores menores de apertura, efecto que se aprecia mejor para L117K donde la zona de mayor densidad corresponde al punto  $\langle 102^\circ, 61^\circ \rangle$  y es posible observar densidades altas en conformaciones que tienen valores de apertura menores de  $95^\circ$ . Para observar con mejor detalle las diferencias en los valores que visita cada mutante, en las figuras 20 y 21 mostramos los diagramas de caja y bigotes para las distribuciones de ángulos de apertura y torsión.



**Figura 19.** Histogramas bidimensionales de ángulos de apertura y torsión para LAO y mutantes puntuales. Los valores de apertura y torsión de la estructura cristalográfica de LAO silvestre abierta se marcan con una cruz en todos los paneles



**Figura 20.** Diagramas de caja y bigotes de las distribuciones de ángulo apertura para LAO y mutantes puntuales



**Figura 21.** Diagramas de caja y bigotes de las distribuciones de ángulo de torsión para LAO y mutantes puntuales

Para cada diagrama de caja y bigotes, obtuvimos la mediana así como el rango intercuartil (IQR) que, se relaciona con la dispersión de una distribución. Estas propiedades las concentramos en la Tabla I. Las mutantes presentan valores mayores de IQR tanto en las distribuciones de apertura como en las de torsión, lo que confirma el aumento de la dispersión en la exploración que observamos en los histogramas bidimensionales. La expansión del IQR en las distribuciones de apertura es muy parecida entre las mutantes, oscilando entre 5 y 6° aproximadamente. Para las distribuciones de torsión, el incremento del IQR es variable dentro de las mutantes, siendo L117R la que presenta una expansión mayor (9.8°) y L117K la que más se parece a la nativa (la expansión fue de 2.68°). Las medianas de las distribuciones también mostraron cambios en las mutantes. Para el ángulo de apertura, todas las medianas disminuyeron respecto a la nativa, aunque en grados distintos: L117R disminuyó 3.46°, L117Q lo hizo en 7.46° y L117K en 11.71°. Las medianas asociadas a las distribuciones de torsión tuvieron comportamientos dispares, ya que el valor para L117Q (65.29°) es prácticamente idéntico al de la silvestre (65.93°), L117R presentó un ligero incremento (2.07°) y L117K mostró una disminución (5.36°) para esta propiedad. En conjunto, los cambios en estas propiedades de las distribuciones nos dejan claro que visitadas por las mutantes no suelen ser muestreadas por la versión silvestre de LAO. Adicionalmente, todas las mutantes exhiben la capacidad de visitar estados cerrados y parcialmente cerrados, algo que bajo las mismas condiciones no ocurre en la nativa en la escala de tiempo que muestreamos en nuestras simulaciones. No pasa desapercibido que L117K exhibe los menores valores de apertura entre mutantes. En las secciones que vienen vamos a indagar los motivos por los cuales las mutantes visitan estados diferentes a los de la nativa y los mecanismos que hacen a L117K capaz de muestrear mejor que nadie estas nuevas conformaciones que se cierran espontáneamente.

#### **5.4 Los residuos que se mueven diferente en L117K se ubican en las bisagras**

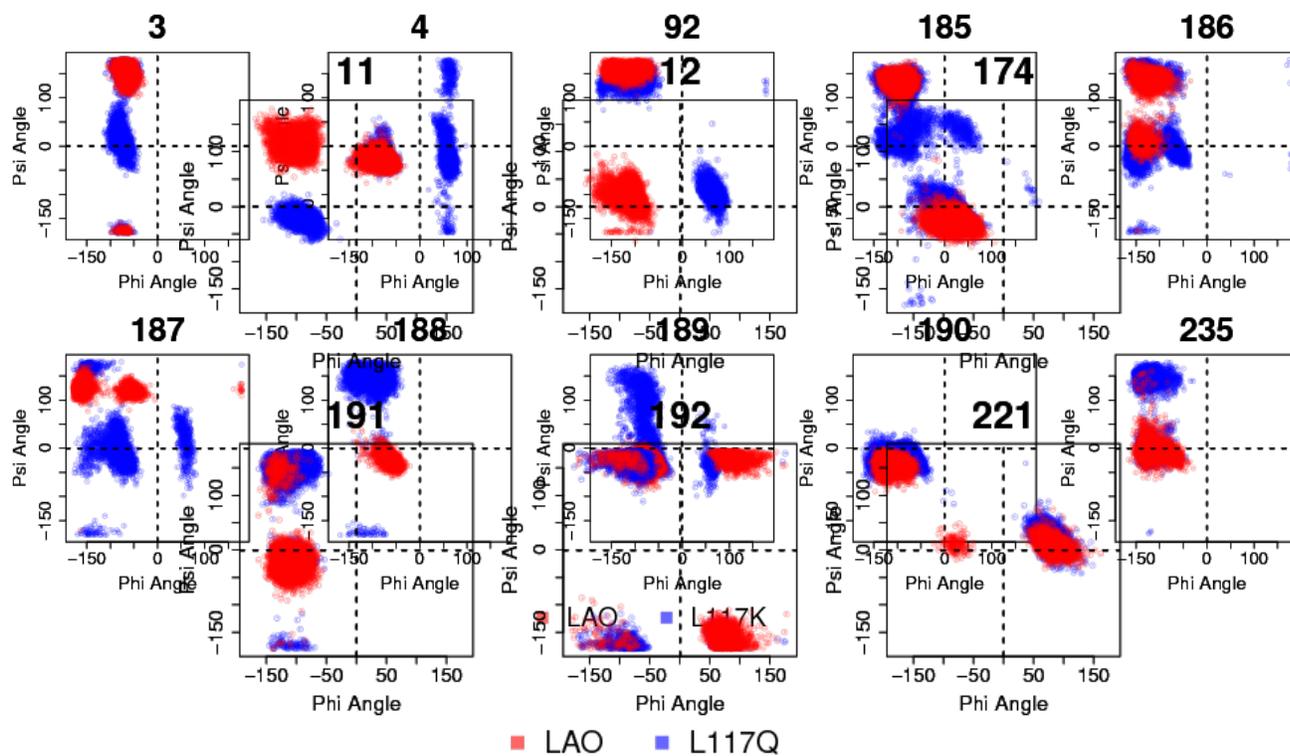
Como primer acercamiento para entender los orígenes de las diferencias que existen entre las conformaciones visitadas por las mutantes y la versión nativa de LAO, comparamos las gráficas de Ramachandran de cada uno de los residuos entre las proteínas que simulamos. Estas gráficas describen el espacio conformacional que un residuo visita a nivel de cadena principal. En las figuras 22 a 24 mostramos los gráficos de Ramachandran de los residuos que mostraron tener valores distintos a la referencia exclusivamente para una sola de las mutantes. Para esto, usamos como criterio los valores de la divergencia de Kullback-Leibler como definimos en la sección de metodologías. La localización de los residuos en la estructura de las mutantes se visualiza en la figura 25.

**Tabla I.** IQR y mediana de las distribuciones de ángulos de apertura y torsión para LAO y sus mutantes

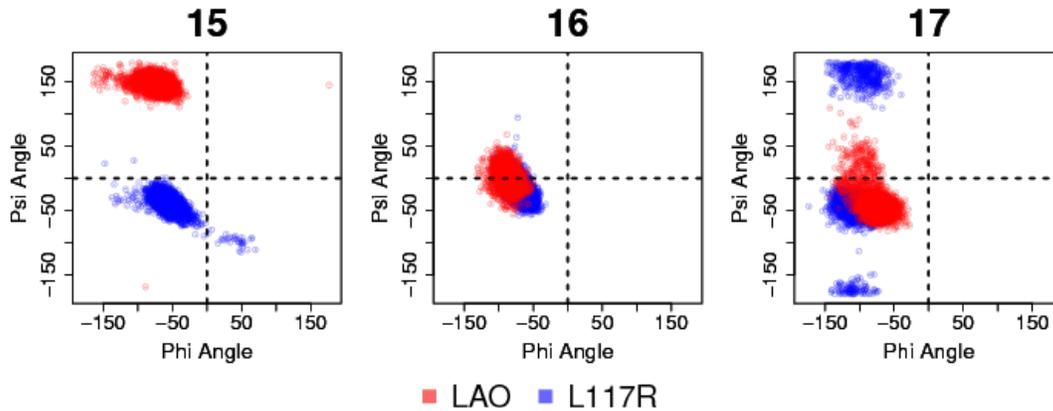
<i>Proteína</i>	<b>Distribuciones de apertura</b>		<b>Distribuciones de torsión</b>	
	<i>IQR (°)</i>	<i>Mediana (°)</i>	<i>IQR (°)</i>	<i>Mediana (°)</i>
<b>LAO</b>	5.95	114.43	11.27	65.93
<b>L117K</b>	10.38	102.71	13.95	60.57
<b>L117R</b>	11.20	110.96	20.35	68.07
<b>L117Q</b>	11.27	106.97	15.09	65.29

Sorpresivamente, en ninguna de las mutantes el residuo 117 ni los que se encuentran en su periferia mostraron diferencias significativas. L117K es la mutante que presenta la mayor cantidad de residuos con movilidad diferencial (10 en total). Estos se distribuyen mayoritariamente en una zona continua de 7 residuos localizados en las bisagras, lo cual muestra correspondencia con la capacidad de muestrear conformaciones cerradas que tiene esta mutante. En L117Q, seis residuos tienen movilidad diferencial y se encuentran dispersos en diversas zonas de la estructura (solo dos de ellos en la bisagra), mientras que los tres residuos que mostraron movilidad diferencial en L117R se encuentran en un loop del lóbulo inferior que se ha asociado con la capacidad que tiene LAO de interactuar con el transportador. En la figura 25 mostramos que solo dos residuos del extremo C-terminal (Y236 y G237) presentaron movilidad diferencial significativa para todas las mutantes. Si bien este análisis nos permite corroborar que las diferencias en las conformaciones que las mutantes visitan se relacionan con cambios en el espacio conformacional que ciertos residuos muestrean, no explica por qué existe esta movilidad diferencial. Para saber eso necesitamos conocer cómo se modificaron las conectividades internas en las proteínas que simulamos, lo que requiere como paso preliminar encontrar cuáles son las estructuras significativas asociadas a cada conformación visitada para las mutantes, para lo que usaremos algoritmos de agrupamiento, cuya implementación y resultados vamos a describir en la sección que sigue.

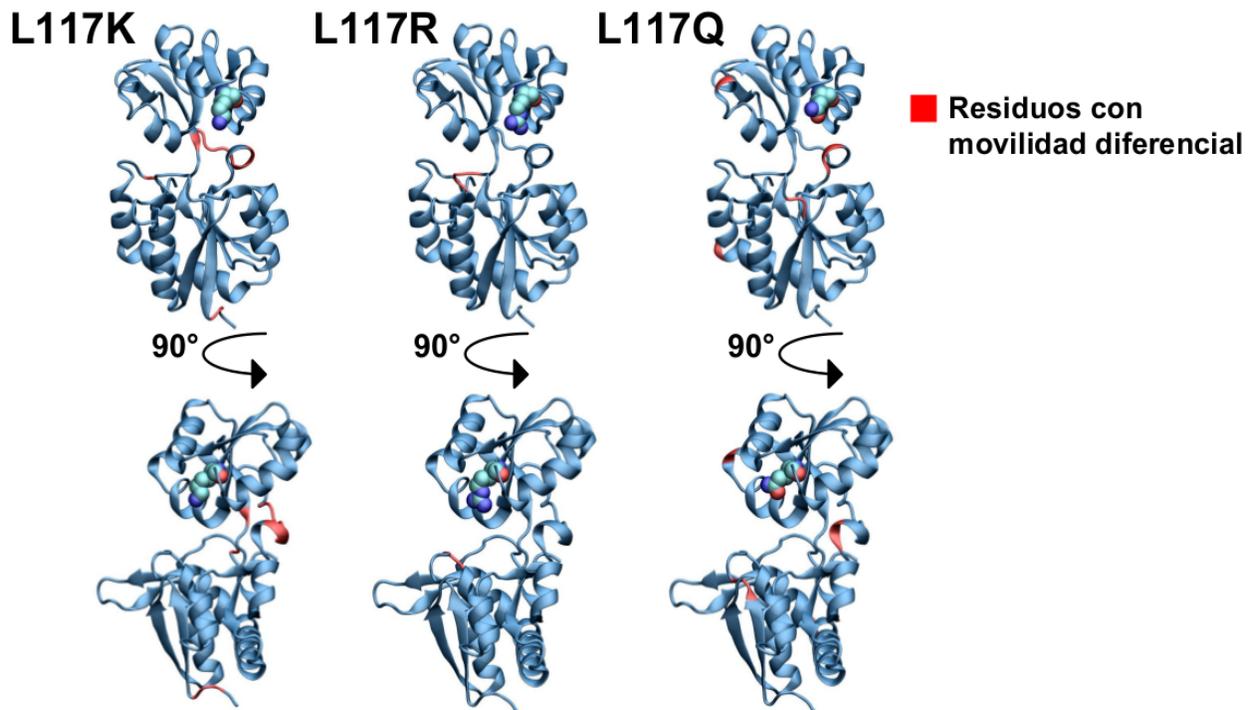
**Figura 22.** Gráficos de Ramachandran de residuos que tienen movilidad diferencial exclusivamente en L117K



**Figura 23.** Gráficos de Ramachandran de residuos que tienen movilidad diferencial exclusivamente en L117Q



**Figura 24.** Gráficos de Ramachandran de residuos que tienen movilidad diferencial exclusivamente en L117R

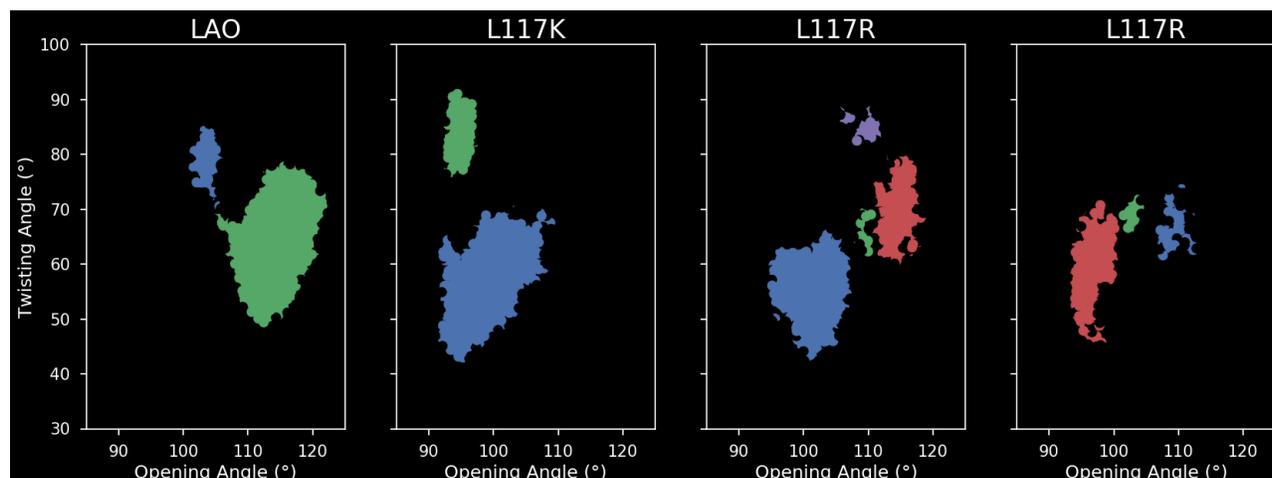


**Figura 25.** Localización de los residuos con movilidad diferencial en la estructura de las mutantes. El residuo 117 se resalta en una representación de esferas.

### 5.5 Agrupar por apertura y torsión revela las conformaciones significativas

Para estudiar cómo se modificaron las conectividades internas de las mutantes a lo largo de las simulaciones necesitamos encontrar grupos de estructuras que sean parecidas entre sí para que las diferencias que encontremos sean significativas y fácilmente identificables. Para esto, utilizamos el algoritmo de agrupamiento HDBSCAN como se encuentra implementado en la biblioteca del mismo nombre de Python. Las propiedades que usamos para llevar a cabo el

agrupamiento fueron los valores de apertura y torsión que describimos previamente. HDBSCAN es un algoritmo que funciona reconociendo zonas de alta densidad en la distribución de datos, de manera que no solo clasifica las observaciones sino que permite discernir entre las que pertenecen a una zona de alta densidad de las que no. En la figura 26 mostramos el resultado del agrupamiento para cada una de las proteínas que simulamos. En negro, se resaltan los puntos que no pertenecen a ningún grupo y fueron clasificados como ruido. A diferencia de figuras anteriores, los ángulos de apertura y de torsión se reportan en radianes debido a que el tipo de distancia que utilizamos funciona con esas unidades.



