

**UNIVERSIDAD DE GRANADA**

**FACULTAD DE CIENCIAS**



**DEPARTAMENTO DE QUÍMICA FÍSICA**

**ASPECTOS DINÁMICOS DEL RECONOCIMIENTO DE  
LIGANDOS Y REGULACIÓN FUNCIONAL DE DOMINIOS  
SH3: HIDRATACIÓN, COOPERATIVIDAD Y EQUILIBRIO  
CONFORMACIONAL.**

**Carlos Corbi Verge**

**Tesis Doctoral  
GRANADA  
2012**

Editor: Editorial de la Universidad de Granada  
Autor: Carlos Corbi Verge  
D.L.: GR 727-2013  
ISBN: 978-84-9028-440-7

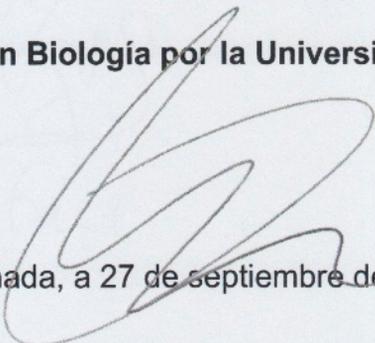


UNIVERSIDAD DE GRANADA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE QUÍMICA FÍSICA

**Memoria presentada para aspirar al grado de Doctor en Química**

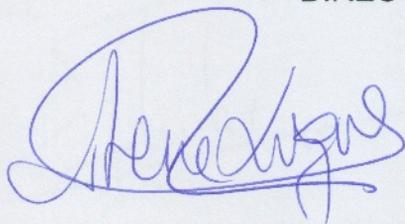
**Fdo: Carlos Corbi Verge**

**Licenciado en Biología por la Universidad de Girona**

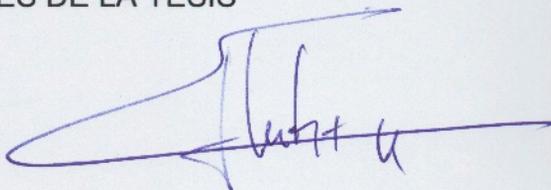


Granada, a 27 de septiembre de 2012

Vº Bº  
DIRECTORES DE LA TESIS



**Fdo: Dra. Irene Luque Fernández**  
Profesor Titular  
Departamento de Química Física  
Universidad de Granada



**Fdo: Dr. José C. Martínez Herrerías**  
Profesor Titular  
Departamento de Química Física  
Universidad de Granada



El doctorando Carlos Corbi Verge y los directores de la tesis Irene Luque Fernández y José C. Martínez Herrerías. Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

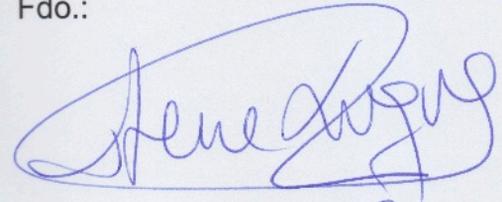
Granada, de septiembre de 2012

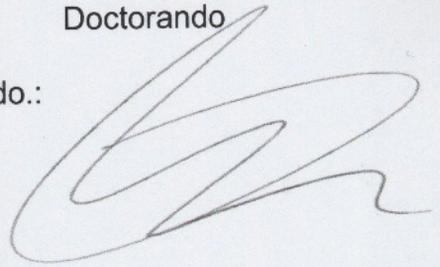
Director/es de la Tesis

Doctorando

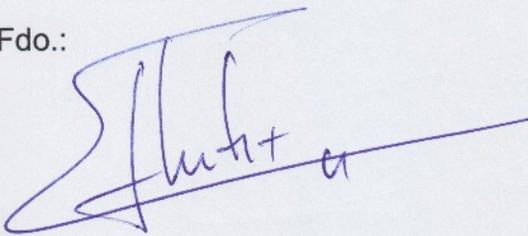
Fdo.:

Fdo.:

  
Irene Luque Fernández



Fdo.:









## **CONTENIDO**

<b>1 INTRODUCCIÓN</b>	<b>1</b>
<b>1.1 LAS QUINASAS DE TIROSINA CSRC Y CABL</b>	<b>4</b>
<b>1.2 DOMINIOS MODULARES: EL DOMINIO SH3</b>	<b>11</b>
<b>1.3 OBJETIVOS</b>	<b>21</b>
<b><u>2 ESTUDIO DE LA DINÁMICA CONFORMACIONAL DE DOMINIOS SH3 Y SUS COMPLEJOS CON LIGANDOS PEPTÍDICOS. EFECTOS COOPERATIVOS Y REESTRUCTURACIÓN DE INTERACCIONES.</u></b>	<b>23</b>
<b>2.1 INTRODUCCIÓN</b>	<b>25</b>
<b>2.2 RESULTADOS Y DISCUSIÓN</b>	<b>32</b>
2.2.1 ANÁLISIS DE LA DINÁMICA ESTRUCTURAL	34
2.2.2 ANÁLISIS DE LAS INTERACCIONES; ENLACES DE HIDRÓGENO Y PUENTES SALINOS	39
2.2.3 EXÁMEN COMPARATIVO DE LAS INTERACCIONES INTERMOLECULARES	40
2.2.4 ANÁLISIS COMPARATIVO DE LAS INTERACCIONES INTRAMOLECULARES	45
<b>2.3 CONCLUSIONES</b>	<b>54</b>
<b><u>3 ZAHORÍ: UN ALGORITMO PARA EL ANÁLISIS AUTOMATIZADO DE HIDRATACIÓN INTERFACIALES MEDIANTE DINÁMICA MOLECULAR</u></b>	<b>56</b>
<b>3.1 INTRODUCCIÓN</b>	<b>58</b>
3.1.1 PAPEL ENERGÉTICO DE LAS MOLÉCULAS DE AGUA EN EL SITIO DE UNIÓN DEL DOMINIO SH3-CABL	60
3.1.2 HUELLA TERMODINÁMICA DE LOS PÉPTIDOS DE DISEÑO PARA SH3-CABL. ANÁLISIS DE LOS POSIBLES CAMBIOS EN EL PATRÓN DE HIDRATACIÓN	62
3.1.3 PATRÓN DE SITIOS DE HIDRATACIÓN EN LA FAMILIA DE DOMINIOS SH3	66
3.1.4 ESTUDIO Y SEGUIMIENTO DE LAS MOLÉCULAS DE AGUA EN LOS SITIOS DE HIDRATACIÓN DE LOS DOMINIOS SH3	71
<b>3.2 RESULTADOS Y DISCUSIÓN</b>	<b>74</b>
3.2.1 ZAHORÍ: UN ALGORITMO PARA EL SEGUIMIENTO DE MOLÉCULAS DE AGUA EN TRAYECTORIAS DE DINÁMICA MOLECULAR	75
3.2.2 ANÁLISIS DE LAS DIFERENCIAS EN ENTALPÍA ENTRE LOS LIGANDOS DE DISEÑO DEL DOMINIO SH3 DE CABL	83
3.2.3 APLICACIÓN DEL ALGORITMO ZAHORÍ AL ESTUDIO DE PATRONES DE HIDRATACIÓN EN DISTINTAS SUBFAMILIAS DE DOMINIOS SH3	94
<b>3.3 CONCLUSIONES</b>	<b>109</b>
<b><u>4 ANÁLISIS DE LA DISTRIBUCIÓN CONFORMACIONAL DEL TÁNDEM SH32 DE CABL Y SU MODULACIÓN POR INTERACCIONES INTRAMOLECULARES</u></b>	<b>111</b>
<b>4.1 REGULACIÓN DE LA QUINASA DE TIROSINA CABL</b>	<b>113</b>
4.1.1 REGULACIÓN DE CABL POR PARTE DE LOS DOMINIOS MODULARES SH3 Y SH2	117
4.1.2 EL PAPEL DE LA REGIÓN N-TERMINAL (NCAP) EN LA REGULACIÓN DE CABL	122
<b>4.2 RESULTADOS Y DISCUSIÓN</b>	<b>126</b>

4.2.1	DINÁMICA CONFORMACIONAL DEL TÁNDEM SH32	127
4.2.2	MODIFICACIÓN DE LA DINÁMICA CONFORMACIONAL DEL TÁNDEM SH32 EN PRESENCIA DE NCAP Y C2Q	135
4.2.3	PAPEL DEL CONECTOR C3-2 EN LA DINÁMICA CONFORMACIONAL DEL TÁNDEM SH32	142
4.2.4	SIMULACIONES DE METADINÁMICA	155
4.2.5	ESTUDIO POR CALORIMETRÍA DIFERENCIAL DE BARRIDO DE LAS CONSTRUCCIONES DEL TÁNDEM SH32	169
<b>4.3</b>	<b>CONCLUSIONES</b>	<b>173</b>
<b>5 MATERIALS AND METHODS</b>		<b>175</b>
<b>5.1</b>	<b>AN INTRODUCTION TO MOLECULAR DYNAMICS OF BIOMOLECULES</b>	<b>177</b>
5.1.1	THE FUNDAMENTALS OF THE METHODOLOGY	177
5.1.2	AN IMPORTANT MATTER: THE TREATMENT OF BOUNDARIES	181
5.1.3	CONSTANT TEMPERATURE AND PRESSURE ENSEMBLE	182
5.1.4	THE MOLECULAR MECHANICS FORCE FIELD IN BIOLOGICAL MACROMOLECULES	185
5.1.5	THE SOLVENT MODELS FOR BIOMOLECULES	189
5.1.6	AN INTRODUCTION TO METADYNAMICS	192
<b>5.2</b>	<b>MOLECULAR DYNAMICS SIMULATIONS CARRIED OUT WITH AMBER</b>	<b>198</b>
<b>5.3</b>	<b>MOLECULAR DYNAMICS SIMULATIONS CARRIED OUT WITH NAMD</b>	<b>199</b>
<b>5.4</b>	<b>MODELLING OF THE 3D STRUCTURES</b>	<b>200</b>
<b>5.5</b>	<b>ANALYSIS OF MOLECULAR DYNAMICS TRAJECTORIES</b>	<b>201</b>
<b>5.6</b>	<b>METADYNAMICS FOR CABL TANDEM.</b>	<b>202</b>
5.6.1	COLLECTIVE VARIABLES.	203
5.6.2	RESTRAINTS	205
5.6.3	ANALYSIS OF THE METADYNAMICS: ESTIMATING THE GIBBS FREE ENERGY	207
<b>6 RESUMEN Y CONCLUSIONES</b>		<b>209</b>
<b>6.1</b>	<b>RESUMEN</b>	<b>211</b>
<b>6.2</b>	<b>CONCLUSIONES</b>	<b>215</b>
<b>7 SUMMARY AND CONCLUSIONS</b>		<b>219</b>
<b>7.1</b>	<b>SUMMARY</b>	<b>221</b>
<b>7.2</b>	<b>CONCLUSIONS</b>	<b>225</b>
<b>8 BIBLIOGRAFÍA</b>		<b>229</b>
<b>9 APÉNDICES</b>		<b>248</b>
<b>9.1</b>	<b>CLONADO, EXPRESIÓN Y PURIFICACION DE LOS CONSTRUCTOS DEL TÁNDEM CABL</b>	<b>250</b>
9.1.1	MUTANTES	251
<b>9.2</b>	<b>CALORIMETRÍA DIFERENCIAL DE BARRIDO</b>	<b>252</b>
<b>9.3</b>	<b>SUPERFICES DE ENERGIA LIBRE</b>	<b>253</b>