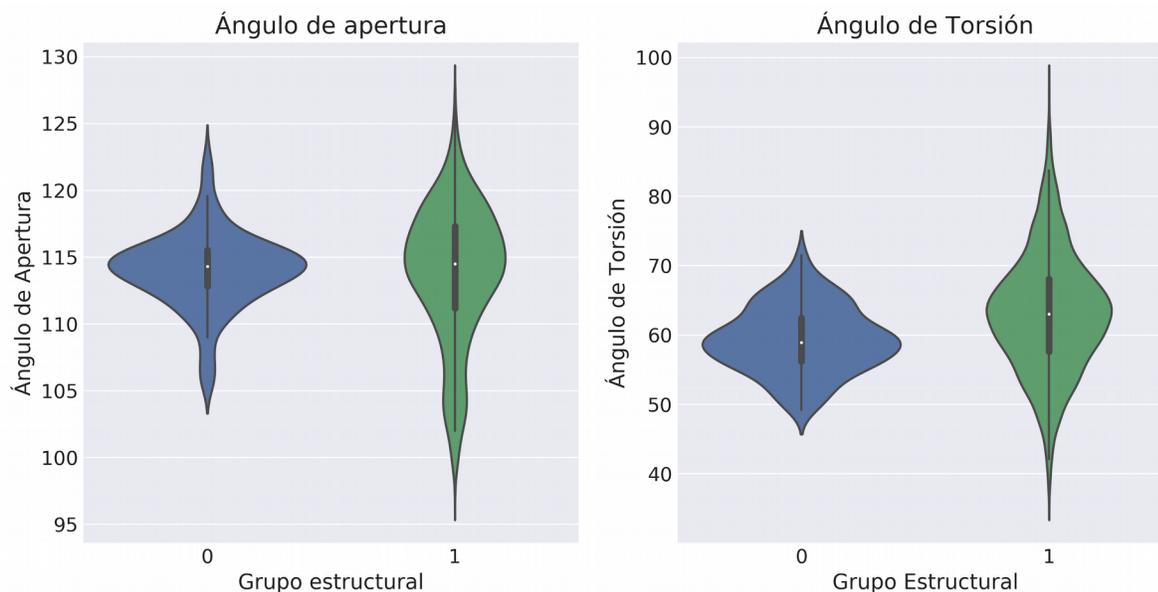
**Figura 26.** Clasificación de las estructuras provenientes de la simulación usando el algoritmo de agrupamiento basado en densidad HDBSCAN. En negro se muestran los puntos que no fueron catalogados dentro de algún grupo y se consideraron ruido

El número de grupos estructurales encontrados para cada proteína simulada, el tamaño de cada uno de estos grupos, así como el número de puntos que fueron clasificados como ruido para cada mutante, se muestran en la Tabla II. La numeración de los grupos comienza a partir del cero ya que Python empieza a contar desde cero. Tuvimos dos patrones en el agrupamiento. En el primero de ellos, constituido por LAO y L117K, observamos dos grupos claramente definidos, siendo uno de ellos mayoritariamente poblado ( 48.98 % de la población estructural de la simulación). En el segundo, que comparten L117R y L117R, hay varios grupos de tamaño pequeño y predominan los puntos que son clasificados como ruido. Este resultado muestra que L117K es la única mutante que tiene al menos un grupo (grupo 0) estructural altamente poblado, lo cual se relaciona con la existencia de un subestado definido dentro la simulación parecido al existente en la referencia silvestre. El grupo más poblado de L117R (grupo 0), aunque se encuentra en la misma zona del espacio conformacional que el

mayoritario de L117K, presenta una población considerablemente menor que éste (20.32% de las estructuras) y ninguno de los grupos de L117Q representan un porcentaje mayor al 12% de la estructuras.

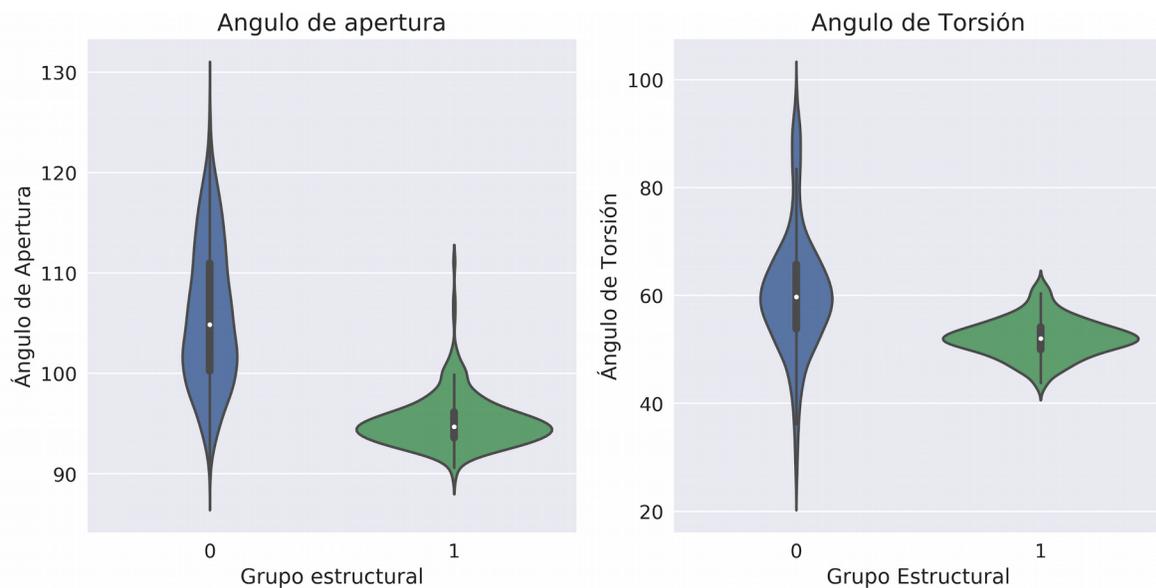
**Tabla II.** Número de miembros en los grupos estructurales en los que fueron clasificadas las proteínas simuladas

Grupo estructural	LAO	L117K	L117R	L117Q
Ruido	1398	2292	2972	4002
0	158	2449	1016	329
1	3444	259	82	113
2	-	-	804	556
3	-	-	126	-
<b>Total de estructuras</b>	<b>5000</b>	<b>5000</b>	<b>5000</b>	<b>5000</b>

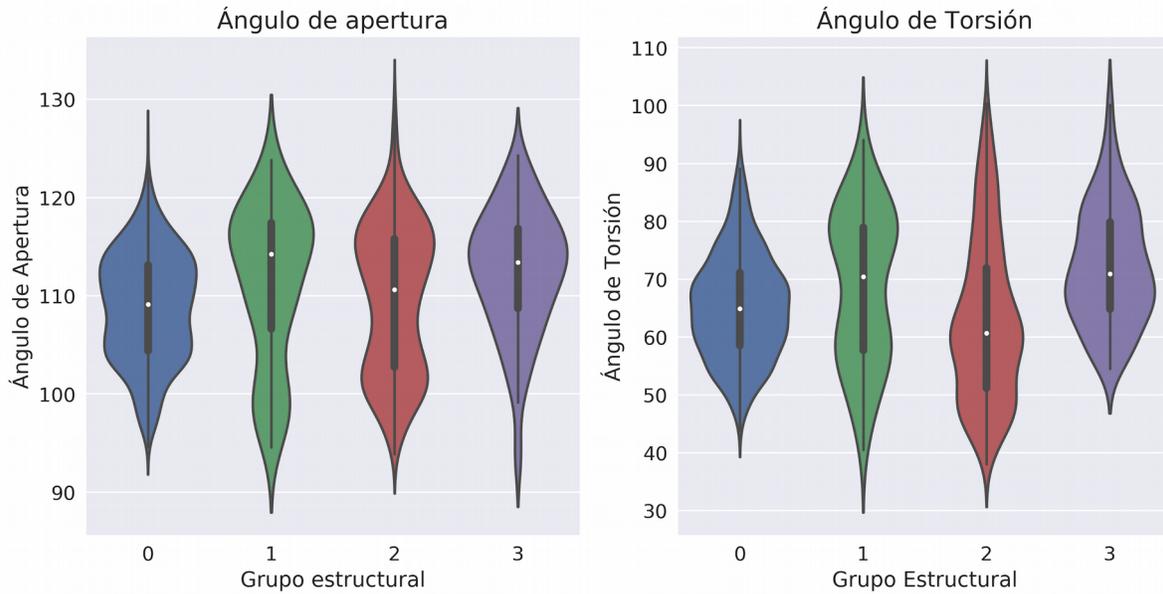


**Figura 27.** Diagramas de violín para las distribuciones de apertura y torsión para los grupos estructurales de LAO.

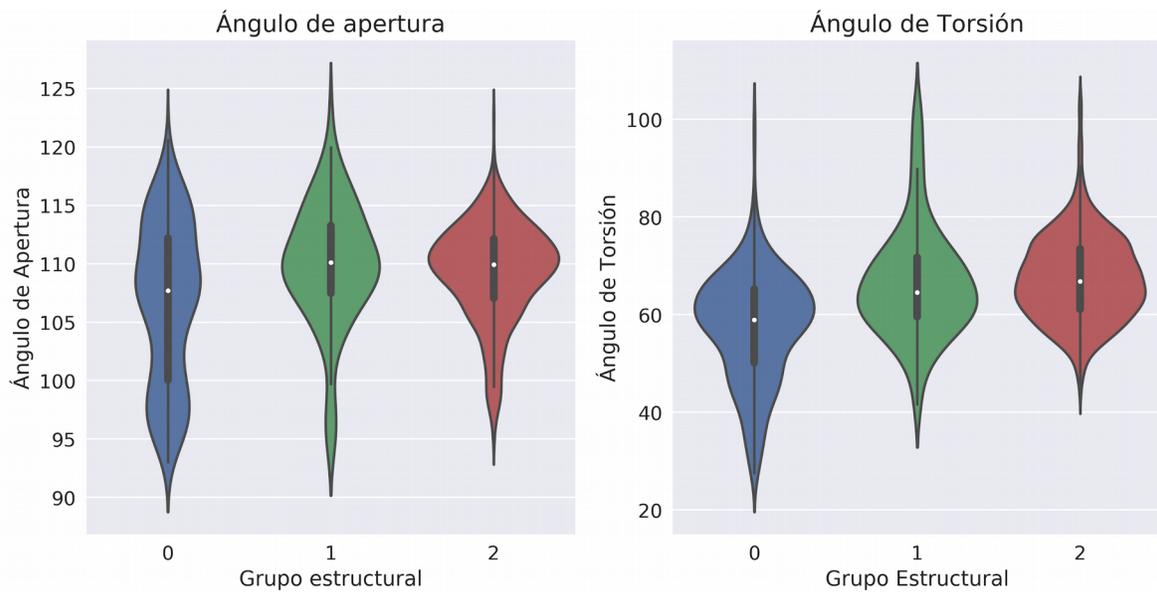
En las figuras que van de la 27 a la 30 mostramos los diagramas de violín para cada uno de los grupos que se encontraron para las proteínas simuladas. Este tipo de diagrama, al igual que los de caja y bigotes, nos permite comparar rápidamente la forma y los valores asociados a distintas distribuciones de apertura y torsión de las estructuras que pueblan cada grupo.



**Figura 28.** Diagramas de violín para las distribuciones de apertura y torsión para los grupos estructurales de L117K.



**Figura 29.** Diagramas de violín para las distribuciones de apertura y torsión para los grupos estructurales de L117R

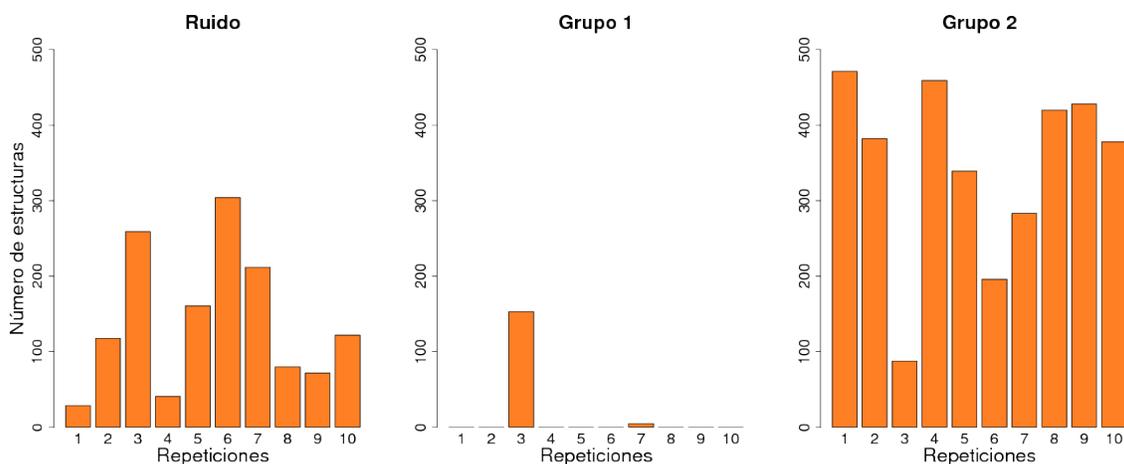


**Figura 30.** Diagramas de violín para las distribuciones de apertura y torsión para los grupos estructurales de L117Q

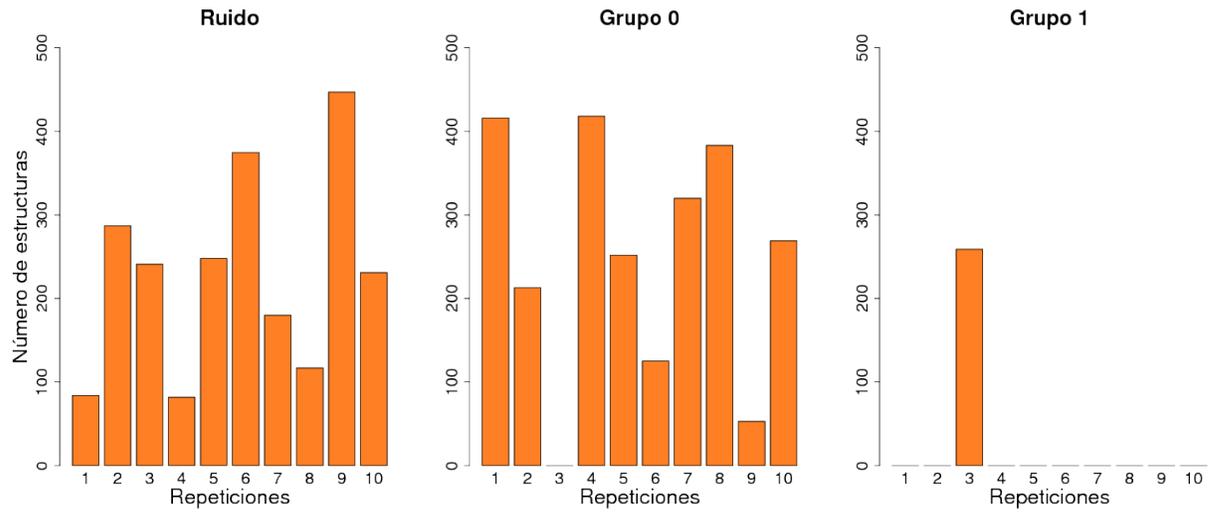
Los diagramas de violín nos muestran que las diferencias entre grupos estructurales es mucho más pronunciada para los casos de L117K y que para L117R y L117Q. Al filtrar por grupos estructurales también se hace más evidente que L117K es la mutante que visita

conformaciones con valores menor de ángulo de apertura. De hecho, los dos grupos estructurales que fueron encontrados para esta mutante con el algoritmo de agrupamiento son los que exhiben los valores más pequeños en sus medianas (104 y 95°, respectivamente). Ninguno de los grupos estructurales de L117Q o L117R exhibe un valor de apertura menor a 107°.

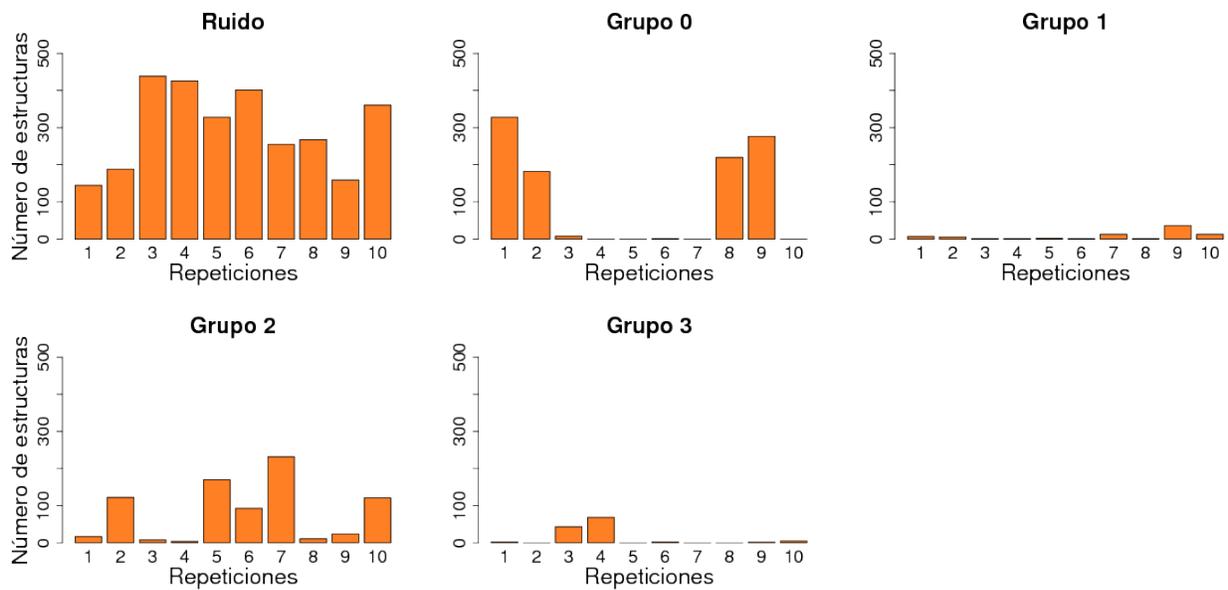
Mostramos la composición por réplica para cada uno de los grupos estructurales de las proteínas que simulamos en las figuras que van 31 a la 34. Este análisis nos permite saber si los grupos se componen de estructuras provenientes de la totalidad de las simulaciones o bien, si esas conformaciones solo pueden ser asociadas a un grupo reducido o una sola de las repeticiones. Las figuras nos muestran que en su mayoría, los grupos en los que se clasificaron las proteínas que simulamos se componen por estructuras de distintas simulaciones. Las excepciones más grandes a este comportamiento fueron 2: El grupo 0 de LAO, el 1 de L117K. En la excepción de LAO, casi la totalidad (96%) de las estructuras que fueron clasificadas en este grupo pertenecen a la repetición número 3. En el caso del grupo 1 de L117K, todas las estructuras que le componen son de la repetición número 3. Hay algunos otros casos donde hay una réplica que contribuye muy por encima a las demás (en el grupo 2 de L117Q, el 57.4 % de las estructuras provienen de la simulación 3). Este patrón de concentración en ciertas repeticiones no quiere decir que debemos desechar la información estructural que nos aportan estos grupos sino que debemos tratar con cuidado la abundancia con la que aparentemente los observamos ya que su prevalencia podría deberse a las limitaciones asociadas al muestreo. Estos resultados recalcan la necesidad de mejorar nuestro muestreo incrementando la cantidad de repeticiones y posiblemente comenzar nuestra exploración a partir de estructuras de estos grupos para ver si las simulaciones nos conducen a otros puntos del espacio conformacional que no hemos explorado o bien, convergen a los grupos estructurales que ya conocemos.