---

**1 INTRODUCCIÓN**

CAPÍTULO 1

## Introducción

Las proteínas son parte esencial de los organismos vivos, estando implicadas en la mayoría los procesos celulares, desde las reacciones bioquímicas fundamentales para el metabolismo, catalizadas por los enzimas, hasta las funciones mecánicas de la actina y la miosina que permiten la contracción de los músculos. A su vez, son también las encargadas de gestionar todos estos procesos de forma adecuada para mantener la homeostasis celular y responder correctamente a estímulos y señales. Para ello, las células han desarrollado una extensa red de receptores y mensajeros que interaccionan de forma específica desencadenando la expresión de nuevos genes, la alteración de las rutas metabólicas, la división celular o incluso la apoptosis celular. De esta forma, cualquier error en este sistema de procesamiento de información puede llevar al desarrollo de patologías como son las enfermedades autoinmunes o el cáncer, lo que convierte a las proteínas implicadas en importantes dianas para el desarrollo de terapias contra innumerables patologías [1].

La base de la actividad de este complejo entramado funcional se basa en el reconocimiento específico entre proteínas y en su activación/desactivación transitoria, con frecuencia mediante la fosforilación o desfosforilación de un residuo que tenga un grupo hidroxilo libre (-OH); normalmente serina, treonina o tirosina. Las quinasas son las proteínas encargadas de transferir ese grupo fosfato desde el adenosín trifosfato (ATP) al residuo de la proteína sustrato, provocando cambios conformacionales que alterarán la actividad de ésta última de diversas maneras, dependiendo de su localización y tipo. Muchas proteínas tienen múltiples sitios susceptibles de fosforilación, incluidas las propias quinasas. La fosforilación puede ser revertida por una clase diferente de enzimas llamadas fosfatasas. La

## CAPÍTULO 1

intervención conjunta y coordinada de quinasas y fosfatasa permite así la activación y desactivación de numerosas funciones celulares [1]

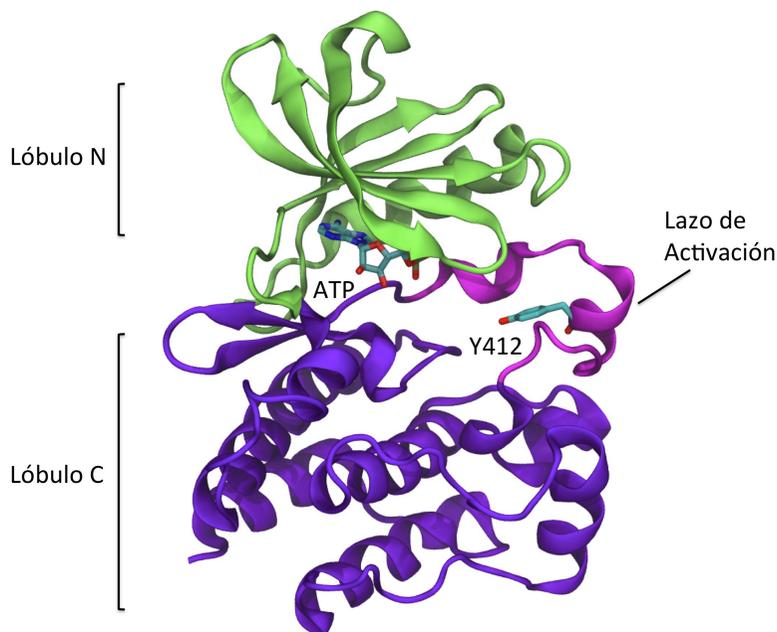
Las quinasas han sido clasificadas en 10 grupos dependiendo de su actividad (no siempre correlacionada con su función biológica), homología de secuencia y estructura. Tenemos así quinasas moduladas por nucleótidos, fosfolípidos y calcio (AGC); quinasas reguladas por calcio y calmodulina (CAMK); caseína quinasas (CK1), homologas a las quinasas de levadura (STE); quinasas que contienen un dominio receptor de guanilato ciclasa (RGC); tirosín quinasas que fosforilan casi exclusivamente residuos de tirosina (TK); quinasas que tienen una alta similitud con las homólogas de tirosina pero que fosforilan generalmente residuos serina y treonina (TKL); quinasas de las subfamilias CDK, MAPK, GSK3 y CLK (CMGC); y finalmente dos grupos más difíciles de definir, donde se agrupan quinasas sin homología estructural y de secuencia con el resto .[2].

Entre todos estos grupos destaca el de las quinasas de tirosina por ser uno de principales protagonistas en los procesos celulares y por ser estas quinasas de las más abundantes en el genoma, según se ha demostrado en la búsqueda por homología de secuencia del genoma humano, donde se encontraron 518 secuencias que codifican quinasas, siendo las quinasas de tirosina el grupo más ampliamente representado, con 90 miembros [2].

### **1.1 Las quinasas de tirosina cSrc y cAbl**

En 1979, J. Michael Bishop y Harold E. Varmus descubrieron que en el genoma de pollos sanos se encontraba un gen con una alta homología de secuencia con el gen vSrc del virus del sarcoma de Rous (RSV). Este gen, el vSrc, es indispensable para que el virus RSV posea capacidad tumoral. Un

profundo estudio comparativo demostró que vSrc, a diferencia de su homólogo aviar cSrc, carecía de la región C-terminal, de forma que la proteína se encontraba constitutivamente activa, al contrario de la forma celular que sólo se activaba bajo ciertas circunstancias. La proteína codificada por estos genes tenía la capacidad de transferir un grupo fosfato a un residuo de tirosina, es decir, se trataba de una quinasa de tirosina. Esta actividad descontrolada era la causante de la transformación de las células sanas a tumorales, de forma que mutaciones en el gen cSrc podían producir cáncer. Se establece así un nuevo paradigma en el panorama de la biología celular, acuñando los términos oncogén (vSrc) y proto-oncogén (cSrc) [3-5].

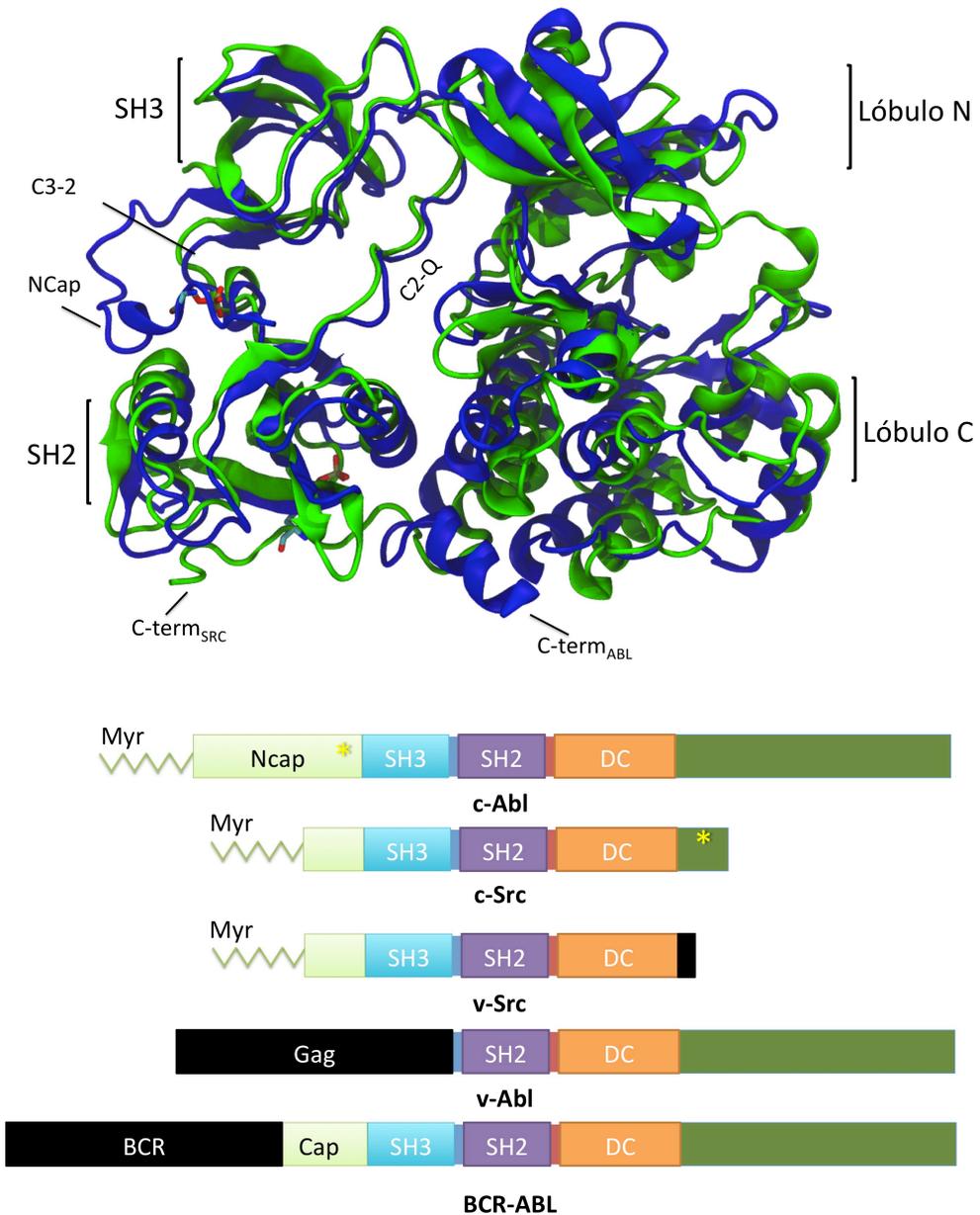


**Figura 1.1 Esquema del dominio catalítico de las quinasa de tirosina de cSrc. Código PDB: 2FO0.** Se ha coloreado en verde la sección correspondiente al lóbulo N y en lila el lóbulo C. En fucsia está representado el lazo de activación, que pertenece al lóbulo C. Se ha representado en varillas el residuo Tyr 412, que es fosforilado para la activación del dominio catalítico. Para detallar dónde se incorpora el ATP, también se ha representado el ATP en varillas