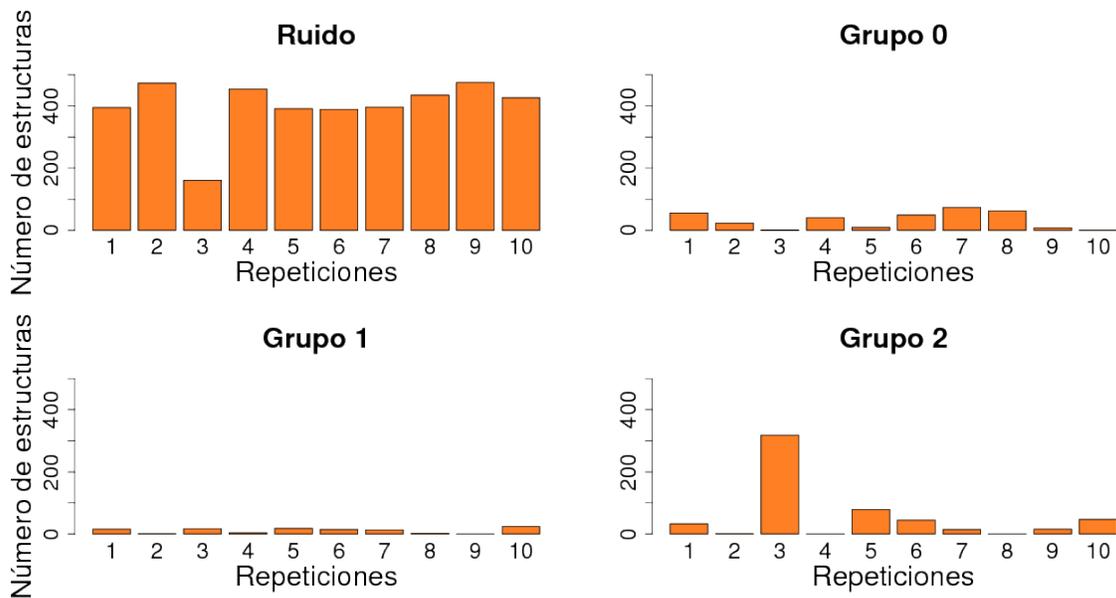
**Figura 31.** Contribución por réplica de los grupos estructurales de LAO



**Figura 32.** Contribución por réplica de los grupos estructurales de L117K



**Figura 33.** Contribución por réplica de los grupos estructurales de L117R

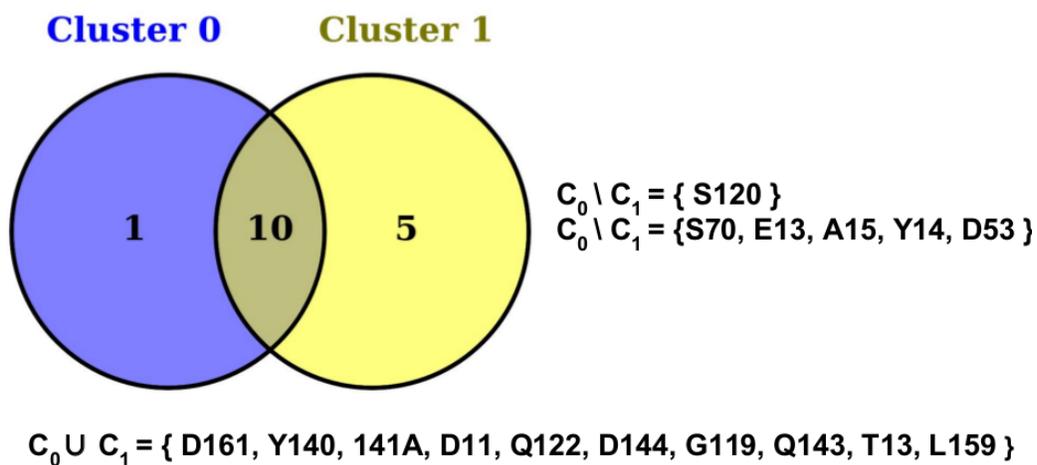


**Figura 34.** Contribución por réplica de los grupos estructurales de L117Q

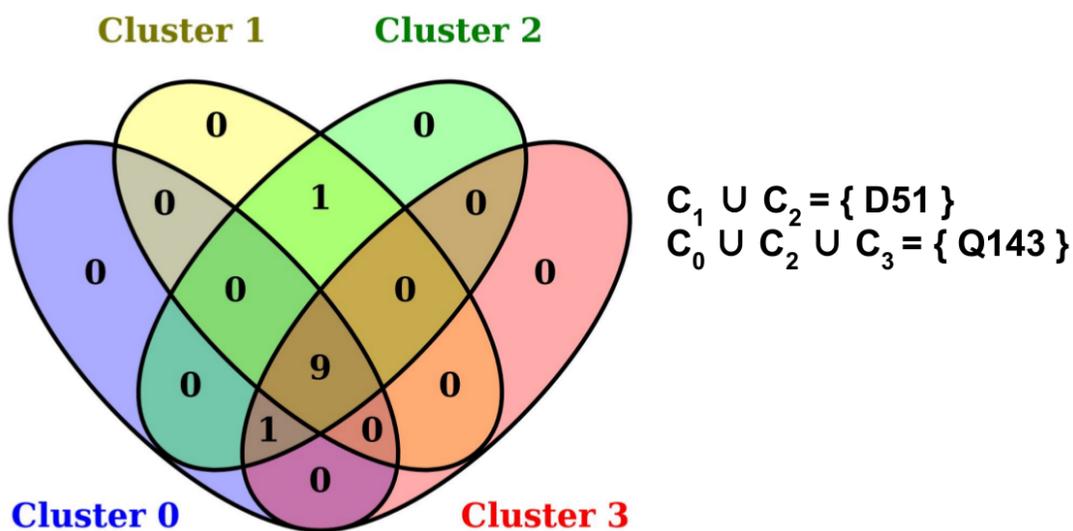
### 5.6 Las mutaciones en el residuo 117 brindan nuevas posibilidades de interacción

Recurrimos a los cambios en las conectividades internas para explicar las diferencias en los valores visitados de apertura y torsión que existen en los grupos en los cuales clasificamos a las estructuras provenientes de las simulaciones. Como primer acercamiento, indagamos en las interacciones que establece el residuo 117 en la silvestre y las mutantes. Para esto, calculamos la energía de interacciones no enlazadas, es decir, la suma de la energía asociada a interacciones electrostáticas y a las fuerzas de Van der Waals. Los valores se obtuvieron usando la función *lie* de *cpptraj* como lo describimos en la sección de metodologías. En las tablas III a VI reportamos los residuos que exhiben una energía de interacción no covalente promedio menor de -2.0 kcal/mol, para cada uno de los grupos estructurales. El residuo 117 (L117) en la versión silvestre de LAO establece una cantidad limitada de interacciones con residuos localizados en su periferia y que se conservan en los dos grupos estructurales, lo cual denota la robustez de esas relaciones químicas. Las mutantes expanden el potencial de interacción del residuo 117, aunque en grados distintos. L117Q conserva la totalidad de las interacciones que la silvestre establece más otras adicionales con residuos cercanos que se encuentran en el lóbulo superior. L117R y L117K muestran una expansión notable en la capacidad de interacción de su residuo 117, que tiene la capacidad de interactuar con residuos que se ubican en el lóbulo inferior como es el caso de D11, que es una interacción que ambas mutantes establecen en todos sus grupos estructurales y que tiene un valor considerablemente

negativo. La manera en que se distribuyen las interacciones en los distintos grupos estructurales no es homogénea, como puede verse en los diagramas de Venn que mostramos en las figuras que van de la 35 a la 37. En estas representaciones omitimos a LAO silvestre porque sus dos grupos estructurales presentan las mismas interacciones. Acompañamos los diagramas con valores de intersecciones y diferencias que consideramos significativas. L117R y L117Q presentan interacciones muy similares en todos sus grupos estructurales. Ninguno de los grupos estructurales presenta interacciones únicas y la mayoría de éstas se concentran en la intersección de todos los grupos, lo que quiere decir que se presentan en todos los clusters. Por otro lado, L117K si muestra una diferencia marcada en las interacciones que cada uno de sus grupos estructurales establece. El grupo 1 establece 5 interacciones exclusivas con residuos del lóbulo inferior. Adicionalmente, carece de la interacción que establecen K117 y S120, que se encuentra presente en todos los grupos estructurales de las demás proteínas que simulamos. La ausencia de esta interacción parece dejar libre a K117 para interactuar con una mayor cantidad de residuos, desencadenando reconfiguraciones adicionales de la conectividad interna lo cual permite que esta mutante visite estados con valores menores del ángulo de apertura. Para probar esta suposición tenemos que encontrar una manera de estudiar la globalidad de las conexiones internas que se establecen en esta mutante para poderlas comparar con el contexto silvestre. Indagaremos en eso en secciones posteriores pero antes, vamos a evaluar si la electrostática del residuo introducido es la fuerza que desencadena estas nuevas interacciones o es necesario considerar otros elementos asociados a la modificación.

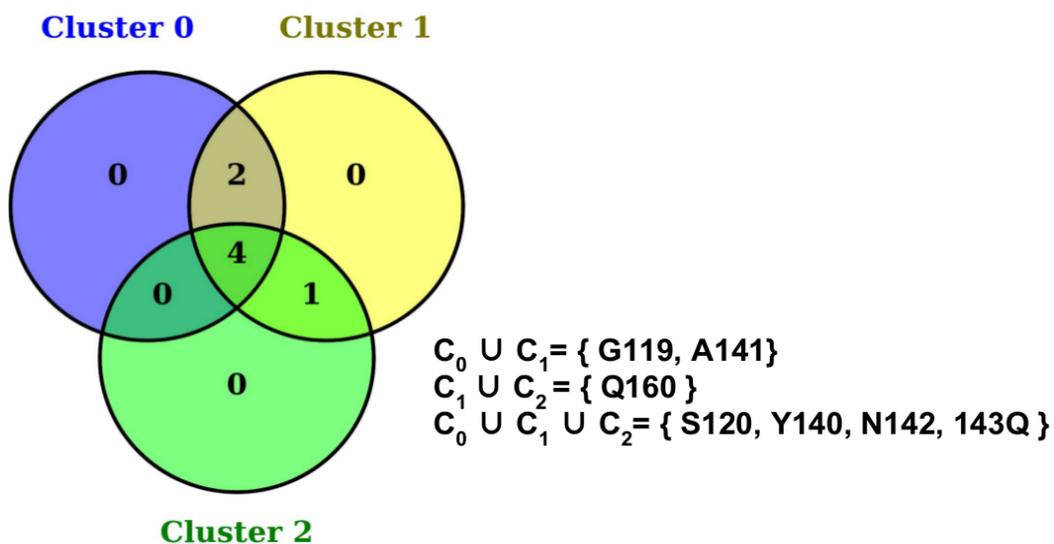


**Figura 35.** Diagrama de Venn de las interacciones que establece el residuo 117 en los distintos grupos estructurales de L117K



$$C_0 \cup C_1 \cup C_2 \cup C_3 = \{ D161, Y140, D144, A141, D11, Q122, G119, S120, L159 \}$$

**Figura 36.** Diagrama de Venn de las interacciones que establece el residuo 117 en los distintos grupos estructurales de L117R



**Figura 37.** Diagrama de Venn de las interacciones que establece el residuo 117 en los distintos grupos estructurales de L117Q

**Tabla III.** Residuos con las menores de energía interacción con L117 en LAO

	Grupos estructurales	
	Cero (Kcal/mol)	Uno (Kcal/mol)
G119	-8.70	-8.75
S120	-5.55	-5.38
Y140	-2.18	-2.15

**Tabla IV.** Residuos con las menores de energía interacción con Q117 en L117Q

	Grupos estructurales		
	Grupo Cero	Grupo Uno	Grupo Dos
S120	-7.972	-8.183	-8.174
Y140	-6.014	-6.124	-5.640
Q143	-4.506	-4.349	-4.220
A141	-----	-2.929	-----
N142	-3.035	-2.768	-3.218
Q160	-----	-2.262	-3.329
G119	-2.002	-2.146	-----

**Tabla V.** Residuos con las menores de energía interacción con K117 en L117K

	Grupos estructurales	
	Grupo cero (Kcal/mol)	Grupo uno (Kcal/mol)
<b>D11</b>	-5.97	-38.25
<b>D161</b>	-24.28	-23.49
<b>Y140</b>	-8.81	-9.63
<b>A141</b>	-8.08	-6.35
<b>Q122</b>	-3.72	-4.11
<b>S70</b>	-----	-3.20
<b>G119</b>	-3.15	-3.06
<b>S120</b>	-2.06	-----
<b>E123</b>	-----	-2.87
<b>D144</b>	-3.25	-2.76
<b>A15</b>	-----	-2.53
<b>Y14</b>	-----	-2.34
<b>D53</b>	-----	-2.30
<b>T13</b>	-2.53	-----
<b>Q143</b>	-2.89	-2.30
<b>L159</b>	-2.46	-2.10

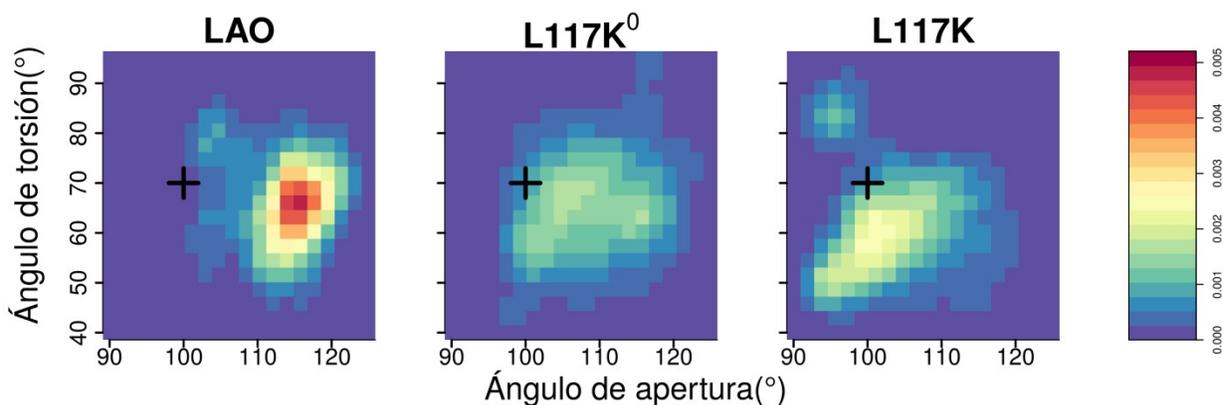
**Tabla VI.** Residuos con las menores de energía interacción con R117 en L117R

	Grupos estructurales			
	Grupo Cero (Kcal/mol)	Grupo Uno (Kcal/mol)	Grupo Dos (kcal/mol)	Grupo Tres (Kcal/mol)
D11	-7.254	-11.774	-21.465	-3.308
Y140	-9.460	-9.452	-9.547	-9.543
A141	-7.277	-7.990	-7.296	-6.197
D51	-----	-5.591	-4.437	----
D161	-18.578	-4.729	-7.672	-5.168
G119	-3.057	-4.172	-4.564	-3.156
Q122	-4.296	-4.080	-3.835	-4.894
D144	-3.605	-3.119	-2.415	-----
S120	-2.423	-2.600	-2.623	-3.622
Q143	-3.604	-----	-2.344	-2.773
L159	-2.342	-2.073	-2.063	-2.577

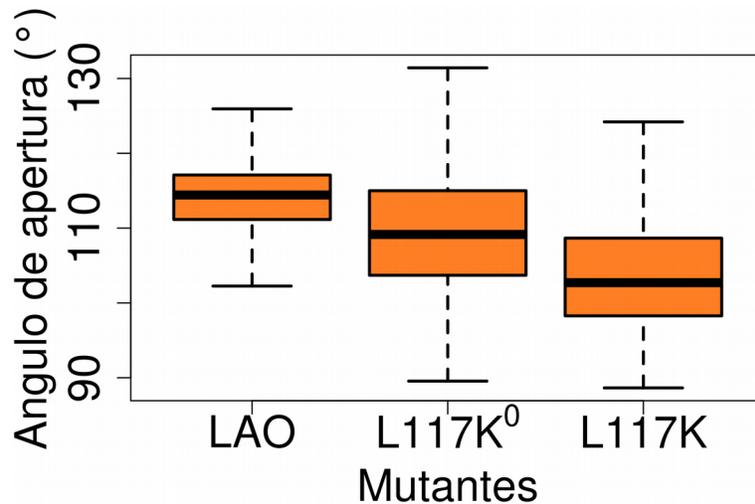
### 5.7 Simular con una lisina neutral reduce el cierre espontáneo de L117K

¿Es suficiente introducir un residuo con una cadena lateral con carga positiva para expandir la capacidad de interacción del residuo 117? L117R y L117K parecen contestar afirmativamente esta pregunta: en ambas mutantes se expanden las capacidades de interacción del residuo 117 y eso se relaciona con una disminución de sus valores de apertura. No obstante, las conformaciones que visita L117R no llegan al grado de reducción de la apertura de L117K, lo que nos sugiere que necesitamos elementos adicionales para explicar las diferencias. Para corroborar si en efecto la carga positiva se queda corta explicando la ganancia de interacciones del residuo 117 en L117K, simulamos sustituyendo a K117 por una versión alternativa de este aminoácido que carece de un protón y por consecuencia tiene neutralidad de carga en su cadena lateral. Al residuo que resultó de esta modificación le llamamos K<sup>0</sup>117 y por consecuencia a la mutante que salió de este proceso, L117K<sup>0</sup>. Corrimos 10 repeticiones independientes de 50 ns y simulamos usando el mismo esquema que usamos para las otras mutantes y describimos en el apartado de metodologías. Comparamos la huella dinámica de L117K<sup>0</sup> con LAO silvestre y L117K (figura 38). Encontramos que la nueva mutante es incapaz de visitar los valores reducidos de apertura que muestrea L117K, además de tener una

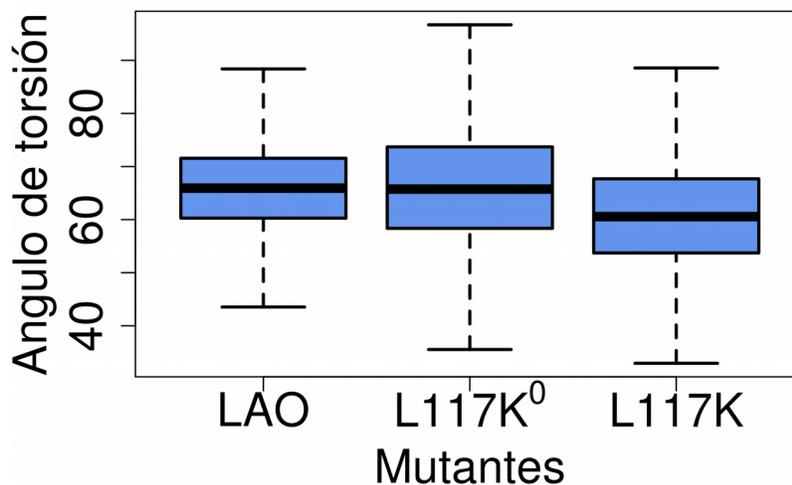
exploración estructural mucho más dispersa que ésta y la nativa. Los diagramas de caja y bigotes de las distribuciones de apertura y torsión (figura 39 y 40, respectivamente) nos indican que L117K<sup>0</sup> tiene un valor mediano de apertura intermedio entre el de LAO y L117K mientras que su valor mediano de torsión es prácticamente idéntico al de la silvestre. Mostramos la manera en la que se agrupan las conformaciones de L117K<sup>0</sup> en la figura 41 y los diagramas de violín de las distribuciones de apertura y torsión de sus grupos estructurales en la figura 42. Las estructuras de L117K<sup>0</sup> se dividen en dos grupos que abarcan el 84.92% de las conformaciones exploradas (el resto entraron en la categoría de ruido). En el grupo mayoritario (uno) cayeron el 83.38 % de las estructuras mientras que el grupo cero solamente el 1.54%. Los diagramas de violín muestran que las conformaciones asociadas a ambos grupos estructurales poseen valores de apertura lejanos a conformaciones cerradas que sí se visitan en L117K. En el grupo más abundante existe una preferencia indistinta para visitar valores de apertura dentro del rango de 100 a 115°. Esto difiere del grupo conformacional mayoritario en LAO silvestre, que muestra un pico de la distribución de apertura de su grupo mayoritario en un 115°.