## CAPÍTULO 1

Casi paralelamente, se descubrió en ratones cómo el producto del gen tumoral del virus del linfosarcoma de Abelson (vAbl) era también una forma alterada de la quinasa celular (cAbl) y, al igual que cSrc, tenía potencial para transformar las células sanas en tumorales, amén de la capacidad catalítica de fosforilar residuos de tirosina [9, 10, 12]. Posteriormente, se observó que una forma defectiva de cAbl, generada por translocación cromosómica, provocaba la fusión del gen BCR con el de cAbl; esta proteína fue detectada y caracterizada en casi todas las formas de leucemia mieloide crónica (CML) [6, 7]. Tanto ésta como la forma vírica de cAbl tienen la actividad quinasa desregulada así como su capacidad de transformar células sanas en cancerígenas, tanto *in vitro* como *in vivo* [8, 9].

Así pues, el descubrimiento de los oncogenes virales de cSrc y cAbl desencadenó un gran interés científico, ya que el poderoso efecto de los oncogenes en el comportamiento celular sugería que la comprensión de sus mecanismos revelaría los principios que gobiernan la regulación celular y la oncogénesis. El esfuerzo investigador se centró inicialmente en entender el mecanismo catalítico de las quinasas de tirosina. Subsiguientes estudios han mostrado cómo el dominio catalítico de las quinasas de tirosina se encuentra muy conservado, tanto a nivel estructural como en sus mecanismos de regulación de la actividad [10]. Dicho dominio está formado por dos lóbulos: el lóbulo que se encuentra en el extremo N-terminal (lóbulo N), que está formado por cinco hebras beta y una sola hélice alfa y el segundo lóbulo, en el extremo C-terminal (lóbulo C), que está mayormente compuesto por hélices alfa y es donde se encuentran la zona de reconocimiento del sustrato peptídico y el centro activo (FIGURA 1.1). El ATP se une a la zona situada entre los dos lóbulos, cerca del sitio de unión de las secuencias diana de la quinasa.



**Figura 1.2 Superposición de las estructuras de las quinasas de tirosina cSrc (en verde) y cAbl (en azul) acompañado del esquema de la organización de dominios de diferentes formas de quinasas de tirosina de SRC y ABL.** En el panel superior; Superposición de las estructuras destacando la estructura secundaria, a partir de los PDB con código 2FO0 (cAbl) y 2SRC (cSrc). En el panel inferior; Comparación de la organización de dominios entre las formas oncogénicas y silvestres de los miembros de las familias de SRC y ABL. Cada dominio se encuentra representado en un color, siendo el negro utilizado para resaltar partes que han sido alteradas. El asterisco amarillo marca la localización de residuos fosforilados.

## CAPÍTULO 1

La mayor parte de las quinasas se activan por medio de un cambio conformacional provocado por la fosforilación de uno de sus residuos. Esta fosforilación puede realizarse por otras quinasas o por copias de ellas mismas, lo que organiza una intrincada red de comunicación [11-13]. El mecanismo más comúnmente observado consiste en la fosforilación del llamado lazo de activación, en el que la carga negativa del grupo fosfato compensa una carga positiva situada en el motivo conservado del lazo de activación, aumentando la actividad catalítica de la proteína. Pero existen otros eventos de fosforilación que pueden desactivar la quinasa o cambiar el sitio de fosforilación [10, 14], estando los dominios que acompañan al dominio catalítico especialmente implicados en su regulación [14]. Los constructos de cAbl y cSrc a los que posteriormente se les había deleccionado la sección N-terminal mostraron una actividad anómala y un cambio en las preferencias de substratos [15, 16]. De forma que, aunque el extremo N-terminal no era imprescindible para que la quinasa de tirosina desarrollara su actividad enzimática, si estaba implicada en la modulación de esta. La inspección de estas regiones en las familias Src y Abl ha puesto de manifiesto que los 100 residuos inmediatamente adyacentes al dominio catalítico se encuentran altamente conservados entre ellas. Esta región se denominó como dominio Src Homology 2 (SH2), entendiéndose que el término SH1 correspondía al dominio catalítico [16]. Posteriormente, a medida que se han ido secuenciando más oncogenes, se ha encontrado que, junto a la región conservada llamada SH2, existe otra algo más pequeña que, al igual que SH2, no parece tener actividad catalítica, pero que también modula el funcionamiento de estas quinasas de tirosina. Este nuevo dominio se llamó, en clara lógica, SH3 [17]. En la FIGURA 1.2 podemos apreciar la disposición general de estos dominios en las quinasas de tirosina de la familia de cSrc, cAbl, tanto en sus formas celulares como en sus formas víricas y oncogénicas.