**Figura 38.** Huella dinámica de L117K<sup>0</sup> y su comparación con la nativa y L117K. Los valores de la conformación abierta cristalográfica de LAO silvestre se muestran con una cruz

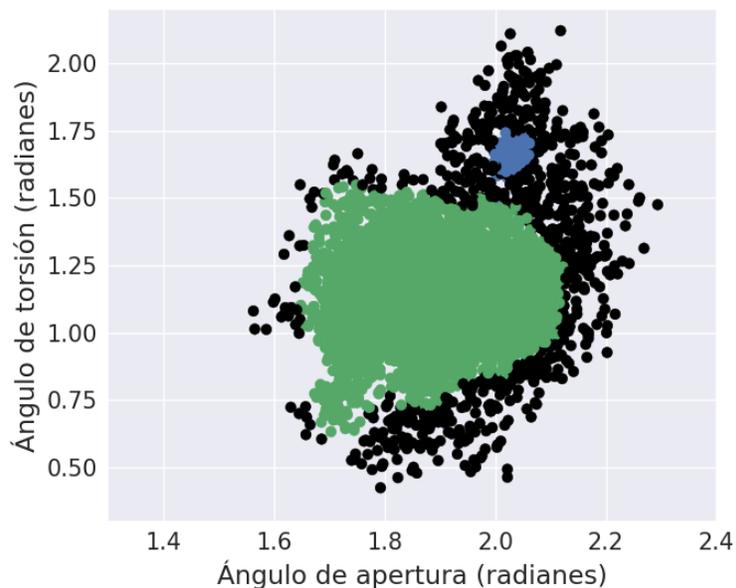


**Figura 39.** Diagramas de caja y bigotes para las distribuciones de apertura de L117K<sup>0</sup> y su comparación con L117K y LAO



**Figura 40.** Diagramas de caja y bigotes para las distribuciones de ángulo de torsión de L117K<sup>0</sup> y su comparación con L117K y LAO

Queda claro que L117K<sup>0</sup> deja de tener un cierre espontáneo pronunciado pero no recupera la firma dinámica de la silvestre por completo. Si sumamos estas observaciones a las de los ángulos de torsión y les asignamos en conjunto un significado estructural podemos concluir que los lóbulos que componen a esta nueva mutante se acercan entre ellos un poco más que los de la nativa pero no se torsionan lo suficiente para reducir su distancia al grado que lo hace L117K. Este conjunto de resultados, más lo que vimos en la dinámica de L117R, nos muestra que la explicación del cierre espontáneo precisa de elementos adicionales a la carga.



**Figura 41.** Distribución de grupos estructurales para L117K<sup>0</sup>

### 5.8 Los efectos de cambiar el residuo 117 se propagan a zonas distantes de la estructura

La manera en la cual las conectividades internas se modificaron en las mutantes va más allá de la conexiones que establece el residuo 117. En este trabajo, usamos mapas de contacto para estudiar la manera en que la conectividad interna de LAO se modificó con las mutaciones que introdujimos. Para generar estas representaciones usamos la suite CONAN, que contiene una serie de aplicaciones para generarlos y analizarlos. Con la información que nos arrojaron los mapas de contacto, identificamos los residuos que participan en un mayor número de interacciones intermoleculares significativas, a los cuales a partir de este segmento del texto vamos a denominar “residuos de alta conectividad”. Los residuos de alta conectividad para cada cluster de cada uno de los sistemas se muestran en las tablas VII a X. La identidad de ellos varía ligeramente entre sistemas y entre grupos estructurales, lo que prueba que a pesar de conservar la misma topología, las proteínas que estudiamos exploran distintos universos de conectividad interna. Usamos esta información para entender qué es lo que hace tan especiales, en términos de sus ángulos de apertura, a las estructuras que pertenecen al cluster 1 de L117K.

**Tabla VII.** Residuos con mayor número de conexiones para LAO

Grupo estructural 0		Grupo estructural 1	
Residuo	Conexiones	Residuo	Conexiones
I68	21	L94	20
L94	20	F210	20
F210	20	L159	20
F191	19	I95	18
L159	18	I68	18
L87	18	L214	18
I73	18	F191	18
L59	18	L150	18
L55	18	W130	18
L214	18	L87	17

**Tabla VIII.** Residuos con mayor número de conexiones para L117K

Grupo estructural 0		Grupo estructural 1	
Residuo	Conexiones	Residuo	Conexiones
F210	20	L159	20
L159	19	W130	19
L94	18	L94	18
I68	18	L59	18
Q37	18	I31	18
L214	18	F210	18
L150	18	S18	18
I146	18	Q160	18
A95	17	L150	18
I67	17	I146	18

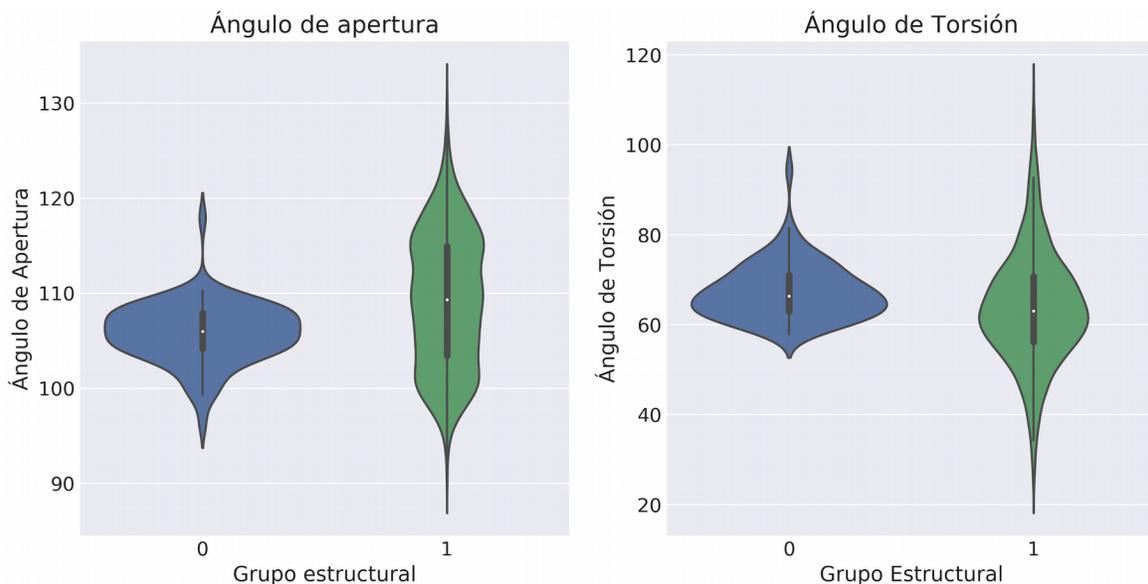
**Tabla IX.** Residuos con mayor número de conexiones para L117R

Grupo estructural 0		Grupo estructural 1		Grupo estructural 2		Grupo estructural 3	
Residuo	conexiones	Residuo	Conexiones	Residuo	Conexiones	Residuo	Conexiones
L94	20	L94	20	L94	20	L94	20
F210	20	L159	20	F210	20	F210	20
L159	20	L59	18	L159	19	I68	19
L214	18	F210	18	L59	18	L159	19
W130	18	L150	18	L150	18	Q37	18
V114	18	W130	18	W130	18	L33	18
I68	17	V114	18	I95	17	L214	18
I67	17	I95	17	I67	17	L150	18
L59	17	I68	17	Q37	17	W130	18
L55	17	I67	17	L33	17	V114	18

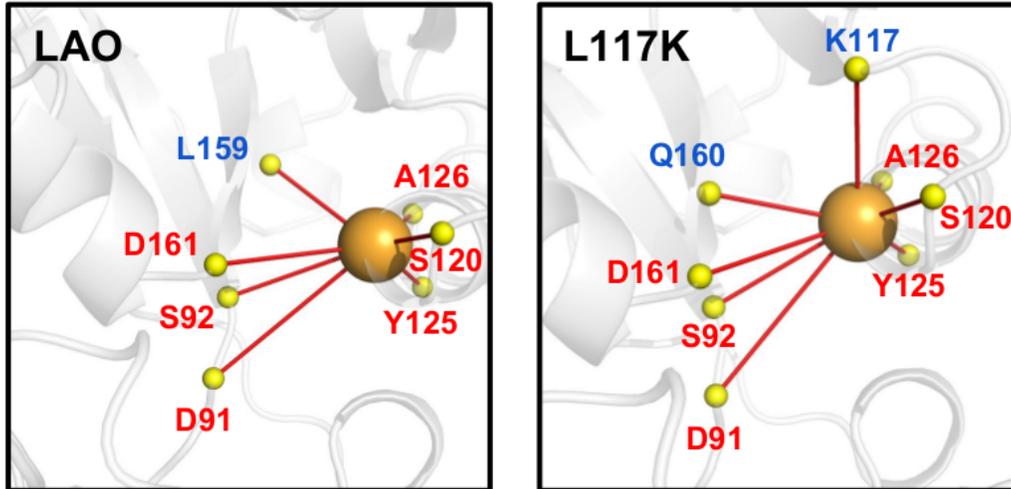
**Tabla X.** Residuos con mayor número de conexiones para L117Q

Grupo estructural 0		Grupo estructural 1		Grupo estructural 2	
Residuo	Conexiones	Residuo	Conexiones	Residuo	Conexiones
L94	20	L94	20	L94	20
L59	18	F210	20	F210	20
F210	18	L59	18	L159	20
L159	18	L159	18	L59	18
W130	18	L150	18	L150	18
I95	17	W130	18	I146	18
I67	17	I95	17	W130	17
L55	17	I68	17	I95	17
L33	17	I67	17	I67	17
I31	17	L55	17	D65	17

Para entender el cierre espontáneo que ocurre en ellas, nuestra apuesta es cazar las interacciones que establecen los residuos que experimentan cambios en sus patrones de conectividad interna. Para esto, ubicamos los residuos de alta conectividad que son únicos de este grupo estructural. Encontramos que el único que clasifica en esta categoría es Q122. Al haberlo identificado, buscamos los residuos con los que presenta una mayor energía de interacción electrostática promedio, calculada por la función *lie* de *cpptraj*. A estos residuos los nombramos a partir de este punto del texto como “pares de interacción significativos”. Los criterios para seleccionarlos están definidos en el área pertinente de la sección de materiales y métodos. Q122 en las estructuras del cluster 1 de L117K tiene 8 pares de interacción significativos: D91, S92, K117, S120, Y125, A126, Q160 Y D161. En el panel derecho de la figura 43 podemos ver dónde se localizan estos residuos en L117K. En el panel izquierdo de esta figura mostramos los pares de interacción energéticamente significativos que establece Q122 en las estructuras del cluster 1 de LAO, arquetipo del estado abierto que usamos de referencia para entender cuáles modificaciones de la conectividad interna abonan al cierre espontáneo.

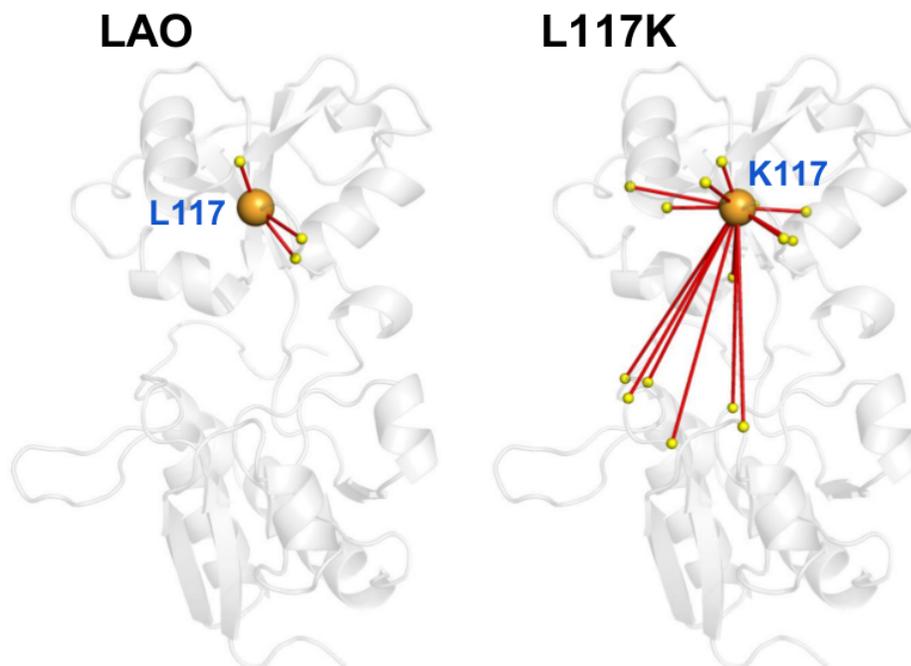


**Figura 42.** Diagrama de violín para las distribuciones de apertura y torsión para las grupos estructurales de L117K<sup>0</sup>. El color de los grupos coincide con el de la figura anterior.

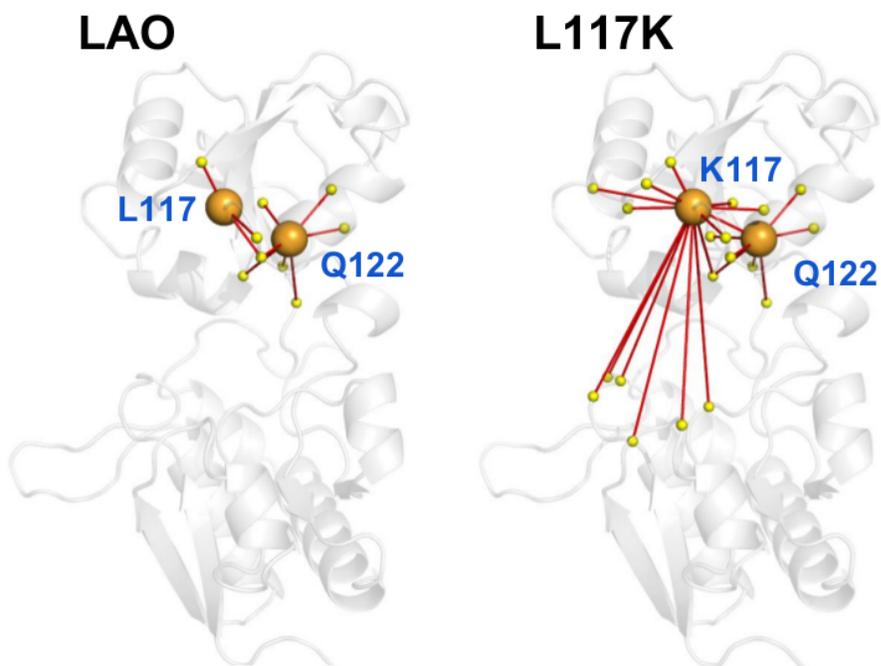


**Figura 43.** Pares de interacción energéticamente significativos de Q122 en L117K (derecha) y LAO (izquierda)

En la figura 43, dibujamos los carbonos alfa de los residuos como esferas. En naranja resaltamos a la Q122 y en amarillo a los residuos que funcionan como pares de interacción significativa. Los residuos en los que coloreamos el nombre de color azul son aquellos que exhiben diferencias entre L117K y la referencia silvestre. La comparación de ambos paneles arroja que Q122 en L117K establece dos nuevas interacciones (Q122-Q160 y Q122-K117) y deja de ser capaz de llevar a cabo una que existe en la silvestre (Q122-L159). La interacción que establece Q122 con K117 abre nuevas posibilidades de comunicación indirecta entre el lóbulo superior e inferior debido a la formación de redes de interacciones no covalentes. Esto ocurre debido a que K117 tiene conexiones energéticamente significativas con diversos residuos del lóbulo inferior debido a que su cadena lateral es notablemente extendida y posee carga positiva, características que difieren por completo de la lisina presente en la referencia silvestre. En la figura 44 comparamos la identidad de los pares de interacción significativos del residuo 117 en la referencia silvestre y en L117K. En la figura 45, dibujamos simultáneamente los pares de interacción de Q122 y K117 para visualizar las redes de conectividad interna que se forman gracias a esta serie de interacciones que son únicas para L117K. En esta figura, los residuos 122 y 117 los resaltamos como esferas de color naranja y sus pares de interacción significativa como esferas amarillas. En LAO, estos residuos están conectados pero de forma indirecta a través de la S120, sin embargo la red de puentes de hidrógeno que existe entre ellos no tiene la capacidad de comunicar zonas distantes de la estructura debido al limitado potencial para interactuar que tiene la L117.