Hoy en día, 30 años después del descubrimiento de cSrc y cAbl como proto-oncogenes, sabemos que la pareja de dominios SH3 y SH2 tiene la capacidad de reconocer motivos de secuencia específicos y están relacionados con el mecanismo de regulación de la actividad enzimática, de forma que el dominio catalítico no tiene la exclusividad, ni en los mecanismos de reconocimiento del sustrato ni en los de regulación de la actividad [18-20]. Por ejemplo, tanto la delección del dominio SH3 como del SH2 provoca la desregulación de la actividad de cAbl, pero cada uno modifica de forma característica el comportamiento de cAbl in vivo [21].

Aunque el mecanismo no está aún completamente detallado y no es exactamente igual para cSrc y cAbl (como veremos con mayor detalle en la Capítulo 4 de esta Memoria), en general, el dominio catalítico se mantiene en su conformación autoinhibida mediante el establecimiento de interacciones intramoleculares, en las que juegan un papel destacado los dominios SH3 y SH2 [22, 23]. Así, es necesaria la disrupción de estas interacciones para que el dominio catalítico pueda llevar a cabo su actividad. Esta disrupción se desencadena por un estímulo que puede ser de diferente naturaleza y que, interactuando en diferentes zonas de la quinasa, desencadena en todos los casos grandes cambios conformacionales [18, 21]. Esto pone de manifiesto que, aunque modular, las estructuras de cAbl y cSrc son altamente cooperativas. Esta cooperatividad podría reflejar la necesidad de estas quinasas de responder a un alto número de estímulos, lo que les confiere la capacidad de estar implicadas en diferentes vías de señalización [24]. Algunos autores proponen la existencia de diferentes estados activos para cada estímulo, lo que permitiría modular la respuesta de la quinasa en función del estímulo activador. Aunque no poseemos suficiente información estructural

## CAPÍTULO 1

del estado activo de ambas quinasas, existen algunas evidencias que apoyarían esta hipótesis [25-27].

Dada la complejidad del mecanismo, el funcionamiento de cAbl y cSrc sigue eludiendo nuestra completa comprensión, convirtiéndose en uno de los principales retos biomédicos actuales, pues no solamente permitiría entender con mayor detalle cómo funciona uno de los elementos claves de la señalización celular, sino que, además, sería de gran ayuda para desarrollar nuevas terapias contra las formas oncogénicas de estas quinasas. Estas terapias son especialmente necesarias en el caso de la leucemia crónica mieloide, donde es común que los pacientes desarrollen resistencia a los fármacos actuales. Dichas formas resistentes, al contrario de lo que cabría esperar, no sólo presentan mutaciones en el sitio activo de cAbl, sino también a los dominios SH3 y SH2 entre otras zonas [28, 29].

Aunque no es el único, uno de los mecanismos de activación de cAbl y cSrc mas estudiados es el desencadenado por el reconocimiento del dominio SH3 de proteínas con motivos ricos en prolina [30, 31]. Formas de cAbl que carecen del dominio SH3, como es el caso de la proteína viral vAbl, se encuentran desreguladas y con capacidad para transformar celular sanas en cancerígenas [32]. Esto sigue su importancia en el mecanismo de regulación y de reconocimiento de sustratos, convirtiéndose en un prometedor campo de estudio para el desarrollo de nuevos fármacos. En este punto, se ha demostrado que el bloqueo del sitio de unión del dominio SH3 de la proteína Lck (de la familia de Src) mediante la modificación del péptido p1 (KERPPPVPNPD) con un derivado del ácido palmítico inhibe el crecimiento de células cancerígenas, tanto *in vitro* como *in vivo* [33].

El dominio SH3, no se encuentra solamente en cSrc y cAbl, sino que está ampliamente distribuido en el proteoma, siendo una de las piezas clave para

entender los mecanismos de señalización. Destacar que sus interacciones con ligandos han sido durante la última década uno de los principales temas de estudio de nuestro grupo de investigación.

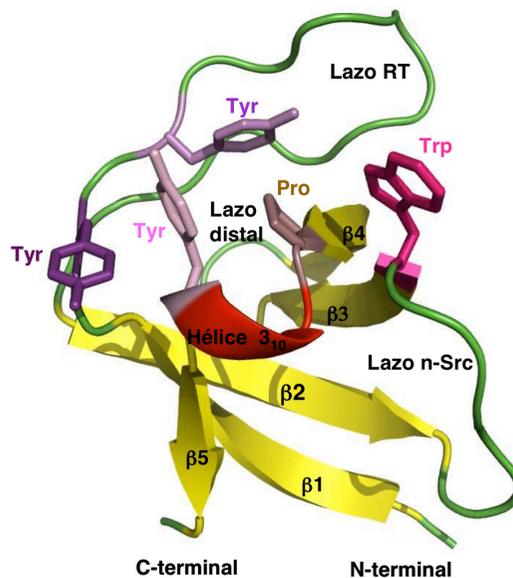
## 1.2 Dominios modulares: el dominio SH3

La presencia de dominios modulares de reconocimiento proteína-proteína, como son los dominios SH2 y SH3, es muy frecuente en diferentes tipos de proteínas [14, 34]. Gracias a las mejoras en las técnicas de biología molecular, se ha podido determinar y clasificar los diferentes tipos de dominios, atendiendo a su estructura y función.