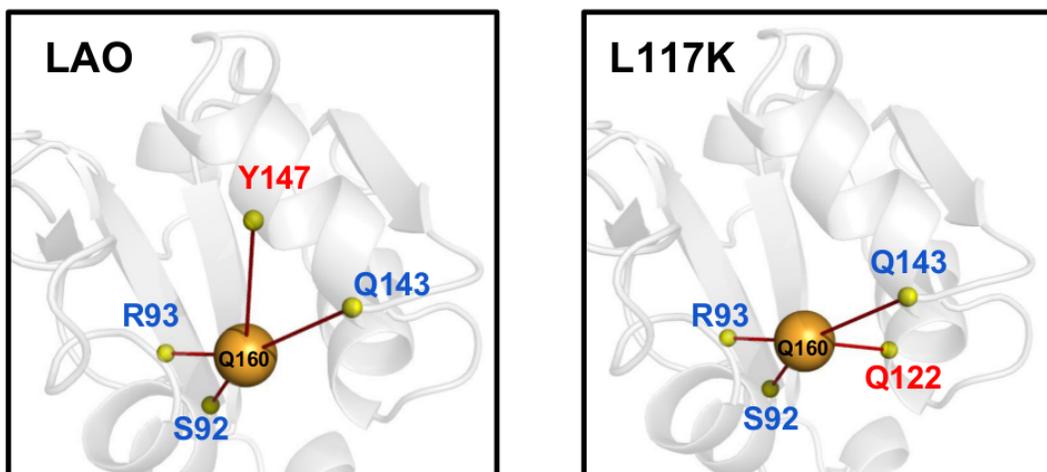


**Figura 44.** Conexiones que establecen el residuo 117 en LAO y L117K, respectivamente.



**Figura 45.** Conexiones que establecen el residuo Q122 y el residuo 117 en LAO y L117K, respectivamente.

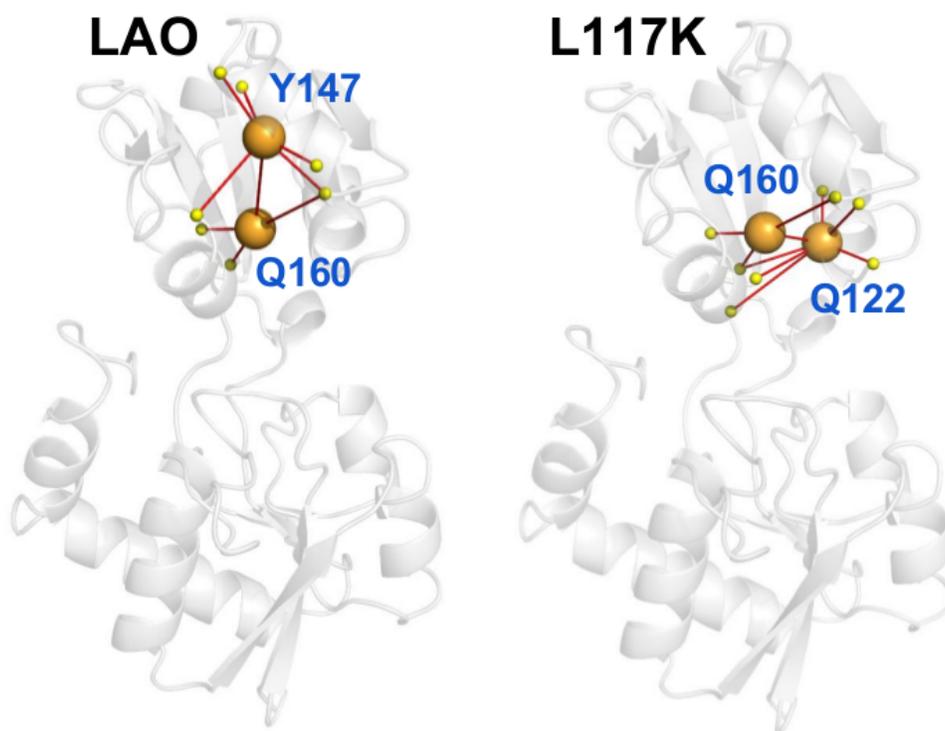
La otra interacción que Q122 establece en el cluster 1 de L117K y se encuentra ausente en la referencia silvestre es la que se da con Q160. En la figura 46 comparamos las interacciones que puede establecer el residuo 160 en la referencia silvestre y en L117K. En ambos casos, Q160 establece 4 interacciones de las cuales 3 se conservan: Con S92, R93 y Q143. La diferencia entre ambos radica en la incapacidad que tiene el residuo 160 en el cluster 1 de L117K en interactuar con Y147. En las figuras 48 y 49 expandimos nuestro foco de observación para analizar el impacto causado por las modificaciones que existen en la conectividad del residuo 160. En la figura 47 mostramos simultáneamente los pares de interacción significativos que existen, para el caso de la referencia silvestre, en Y147 y Q160 y en Q122 y Q160 para el caso del cluster 1 de L117K. Esta figura nos sirve para entender cómo la reconfiguración de las conectividades del residuo 160 causa que las redes de puentes de hidrógeno intermoleculares se propaguen por zonas distintas de la estructura: la silvestre, gracias a los pares de interacción significativos de Y147 propaga interacciones por la zona superior del lóbulo superior, mientras que en el cluster 1 de L117K, Q122 sostiene interacciones con residuos que se encuentran en la zona de la bisagra, hecho que se explica porque los residuos de estas zonas exploran distintas zonas del espacio conformacional, como se ve en los gráficos de Ramachandran de la sección correspondiente y por lo tanto, permiten que las estructuras de este grupo tengan valores de apertura menores.



**Figura 46.** Pares de interacción significativos de Q160 en LAO (izquierda) y L117K (derecha)

En la figura 48 comparamos los compañeros de interacción significativa de Y147 en la referencia silvestre y en el cluster 1 de L117K, para evaluar qué efectos tiene en la conectividad de este residuo la desaparición de su interacción con Q160. En la figura observamos que la mayoría de los pares de interacción que la Y147 tiene en L117K son conservados de la

referencia silvestre: Q143, D144, L150, T151, F169. Los compañeros de interacción nuevos que establece la Y147 en L117K son A174 y Q172. La pérdida de la interacción silvestre le da una menor restricción conformacional a Y147 y le permite explorar zonas de las proteínas más alejadas. En la figura 49 dibujamos de forma simultánea los compañeros de interacción significativos de Y147, Q172 y A174, lo que evidencia que la expansión de la red de puentes de hidrógeno intermoleculares se da hacia zonas distantes gracias a las nuevas habilidades de interacción del residuo 147, abriendo nuevas configuraciones de conectividad y por ende, nuevas posibilidades conformacionales para esta mutante. El impacto que tuvo la modificación del residuo 117 fue mayor al que esperábamos, ya que alteró la forma en que interactúan residuos distantes a esa mutación. En la siguiente sección trataremos de describir cómo se alteró la globalidad de las conexiones utilizando herramientas de teoría de redes para tener una visión más amplia de los cambios que experimentaron las estructuras del cluster 1 de L117K.

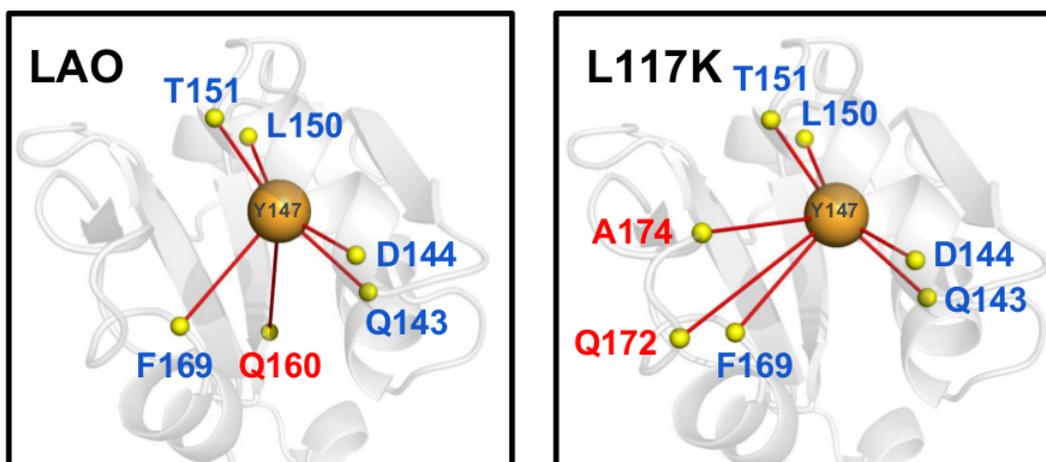


**Figura 47.** Localización en la estructura de los pares de interacción significativos de Y147 y Q160 en LAO (derecha) y Q122 y Q160 para L117K (derecha).

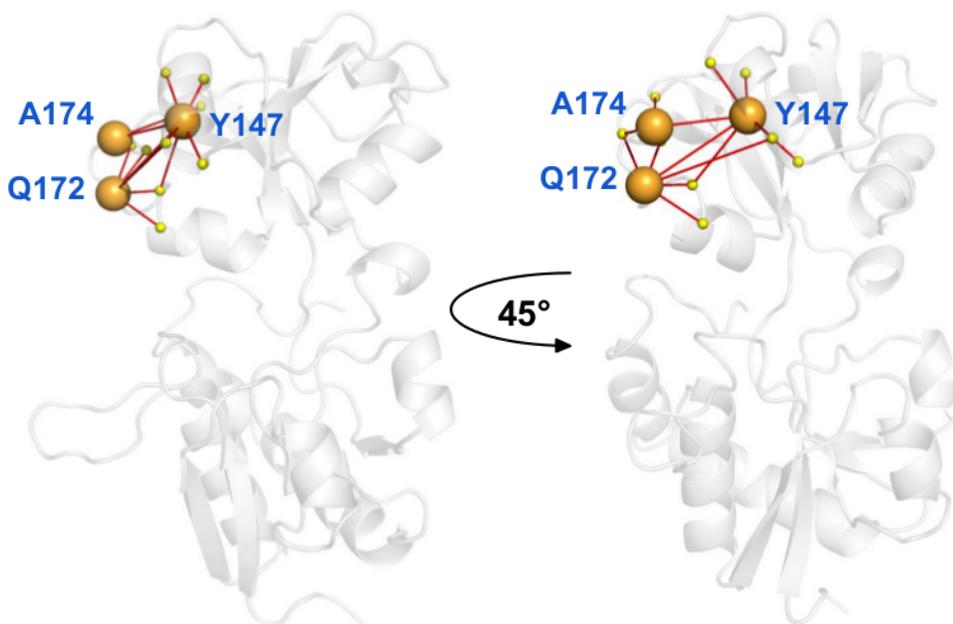
### 5.9 La topología de la red de interacción del grupo 1 de L117K difiere de la referencia

En la sección anterior encontramos diferencias notables en la conectividad interna de las estructuras que pertenecen al cluster 1 de L117K respecto a las de la referencia silvestre, sin embargo, nuestro enfoque dista de ser una descripción global ya que nos centramos en los

cambios más evidentes y cazamos las diferencias una a una. Para evaluar cambios globales, utilizamos la información proveniente de los mapas de contacto para construir matrices de adyacencia de residuos y las representamos gráficamente como grafos no dirigidos para buscar diferencias en los patrones globales usando medidas de conectividad y centralidad. En la figura 50 mostramos los grafos que generamos usando esta estrategia.



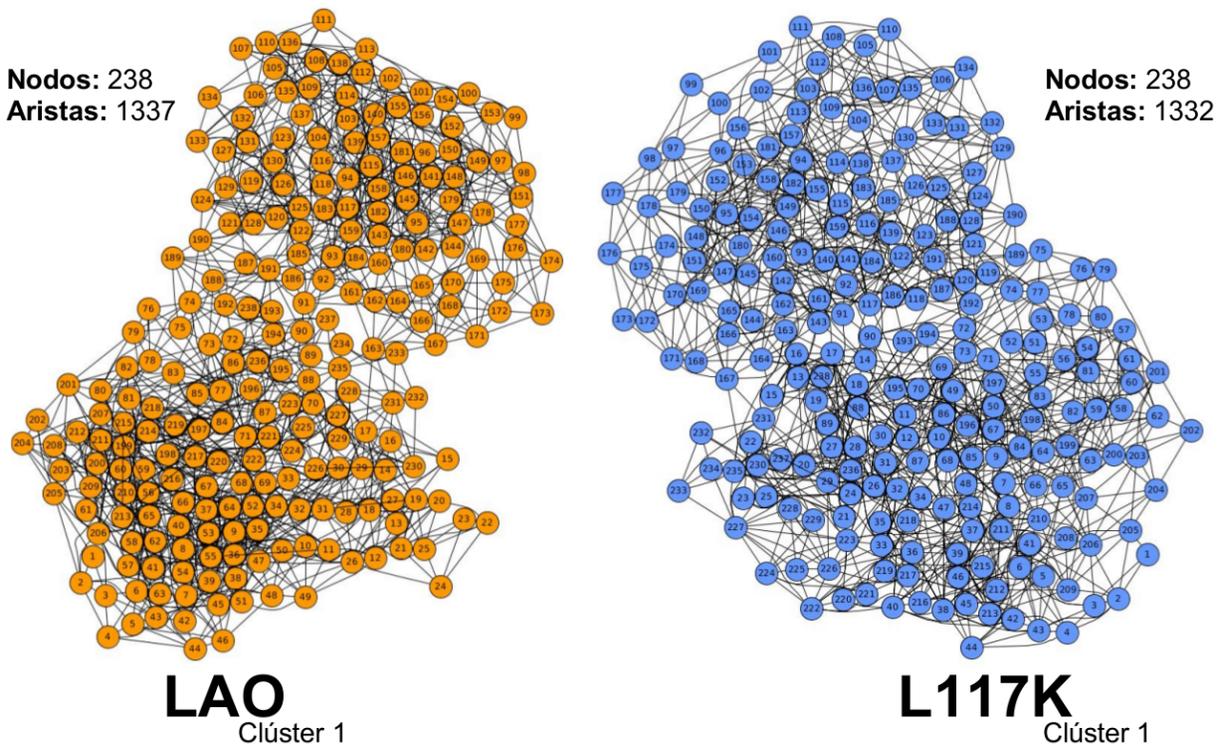
**Figura 48.** Pares de interacción significativos de Y147 en LAO (izquierda) en y L117K (derecha).



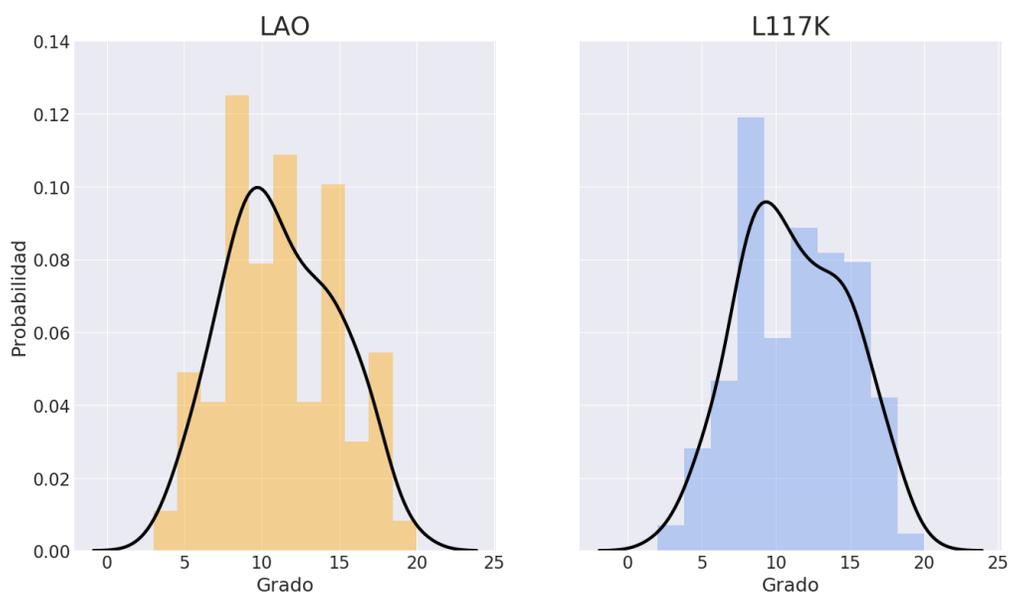
**Figura 49.** Localización de los compañeros de interacción significativos de Y147, Q172 y A174 en L117K.

En el panel izquierdo se encuentra el grafo asociado a las estructuras del cluster 1 de LAO y en el derecho el correspondiente para las estructuras del cluster 1 de L117K. Ambas

redes cuentan con el mismo número de nodos (238, que corresponde al número de residuos) y un número muy similar de conexiones entre ellos, que en este contexto, significan contactos entre los residuos. En ambas, el grado (número de conexiones) de los nodos se distribuye de forma cercana a una distribución Poisson, por lo que ambas redes corresponden a un modelo aleatorio (figura 51). Las propiedades generales de ambas redes se resumen en la tabla 42.



**Figura 50.** Comparación de redes de contactos de residuos de las estructuras del cluster 1 de LAO y del cluster 1 de L177K.