Se denominan dominios modulares aquellos dominios con identidad propia y estructura conservada, cuya función principal es la de reconocimiento de secuencias, entre 3 y 6 aminoácidos, específicas y muy conservadas. Estos dominios son generalmente pequeños, entre 30 y 150 aminoácidos, y se caracterizan por un plegamiento compacto y estable, presentando uno o más sitios de unión para ligandos [35]. Se encuentran en proteínas de todo tipo y no es raro encontrar más de un dominio modular formando parte de una misma proteína [18, 34, 36-39]. Los dominios modulares se clasifican en diferentes familias dependiendo de su homología estructural, de secuencia y del motivo específico que reconocen. Por ejemplo, los dominios SH2 reconocen secuencias con una tirosina fosforilada, los PDZ reconocen secuencias correspondientes a extremos carboxilo terminal y los dominios SH3 reconocen secuencias ricas en prolina (MRP). De este último tipo actualmente se conocen hasta cinco familias: SH3, WW, EVH1, GYF y UEV, siendo los dominios SH3 los más ampliamente representados en virus,

## CAPÍTULO 1

procariotas y eucariotas, formando parte de todo tipo de proteínas [36-38, 40].



**Figura 1.3** Representación de la estructura tridimensional del dominio SH3 de c-Src. La estructura del panel superior fue resuelta por cristalografía de rayos X (Martin-García et al., 2007) y depositada en el Protein Data Bank (PDB) bajo el código 2hda. Las cinco hebras beta se representan como β1, β2, β3, β4 y β5 (amarillo) y la hélice 310 que conecta las hebras β4 y β5 aparece en rojo. El lazo n-Src que conecta las hebras β2 y β3, el lazo RT que conecta las hebras β1 y β2, y el lazo distal que conecta las hebras β3 y β4, también se indican (verde). Los restos aromáticos más conservados que son importantes para la unión de los ligandos se indican en representación de varillas con el código de colores indicado en el texto. Éstos son los dos residuos de tirosina del motivo ALYDY, el primer triptófano del motivo WW, el residuo prolina de la hebra β4 y el residuo tirosina de secuencia SNY de la hélice 310

Los dominios SH3 están formados por unos 60 residuos y comparten una alta identidad de secuencia y una estructura común formada por 5 hebras beta antiparalelas unidas por 3 lazos (denominados comúnmente como RT, n-Src y Distal) y una hélice 3<sub>10</sub>. Figura 1.3. El sitio de unión de los dominios SH3

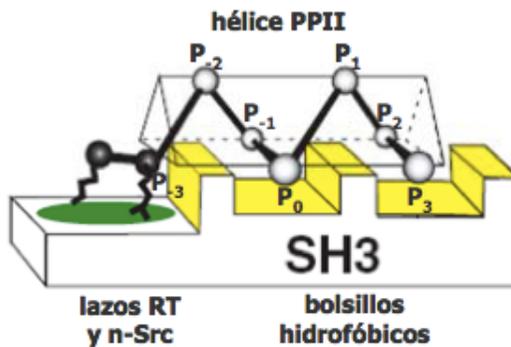
está compuesto por dos hendiduras poco profundas que forman los llamados bolsillos hidrofóbicos, ya que están constituidos por residuos aromáticos muy conservados. Colindante a éstos existe un tercer bolsillo responsable de la especificidad de unión, delimitado por los lazos RT y n-Src, más variables en secuencia entre los diferentes miembros de la familia Figura 1.4.

El motivo central de unión de esta familia está ampliamente descrito y se caracteriza por la secuencia consenso PxxP (donde x es normalmente un aminoácido hidrofóbico) que acomoda cada dipéptido xP en cada uno de los dos bolsillos hidrofóbicos del sitio de unión, de forma que el ligando adopta una estructura secundaria característica en hélice poliprolina tipo II (PPII), Figura 1.5. Esta conformación de tres residuos por vuelta está muy favorecida por la presencia de las prolina y su peculiar cadena lateral ciclada. Los restos adyacentes al motivo central PxxP interaccionan con el tercer bolsillo, delimitado por los lazos RT y n-Src, y están menos conservados, estando relacionados con la especificidad de reconocimiento entre diferentes dominios [41].

La geometría del sitio de unión permite que las secuencias ricas en prolina se puedan unir en dos orientaciones diferentes respecto al sitio de unión, lo que da lugar a dos tipos de ligandos: tipo I, caracterizados por la secuencia consenso [R/K]xxPxxP, donde el ligando va del extremo N-terminal al C-terminal; y tipo II, con la secuencia consenso xPxxPX[R/K] donde el ligando se acomoda en el sentido opuesto, desde el C-terminal al N-terminal Figura 1.6. La orientación del ligando viene determinada por un aminoácido básico localizado a dos residuos de distancia del motivo PxxP (hacia el extremo N-terminal en los ligandos de tipo I y hacia el C-terminal en los ligandos de tipo II) que forma un puente salino con un residuo ácido muy conservado en el dominio, que se encuentra en el lazo RT [41, 42]. Además de