**Figura 51.** Distribución de grado de las redes de LAO y L117K

**Tabla XI.** Comparación de propiedades generales de las redes

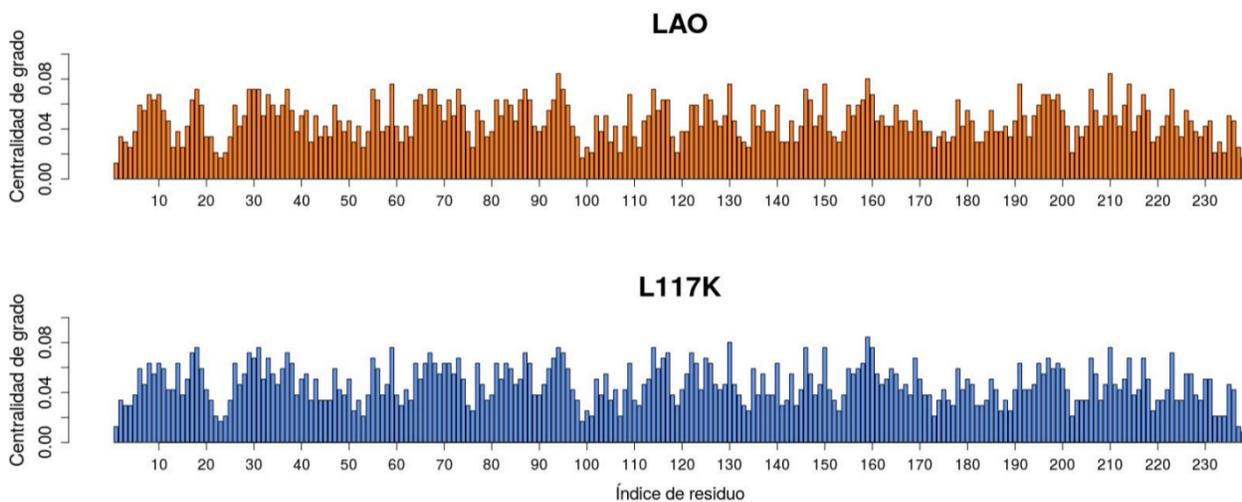
	LAO	L117K
<b>Grado promedio</b>	11.24	11.19
<b>Diámetro de la red</b>	10	9
<b>Longitud característica</b>	4.307	3.978
<b>Eficiencia global promedio</b>	0.299	0.312
<b>Eficiencia local promedio</b>	0.801	0.805
<b>Coefficiente de clustering promedio</b>	0.532	0.522

Los valores de la tabla XI muestran que ambas redes muestran diferencias muy pequeñas en su globalidad. La primera propiedad que comparamos es el grado promedio, que es el número promedio de conexiones que establecen los nodos de las redes y que en el contexto de las redes de interacción de proteínas se relaciona con la estabilidad del sistema. El valor de L117K (11.19) es ligeramente menor al de la versión silvestre de LAO (11.24). La diferencia es tan pequeña que podemos considerar que el efecto de esta mutación en la estabilidad es despreciable, como habíamos indagado por otras propiedades estructurales en secciones anteriores. Posteriormente, obtuvimos propiedades que nos muestran qué tan bien o mal se

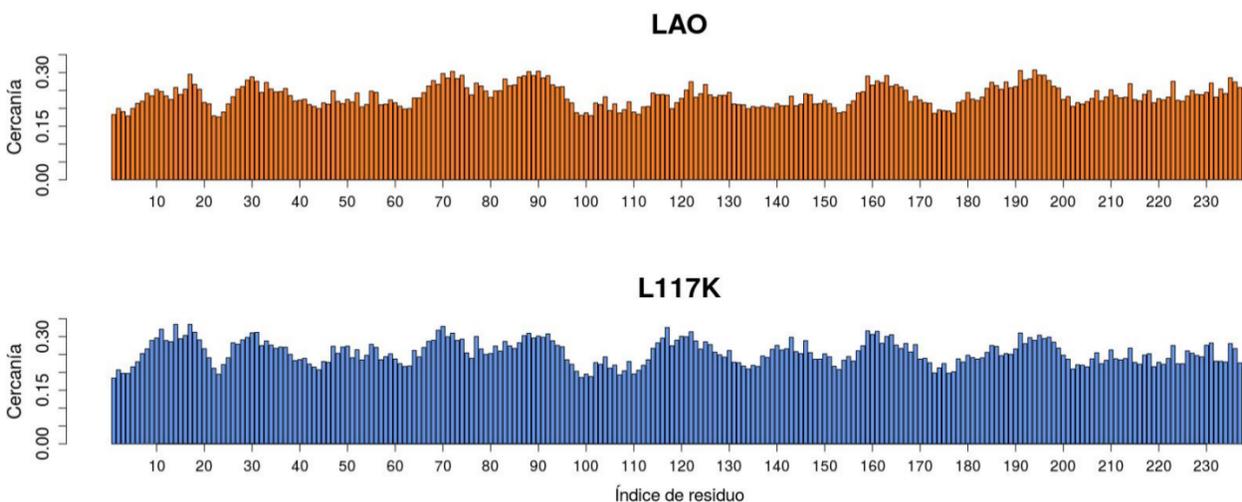
propaga la información en las redes. Estas propiedades son el diámetro, la longitud característica, la eficiencia local promedio y global promedio de las redes. El diámetro de la red, que nos dice qué tan alejados, en número de conexiones, están los dos nodos más distantes, es de 10 y 9 para LAO y L117K, respectivamente. La longitud característica, que nos dice cuál es la distancia promedio entre todas las parejas de nodos es de 4.307 para LAO y 3.978 para L117K. La eficiencia promedio es el inverso multiplicativo del promedio de las distancias entre todas las parejas de nodos en el caso de la eficiencia global promedio o bien, entre los nodos que integran los subgrafos compuestos exclusivos por los vecinos de un nodo (eficiencia local promedio). Esta propiedad se comportó de la siguiente forma: la eficiencia global de LAO fue de 0.299, la de L117K 0.312 mientras que la eficiencia local fue de 0.801 y 0.805 para LAO y L117K, respectivamente. Al igual que con el grado, las diferencias son pequeñas, pero todas indican que la red del cluster 1 de L117K difunde información a través de sus nodos ligeramente mejor que la versión silvestre, lo que en este contexto se relaciona con el establecimiento de vías de interacción adicionales dentro de la molécula propiciadas por la mutación. La última propiedad comparada es el coeficiente de clustering promedio que mide el carácter gregario de los elementos que constituyen la red. En el contexto estructural, mide la disposición a agruparse en módulos de los residuos. Los valores de esta propiedad se parecen bastante entre redes, siendo ligeramente más grande en LAO.

Con las propiedades anteriores tenemos una noción clara de que el comportamiento general de las redes es bastante parecido, no obstante, estas características globales no emergen necesariamente gracias a la participación de los mismos residuos. Para entender cuáles son los nodos cruciales para la conectividad, obtuvimos medidas de centralidad para identificar los residuos críticos de ambas proteínas. Las tres medidas que obtuvimos fueron la centralidad de grado, la cercanía (*closeness centrality*) y la intermediación (*betweenness centrality*). En el contexto de las redes de contactos de proteínas, la centralidad de grado nos sirve para identificar los residuos que establecen el mayor número de interacciones (hubs), ya que se relaciona con el número de conexiones que establece cada residuo. La cercanía y la intermediación son centralidades asociadas a las rutas cortas por lo que sus valores nos permiten identificar residuos cruciales en el establecimiento de vías de interacción intermolecular. La cercanía indica qué tan próximo está un nodo de los demás, por lo que se le relaciona con la aptitud de un residuo a participar en la transmisión de señales dentro de la estructura y se ha identificado que los residuos con altos valores de cercanía suelen formar parte de sitios activos. La intermediación cuantifica la habilidad de un nodo para mediar la comunicación entre los demás: un nodo central en este caso es aquel que en más ocasiones

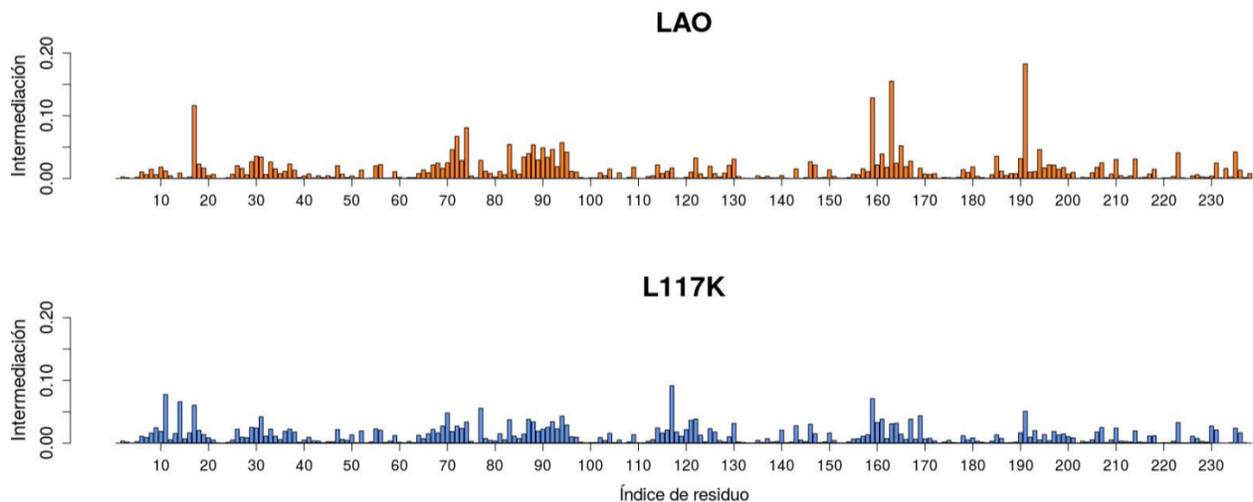
forma parte de la ruta entre parejas de residuos. Los residuos con altos valores de cercanía son potencialmente reguladores en las vías de comunicación alostéricas. En las figuras 52 a 54 mostramos los valores de cada uno de los residuos para estas diferentes centralidades. Estas representaciones nos muestran que los valores de estas centralidades distan de ser homogéneos en los nodos que constituyen ambas proteínas, así que obtuvimos los residuos críticos para cada centralidad. Consideramos críticos aquellos residuos que exhiben un valor igual o mayor que el de la media más 1.8 veces la desviación estándar. En las tablas 43 a 45 los enlistamos y en las figuras 55 a 57 dibujamos diagramas de Venn para conocer qué tanto se superponen los conjuntos de nodos críticos en ambas redes.



**Figura 52.** Centralidad de grado para LAO (panel superior) y L117K (panel inferior)

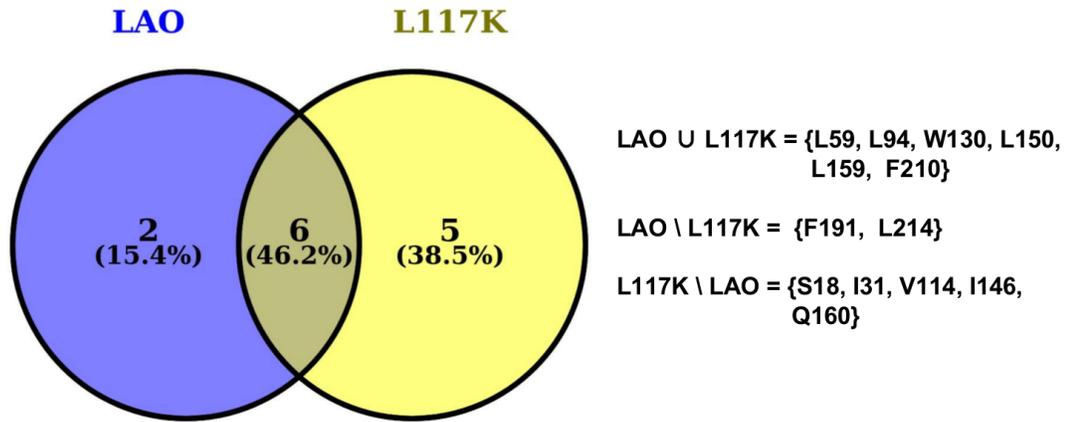


**Figura 53.** Cercanía para LAO (panel superior) y L117K (panel inferior)

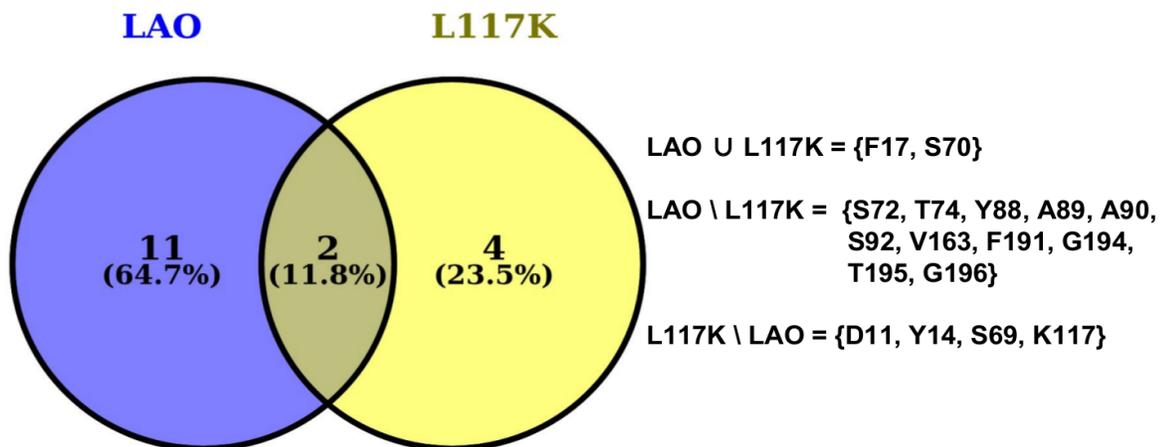


**Figura 54.** Intermediación para LAO (panel superior) y L117K (panel inferior)

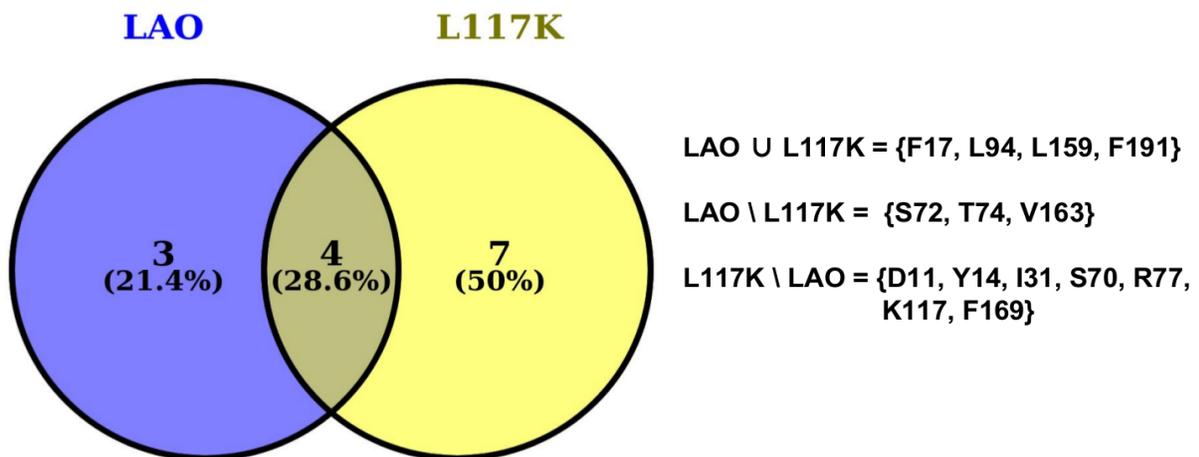
En todas las centralidades que estudiamos existe un grado de solapamiento en los conjuntos de residuos críticos de LAO y L117K. El valor más alto del solapamiento corresponde a la centralidad de grado, lo cual consideramos congruente con el carácter de la criticalidad bajo la óptica de esta centralidad. Los nodos con valores altos de esta medida son *hubs* estructurales y su identidad está determinada en gran medida por la topología de las proteínas, la cual fue indiferente a la mutación. Los conjuntos de residuos críticos en cercanía e intermediación presentaron valores más pequeños de solapamiento entre proteínas. Esto se relaciona con el establecimiento de nuevas vías de comunicación dentro de la mutante. La inserción del nuevo residuo no sólo abre nuevas interacciones sino que cierra interacciones que naturalmente existen en LAO ya que el nuevo jugador modifica la disponibilidad para interactuar de residuos que normalmente ignoran al residuo 117, detonando modificaciones en las conexiones que establecen residuos distantes de forma análoga a la que ocurre en las ciudades cuando el cierre de una avenida importante se muestra capaz de desquiciar el tránsito en calles lejanas.



**Figura 55.** Diagrama de Venn de los residuos de alta centralidad de grado. En el costado derecho se encuentran desglosados

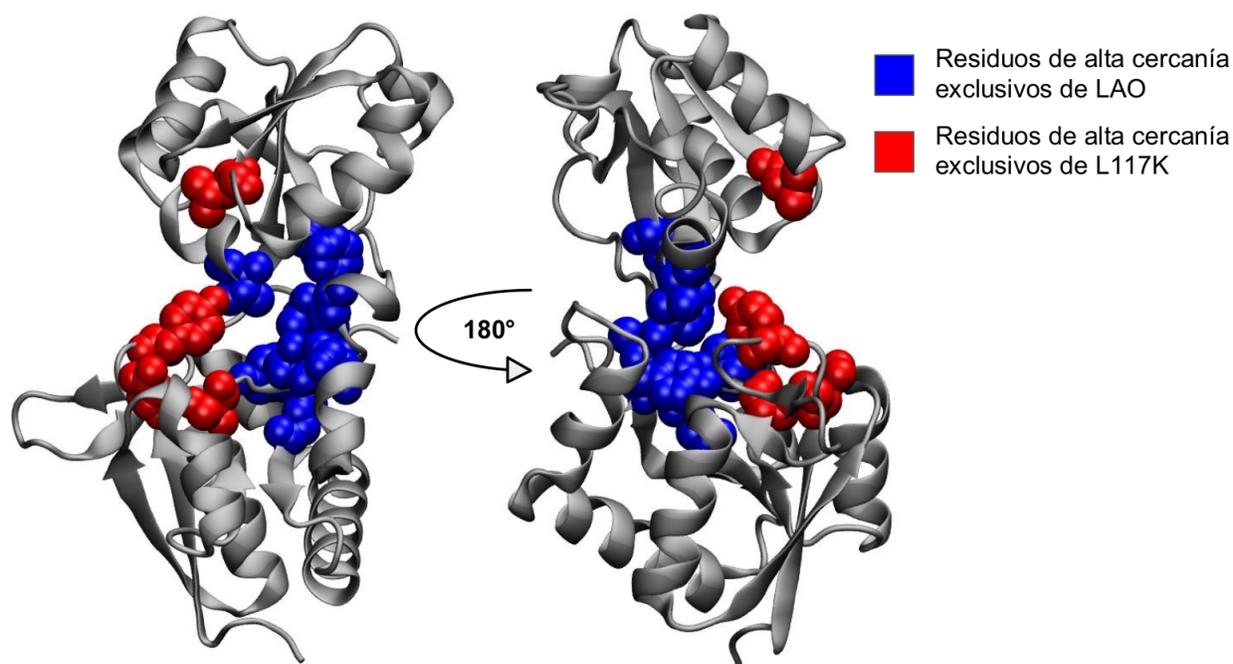


**Figura 56.** Diagrama de Venn de los residuos de alta cercanía. En el costado derecho se encuentran desglosados



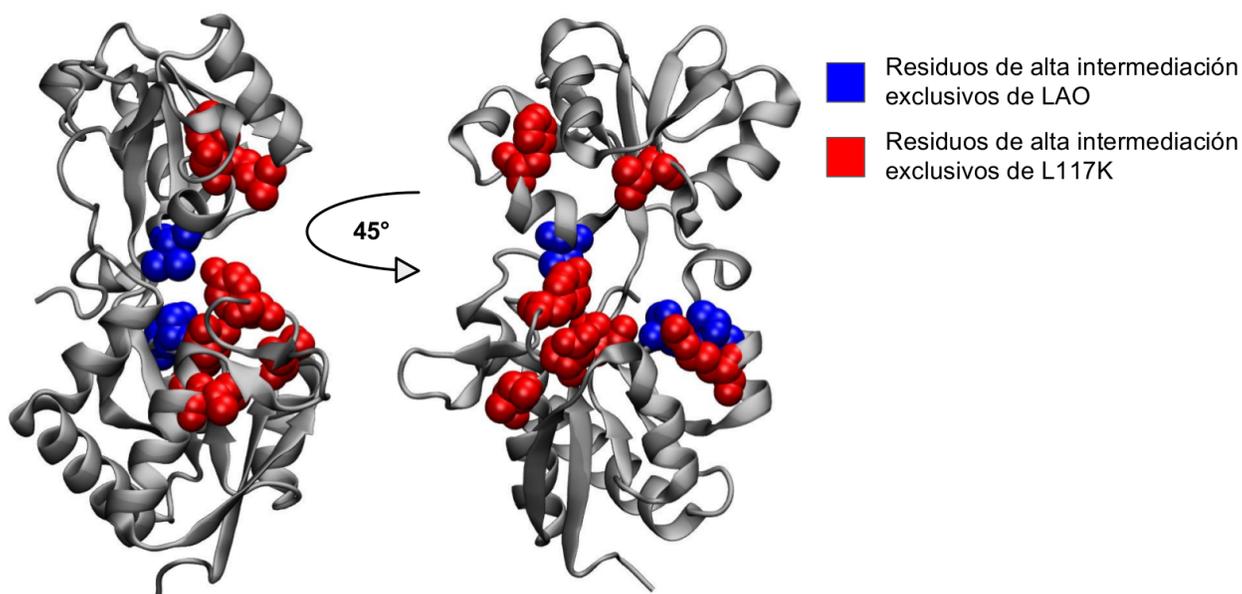
**Figura 57.** Diagrama de Venn de los residuos de alta intermediación. En el costado derecho se encuentran desglosados

Al localizar los residuos que son críticos en cercanía para LAO encontramos que la mayoría de ellos se concentran en la bisagra mientras que los de L117K no están localizados en una zona específica pero en su mayoría son residuos que entre ellos interactúan, como es el caso de D11, Y14 y K117, cuyas relaciones describimos en secciones anteriores (figura 58). Para las PBP, la cercanía parece informarnos cuáles son los residuos que son cruciales para que la conformación predominante se sostenga durante la dinámica. En el caso del cluster 1 de LAO, filtramos estructuras provenientes de un estado abierto y por lo tanto las interacciones que establecen los residuos de la bisagra son las cruciales para limitar su movilidad y por ende evitar el cierre espontáneo. Por otro lado, las estructuras que provienen del cluster 1 de L117K se cierran espontáneamente y los residuos cruciales para que esa conformación siga existiendo son K117 son los residuos con los que interactúa electrostáticamente. Entendemos que la corroboración de esta suposición requiere de indagaciones adicionales que no serán exploradas en este trabajo, no obstante, este resultado es congruente con otros resultados de este trabajo que sugieren que la mutación modificó el espacio conformacional visitado por LAO mediante el establecimiento de nuevas interacciones, sumadas a la desaparición de interacciones importantes que existen en la silvestre. Consideramos que esta centralidad debe ser estudiada en trabajos próximos con PBPs adicionales para poder verificar si esta distribución estructural de los residuos con valores altos de cercanía es una propiedad general.



**Figura 58.** Localización los residuos críticos por cercanía en LAO (azul) y en L117K (rojo). La figura del panel derecho es una rotación de la del izquierdo

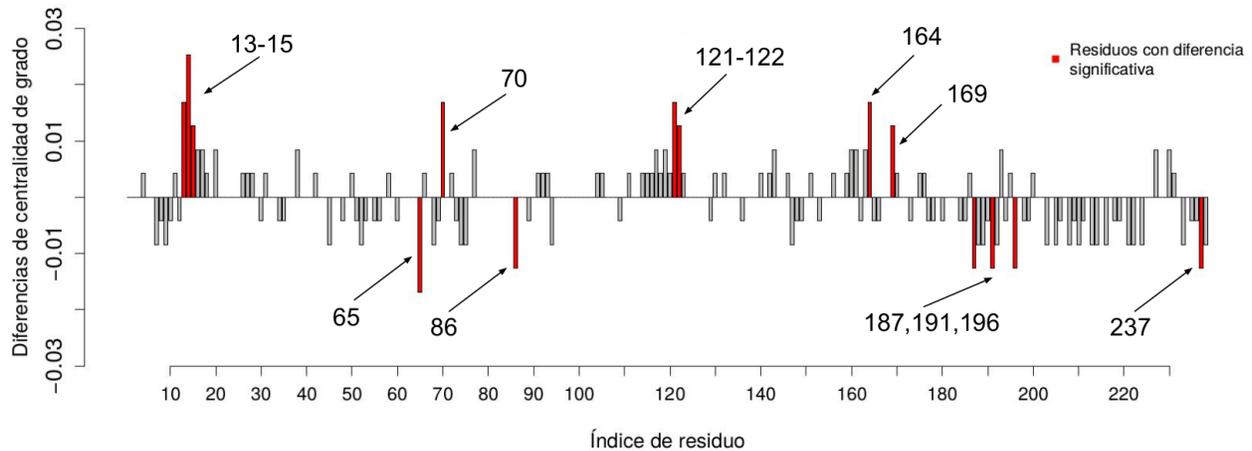
A diferencia de la centralidad anterior, los residuos altos en intermediación no se localizan en ninguna zona específica para ninguna de las dos proteínas, sino que más bien se encuentran dispersos en diferentes partes de la estructura (figura 59). Algunos residuos que fueron identificados como críticos por su intermediación también son críticos por su cercanía, hecho que consideramos se relaciona con la asociación de ambas medidas de centralidad a la comunicación inter-residuo y con la identificación de vías de señalización alostéricas. Los residuos que son exclusivamente críticos por su intermediación están en posiciones contiguas a los críticos por cercanía, lo que sugiere que el rol que desempeñan estos residuos es de soporte y regulación del establecimiento de las interacciones que están asociadas a la conformación predominante en cada una de las proteínas.



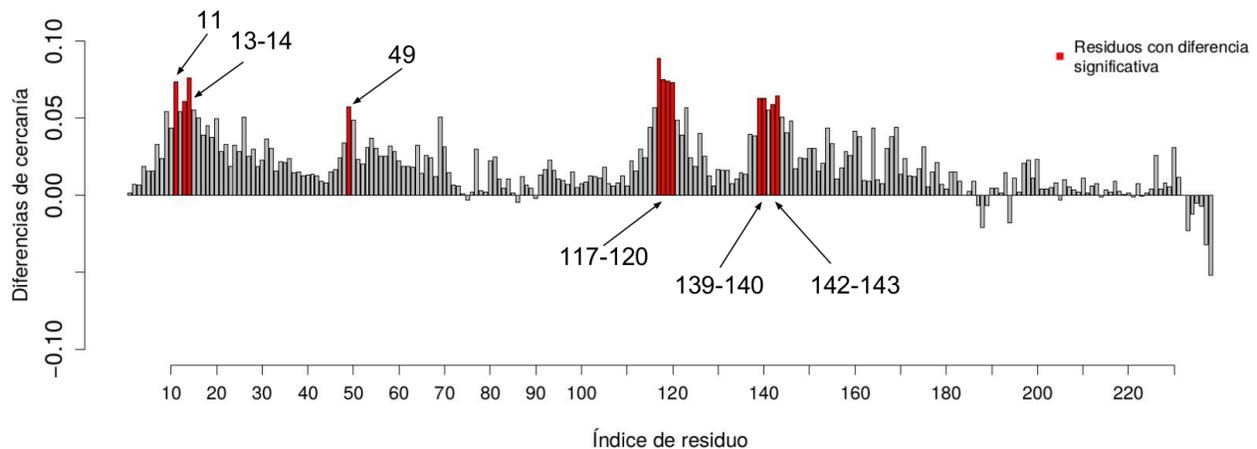
**Figura 59.** Localización de los residuos críticos por intermediación para LAO (azul) y L117K (rojo). La figura del panel derecho es una rotación de la del izquierdo

Además de los hubs, identificamos los residuos que más cambiaron respecto a la referencia silvestre en los valores de estas centralidades. En las figuras 60 a 62 mostramos las diferencias por residuo de cada centralidad. Las diferencias se calcularon restándole el valor de L117K al de LAO, por lo que los residuos cuyo valor de diferencia es positivo son aquellos donde hubo un incremento en la centralidad respecto a la silvestre, mientras que en los que hay un valor de diferencia negativa son aquellos en los que existe una disminución en el valor de centralidad respecto al de referencia. En color rojo marcamos los residuos que tienen una diferencia dos veces mayor a la desviación estándar de la variación y que denominamos como nodos de alta

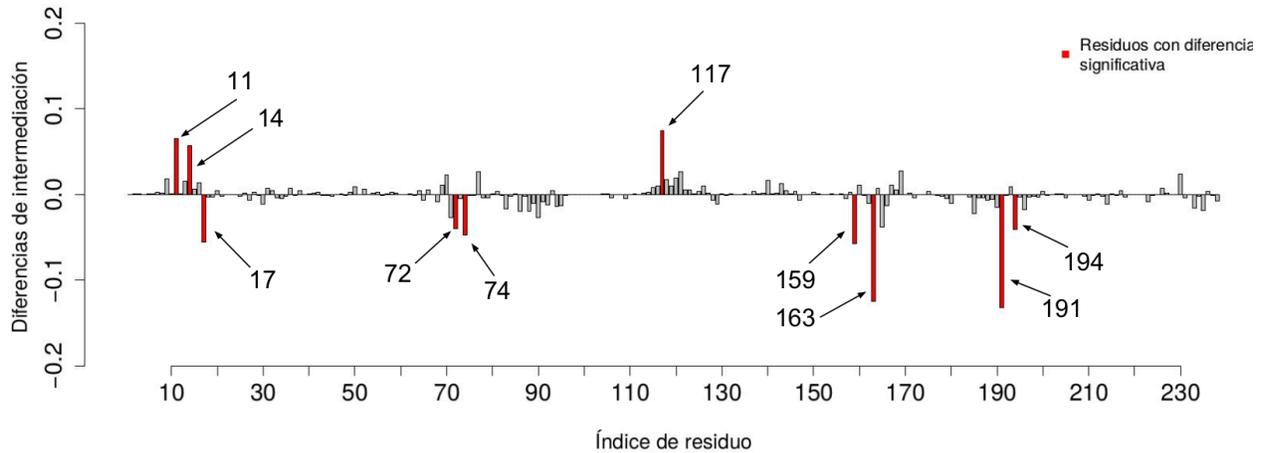
variación. En las descripciones que vienen a continuación, nos centramos en estos nodos de alta variación para explicar diferencias de conectividad que existen entre las redes de la silvestre y de L117K.



**Figura 60.** Diferencias en la centralidad de grado. La diferencia fue calculada restando los valores de LAO a los de L117K. En rojo se resaltan los residuos que presentan una diferencia mayor que el promedio.



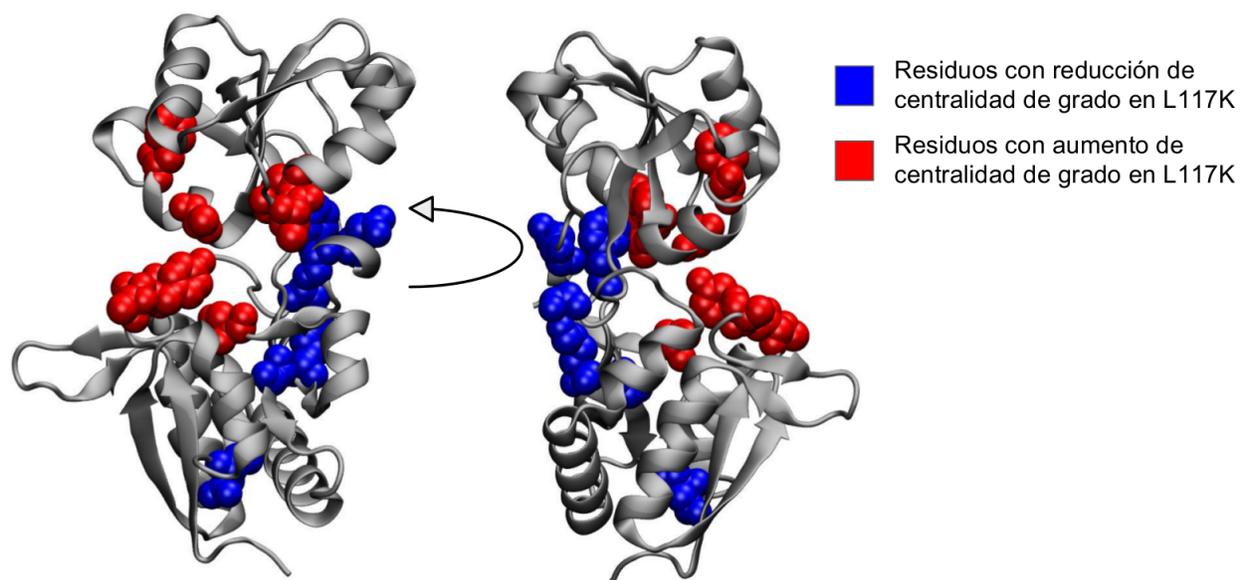
**Figura 61.** Diferencias en la cercanía. La diferencia fue calculada restando los valores de LAO a los de L117K. En rojo se resaltan los residuos que presentan una diferencia mayor que el promedio.



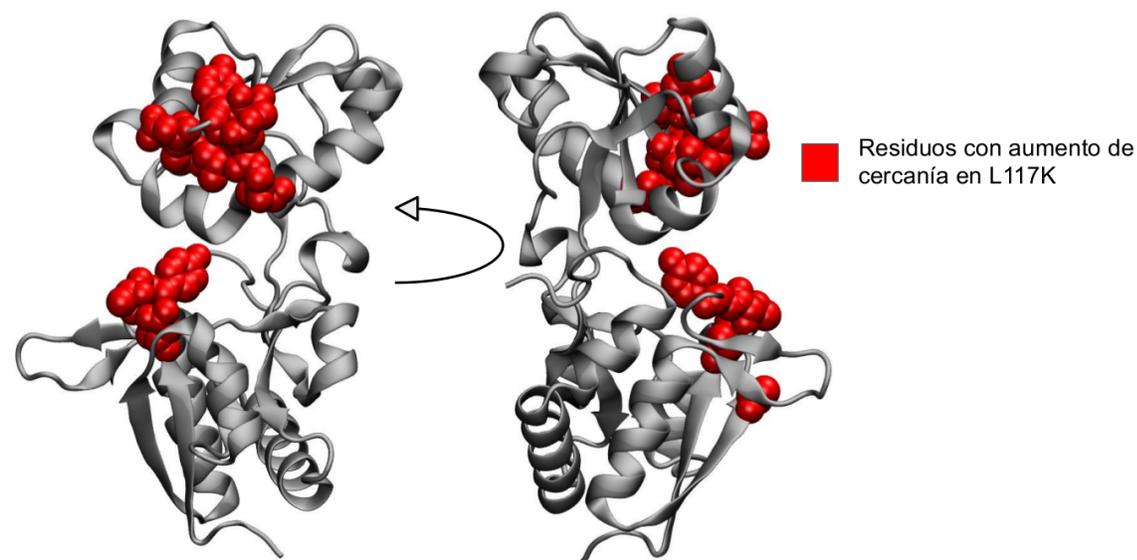
**Figura 62.** Diferencias en la intermediación. La diferencia fue calculada restando los valores de LAO a los de L117K. En rojo se resaltan los residuos que presentan una diferencia mayor que el promedio

Al resaltar sobre la estructura de LAO los residuos que presentan diferencias significativas para las distintas centralidades (figura 63 a 65) se presenta un patrón en el que los residuos cuyos valores de centralidad decrecieron respecto a la referencia silvestre (marcados en color azul) se ubican principalmente en la bisagra mientras que aquellos que aumentaron (resaltados en color rojo) se localizan en zonas expuestas o bien en residuos vinculados con la formación de la interfaz de unión a ligandos. Además, existen residuos que incrementaron su centralidad respecto a la referencia que también entran en la categoría de nodos de alta conectividad, como lo son D11, Y14 (nodos que son críticos para las tres centralidades) y K117 (crítico para intermediación). El hecho de que los residuos que presentan mayor variabilidad hacia arriba en L117K coincidan con aquellos que son críticos por sus valores de centralidad, muestra que los cambios en las conexiones internas que experimenta esta mutante son para darle cabida a las nuevas posibilidades de interacción que se abren gracias a la introducción de K117, un residuo que poco tiene que ver en términos químicos y estéricos con su contraparte silvestre. Debido a la emergencia de estos nuevos patrones de conectividad, donde los residuos de la bisagra son menos relevantes, las restricciones de apertura y torsión se relajaron, ampliando la colección de conformaciones que se pueden visitar en ausencia del ligando. Aunque desconocemos el rol particular que juegan en la adquisición de nuevas capacidades de interacción cada una de las conformaciones asociadas a esta heterogeneidad conformacional, cada vez queda más claro que la modificación del ensamble estructural es una herramienta de innovación funcional en proteínas por lo cual es crucial evaluar el impacto que tienen en la dinámica las mutaciones que introdujimos. Nuestro trabajo se suma de forma tangencial a las evidencias que se

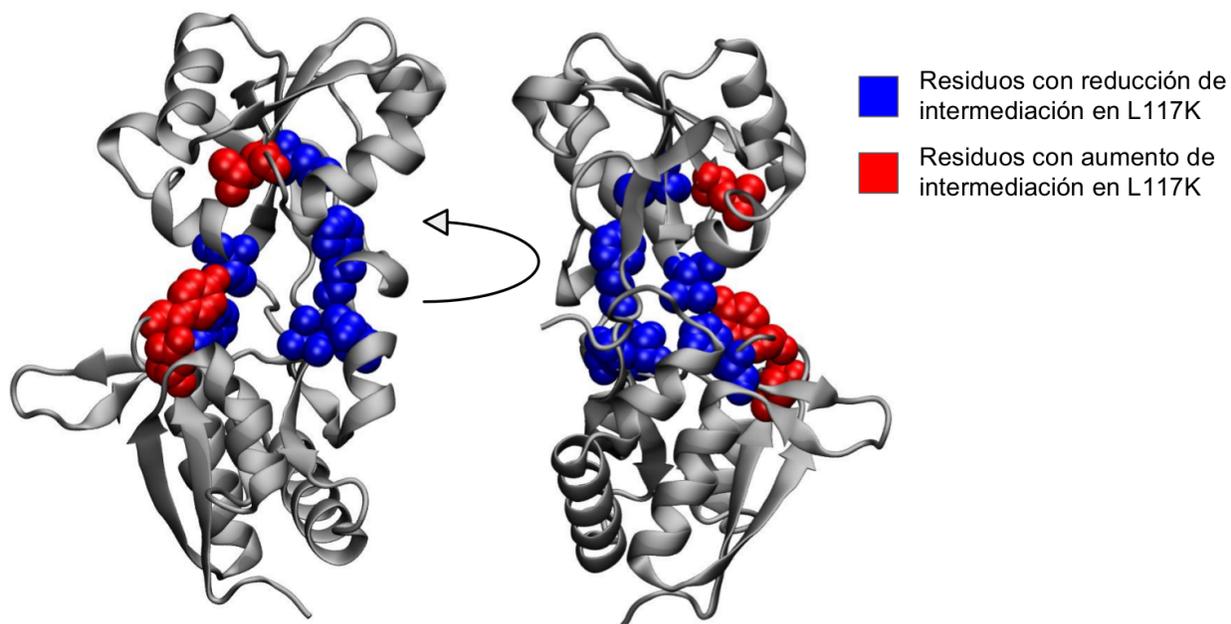
acumulan al respecto y que dejan claro que sin la consideración de ese aspecto carecemos de una imagen completa que nos permita entender el éxito o el fracaso de nuestros algoritmos de rediseño. Hay que mantener los ojos abiertos cuando evaluamos los resultados de modificar la secuencia de nuestras proteínas para conferirles nuevas habilidades sin otra pista para ello que nuestra temeridad, a la que muchas veces le decimos intuición química.



**Figura 63.** Residuos con cambios significativos en la centralidad de grado. En azul se muestran los que tienen una reducción y en rojo los que presentan un aumento respecto a la referencia silvestre.



**Figura 64.** Residuos con cambios significativos en la cercanía. En rojo se muestran aquellos residuos que tienen un aumento en el valor de esta centralidad respecto a la referencia silvestre.



**Figura 65.** Residuos con cambios significativos en la cercanía. En azul se muestran los que tienen una reducción y en rojo los que presentan un aumento en el valor de esta centralidad respecto a la referencia silvestre.

# 6 Conclusiones

Por fines de claridad, desglosamos en distintos apartados las conclusiones a las cuales llegamos al finalizar este trabajo.

## 6.1 Evaluación del impacto en la estructura de las distintas mutaciones

- Los parámetros de viabilidad estructural que usamos, concretamente los histogramas del número de puentes de hidrógeno de cadena principal y las distribuciones de residuos en distintas conformaciones de estructura secundaria, nos muestran que las mutaciones que introdujimos modifican muy poco la topología de las proteínas simuladas, por lo que consideramos que las diferencias dinámicas entre mutantes no responden a eventos de pérdida de estructura o desplegamiento, sino al muestreo conformacional característico de cada una de ellas.
- A nivel de cadena principal, los cambios son imperceptibles y presentan un grado muy parecido entre todas las mutantes, si comparamos usando el porcentaje de intersección de los histogramas de las mutantes con la referencia silvestre.
- A nivel de estructura secundaria observamos mayores diferencias respecto a la referencia silvestre y mayor variabilidad entre mutantes.
- L117Q es la mutante cuyas distribuciones se alteran menos respecto a la referencia.
- L117K es la que experimenta las mayores diferencias. Al diseccionar por lóbulos, vemos que las distribuciones de esta mutante difieren más en el lóbulo superior donde se localiza la mutación y tienen un impacto notorio en el porcentaje de residuos en conformación beta. No obstante, el histograma de residuos sin asignación tiene un porcentaje de intersección del 85% con el de la silvestre, por lo que las diferencias de las distribuciones se relacionan con reconfiguraciones y no con una pérdida de estructura significativa.

## 6.2 Impacto de las mutaciones puntuales en la dinámica

- Todas las mutantes que simulamos presentan cambios notables en sus huellas dinámicas al ser comparadas con la versión silvestre de LAO.
- Las proteínas simuladas presentan una reducción de sus valores medianos de ángulo de apertura, lo que señala que el residuo nativo en la posición 117 (L117) es crucial para que LAO permanezca en los valores característicos de esta propiedad en ausencia de ligandos.
- L117K experimenta la mayor reducción de apertura, al grado de que puede visitar estados cerrados y parcialmente cerrados en las condiciones de simulación.
- La reducción de la apertura en L117K también viene acompañada de diferencias en el valor mediano de la torsión, el cual no se modifica para las otras mutantes, lo que significa que para visitar estos nuevos estados cerrados, los lóbulos tienen que cambiar su orientación relativa entre ellos.

## 6.2 Localización de residuos con movilidad diferencial

- La comparación de las gráficas de Ramachandran de los residuos de las mutantes respecto a los de la silvestre mediante la utilización de la prueba bidimensional de Kolmogorov-Smirnov nos permitió identificar los residuos que presentaron diferencias de movilidad a nivel de cadena principal de forma automática y ágil.
- L117K es la mutante que tiene residuos con mayor movilidad diferencial y la mayoría de ellos se localizan en la bisagra, lo cual se relaciona de forma directa con la visita de los estados cerrados y parcialmente cerrados que mencionamos en el apartado anterior.
- La cantidad de residuos de movilidad diferencial para las otras dos mutantes es menor (L117R tiene 3 y L117Q tiene 6) bajo el mismo criterio de selección, lo que quiere decir que las diferencias de movilidad a nivel de cadena principal son menores en ellas. La localización de éstos no exhibe un patrón claro que nos permita adjudicárselo a alguna propiedad dinámica.

## 6.3 Agrupamiento y conectividades de conformaciones relevantes

- El algoritmo HDBSCAN nos permitió agrupar conformaciones correspondientes a las zonas de alta densidad de los histogramas bidimensionales de apertura y torsión mediante la definición de solo un parámetro, el tamaño mínimo cluster.
- A diferencia de la mayoría de los algoritmos utilizados para la generación de grupos estructurales, HDBSCAN determinó de forma automática el número de clusters

significativos para cada simulación y nos permitió separar las conformaciones relevantes de las que fueron consideradas ruido.

- Respecto a los grupos estructurales identificados, el grupo 1 de LAO es el más grande de todos, ya que contiene 3444 de las 5000 estructuras totales y está compuesto por las conformaciones que constituyen el estado abierto arquetípico de LAO.
- Otro grupo relevante identificado en el clustering es el grupo 1 de L117K que presenta valores de apertura notoriamente menores a los grupos estructurales. Al comparar las conexiones que la K117 en este cluster nos damos cuenta que la mayoría de ellas son interacciones únicas y se dan con residuos del lóbulo inferior.
- Aunque la energía electrostática es una fuerza impulsora de las interacciones que establece K117 en las estructuras de ese grupo estructural, la estérica de la cadena lateral de este residuo también tiene un rol importante, ya que cuando simulamos con una lisina neutral no se restableció la firma dinámica de la silvestre porque no se formaron las interacciones características asociadas a ese grupo estructural.
- La ausencia de la interacción del residuo 117 con S120, que está presente en la versión silvestre de LAO y las otras mutantes pero no en el clúster 1 de L117K, permite que el residuo 117 tenga menores restricciones conformacionales, lo que le permite interactuar con residuos del lóbulo inferior.

### **6.5 Impactos en la conectividad interna**

- La representación de las interacciones usando mapas de contactos nos permitió identificar que las diferencias en las conexiones internas que tienen las estructuras del grupo 1 de L117K se propagan por zonas distantes a la mutación.
- Encontramos que Q122 es el único residuo de alta conectividad que es exclusivo para el grupo 1 de L117K y al comparar sus contactos con los que establece en la versión silvestre ubicamos que pierde una interacción, Q122-L159, y gana dos: Q122-K117 y Q122-Q160.
- Las dos interacciones que se ganan expanden las vías de comunicación indirectas que Q122 puede establecer ya K117 es un residuo que se comunica con zonas distantes de la proteína.
- Construimos redes de interacción de residuos usando como matriz de adyacencia los mapas de contactos y al medir las propiedades globales de la red encontramos que tanto la versión silvestre de LAO como las estructuras del cluster 1 de L117K presentan un número de conectividades similar, una arquitectura que corresponde con redes

aleatorias y un valor de grado similar, por lo que bajo esta perspectiva son sistemas de una estabilidad comparable.

- Las propiedades generales asociadas con la conectividad sugieren que el cluster 1 de L117K corresponde a la de un sistema que propaga ligeramente de mejor manera la información que la referencia silvestre.
- Los conjuntos de residuos nodos con alta centralidad de grado comparten una alta cantidad de elementos porque son estructuras que se constituyen como hubs estructurales.
- Los conjuntos de residuos con alta cercanía varían entre LAO y L117K y nos indicaron aquellos residuos que son críticos para el establecimiento de la conformación deseada. Para LAO éstos se concentran en la bisagra mientras que para L117K son aquellos que forman parte de interacciones que describimos como importantes para el cierre.
- Los residuos altos en cercanía suelen preceder a los residuos que tienen valores altos en intermediación, por lo que su rol es de soporte a los anteriores.
- Los residuos que disminuyeron sus valores de centralidad en la red de L117K respecto a la referencia silvestre se localizan en la bisagra, lo que indica que estos residuos redujeron su capacidad de establecer interacciones en la estructura.
- Los residuos que aumentaron sus valores de centralidad en la red de L117K respecto a la referencia silvestre, se localizan en zonas expuestas o la interfaz y también están asociados en interacciones que vinculamos al cierre espontáneo.

# 7 Perspectivas

Como suele ocurrir, al finalizar el trabajo nos quedamos con más preguntas que respuestas. Pensamos que nuestros resultados son un punto de partida para profundizar sobre preguntas asociadas con la diversidad conformacional y la adquisición de nuevas funciones en LAO. En primer lugar, proponemos los siguientes experimentos para indagar más en la pertinencia de nuestros resultados:

- Cristalizar la conformación abierta para conocer cual es la conformación predominante en ausencia de ligandos en L117K
- Estudiar de manera experimental la dinámica de L117K haciendo experimentos de resonancia magnética nuclear y comparar las conformaciones encontradas de esta manera con las que nosotros ubicamos en las simulaciones.

Y además, proponemos estos experimentos que toman nuestros datos como punto de partida y permiten explorar nuevos aspectos de la manera en la cual funciona L117K y LAO-BP:

- Caracterizar con simulaciones adicionales el cluster 1 de L117K. Ya que este cluster es poco poblado y todas las estructuras que lo componen provienen de la misma repetición, por lo que es importante lanzar simulaciones adicionales tomando estructuras de este grupo como conformaciones iniciales para incrementar la precisión de la descripción de este estado y conocer si una vez alcanzado es un estado estable o no.
- Incrementar el muestreo de las distintas mutantes utilizando dinámicas moleculares extendidas para tener una caracterización completa del paisaje conformacional de las proteínas simuladas así como paisajes detallados de energía libre para encontrar cómo las mutaciones puntuales afectan su composición.
- Ya que L117K es un residuo de alta intermediación, podríamos generar mutantes adicionales cambiando residuos que también sean nodos claves en esta propiedad y

evaluar el impacto que estos cambios tienen en la diversidad conformacional y capacidades de unión de LAO, utilizando esta información como estrategia de diseño.

- Utilizar metodologías de muestreo de rutas (*path sampling*) para describir cómo ocurre la unión de L117K con glutamina y determinar qué rol juegan en este nuevo proceso las conformaciones cerradas y parcialmente cerradas que ubicamos en nuestro trabajo.

## Referencias

1. Berg, J. M., Tymoczko, J. L., & Gatto Jr, G. J. (2002). Stryer: Biochemistry. *WH Freeman and Company*, 5, 306-307.
2. Sela, M., White Jr, F. H., & Anfinsen, C. B. (1957). Reductive cleavage of disulfide bridges in ribonuclease. *Science*, 125(3250), 691-692.
3. Frauenfelder, H., Sligar, S. G., & Wolynes, P. G. (1991). The energy landscapes and motions of proteins. *Science*, 254(5038), 1598-1603.
4. Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities of proteins. *Nature*, 450(7172), 964.
5. Gruebele, M. (2014). Protein dynamics in simulation and experiment.
6. Alder, B. J., & Wainwright, T. E. (1959). Studies in molecular dynamics. I. General method. *The Journal of Chemical Physics*, 31(2), 459-466.
7. McCammon, J. A., Gelin, B. R., & Karplus, M. (1977). Dynamics of folded proteins. *Nature*, 267(5612), 585.
8. Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., ... & Wrighers, W. (2010). Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002), 341-346.
9. Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H., & Shaw, D. E. (2012). Biomolecular simulation: a computational microscope for molecular biology. *Annual review of biophysics*, 41, 429-452.
10. Leach, A. R., & Leach, A. R. (2001). *Molecular modelling: principles and applications*. Pearson education.
11. Dwyer, M. A., & Hellinga, H. W. (2004). Periplasmic binding proteins: a versatile superfamily for protein engineering. *Current opinion in structural biology*, 14(4), 495-504.
12. Oh, B. H., Kang, C. H., De Bondt, H., Kim, S. H., Nikaido, K., Joshi, A. K., & Ames, G. F. (1994). The bacterial periplasmic histidine-binding protein. structure/function analysis of the ligand-binding site and comparison with related proteins. *Journal of Biological Chemistry*, 269(6), 4135-4143.
13. Silva, D. A., Domínguez-Ramírez, L., Rojo-Domínguez, A., & Sosa-Peinado, A. (2011). Conformational dynamics of L-lysine, L-arginine, L-ornithine binding protein reveals ligand-dependent plasticity. *Proteins: Structure, Function, and Bioinformatics*, 79(7), 2097-2108.
14. Silva, D. A., Bowman, G. R., Sosa-Peinado, A., & Huang, X. (2011). A role for both conformational selection and induced fit in ligand binding by the LAO protein. *PLoS computational biology*, 7(5), e1002054.
15. Banda-Vázquez, J., Shanmugaratnam, S., Rodríguez-Sotres, R., Torres-Larios, A., Höcker, B., & Sosa-Peinado, A. (2018). Redesign of LAOBP to bind novel l-amino acid ligands. *Protein Science*, 27(5), 957-968.
16. Biel, J. T., Thompson, M. C., Cunningham, C. N., Corn, J. E., & Fraser, J. S. (2017). Flexibility and design: conformational heterogeneity along the evolutionary trajectory of a redesigned ubiquitin. *Structure*, 25(5), 739-749.
17. Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, 9(1), 40.
18. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham III, T. E., DeBolt, S., ... & Kollman, P. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy

- calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, 91(1-3), 1-41.
19. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2), 926-935.
  20. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., & Shaw, D. E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8), 1950-1958.
  21. Martín-García, F., Papaleo, E., Gomez-Puertas, P., Boomsma, W., & Lindorff-Larsen, K. (2015). Comparing molecular dynamics force fields in the essential subspace. *PLoS One*, 10(3), e0121114.
  22. Miyamoto, S., & Kollman, P. A. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry*, 13(8), 952-962.
  23. Roe, D. R., & Cheatham III, T. E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation*, 9(7), 3084-3095.
  24. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry*, 32(10), 2319-2327.
  25. DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, 40, 82-92.
  26. Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1), 33-38.
  27. Oliveros, J.C. (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>
  28. Stokely, M., & Hesterberg, T. (2013). Package 'HistogramTools'.
  29. Heinig, M., & Frishman, D. (2004). STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic acids research*, 32(suppl\_2), W500-W502.
  30. Pastore, M. (2018). Overlapping: a R package for Estimating Overlapping in Empirical Distributions. *Journal of Open Source Software*, 3(32), 1023.
  31. Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1), 79-86.
  32. Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D., Li, S., & Li, M. S. (2015). Package 'FNN'. Accessed June, 1.
  33. Ripley, B. D. (2002). Modern applied statistics with S. *Statistics and Computing*, fourth ed. Springer, New York.
  34. Silva, D. A., Domínguez-Ramírez, L., Rojo-Domínguez, A., & Sosa-Peinado, A. (2011). Conformational dynamics of L-lysine, L-arginine, L-ornithine binding protein reveals ligand-dependent plasticity. *Proteins: Structure, Function, and Bioinformatics*, 79(7), 2097-2108.

35. Campello, R. J., Moulavi, D., & Sander, J. (2013, April). Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 160-172). Springer, Berlin, Heidelberg.
36. Hdbscan.readthedocs.io. (2019). *The hdbscan clustering Library — hdbscan 0.8.1 documentation*. [online] Available at: <https://hdbscan.readthedocs.io/en/latest/index.html> [Accessed 31 Jan. 2019]
37. Pietal, M. J., Bujnicki, J. M., & Kozłowski, L. P. (2015). GDFuzz3D: a method for protein 3D structure reconstruction from contact maps, based on a non-Euclidean distance function. *Bioinformatics*, *31*(21), 3499-3505.
38. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry*, *32*(10), 2319-2327.
39. Mercadante, D., Gräter, F., & Daday, C. (2018). CONAN: A Tool to Decode Dynamical Information from Molecular Interaction Maps. *Biophysical journal*, *114*(6), 1267-1273.
40. Di Paola, L., De Ruvo, M., Paci, P., Santoni, D., & Giuliani, A. (2012). Protein contact networks: an emerging paradigm in chemistry. *Chemical Reviews*, *113*(3), 1598-1613.
41. Hagberg, A., Swart, P., & S Chult, D. (2008). *Exploring network structure, dynamics, and function using NetworkX* (No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Lab.(LANL), Los Alamos, NM (United States).
42. yWorks, t. (2019). yWorks - The Diagramming Company. Retrieved from <https://www.yworks.com/>