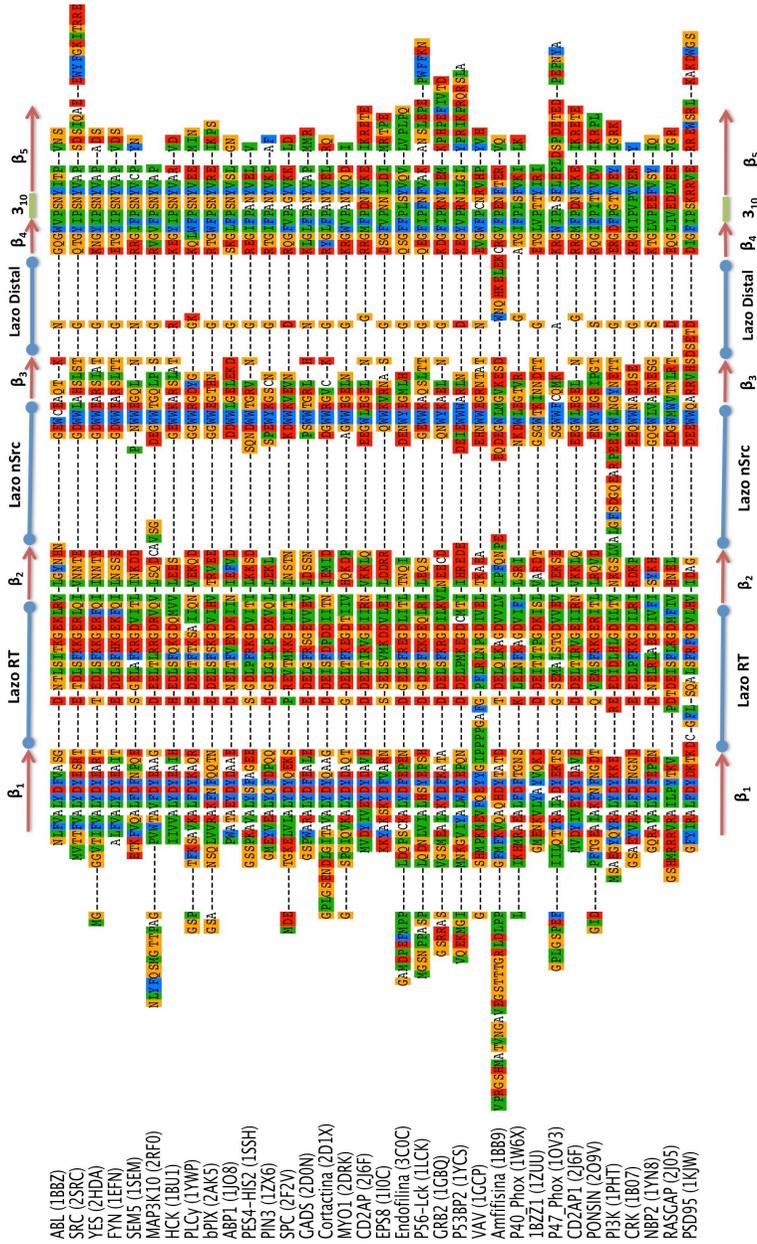
## CAPÍTULO 1

estas interacciones electrostáticas favorables, también es importante la formación de interacciones de van der Waals de la cadena lateral del residuo básico del ligando con un triptófano altamente conservado del sitio de unión [43]. Algunos miembros de la familia de dominios SH3 prefieren una orientación sobre la otra mientras que otros no hacen distinción. Por ejemplo, se ha podido comprobar cómo los dominios SH3 de la familia de c-Src forman frecuentemente complejos tipo I y II, mientras que los complejos de cAbl son mayoritariamente de tipo I [44]



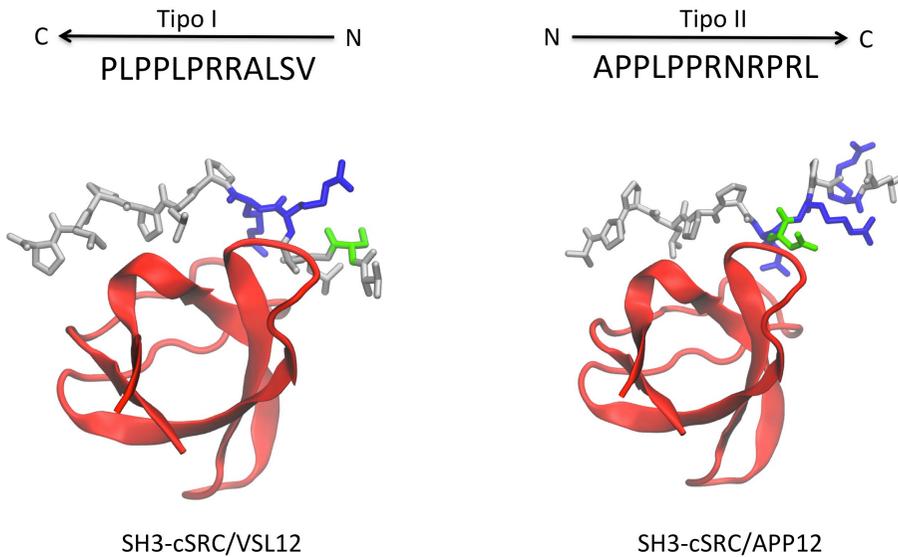
**Figura 1.5 Representación esquemática del modo de unión mostrado por dominios SH3.** Concretamente se representa el sitio de unión ocupado con un ligando tipo PxxP en orientación tipo I. El sitio de unión contiene dos bolsillos hidrofóbicos xP formados por restos aromáticos conservados (en amarillo) y un bolsillo de especificidad adyacente que es más variable (en verde). Las letras P numeradas en el ligando indican las posiciones respecto de la secuencia consenso P<sub>0</sub>X<sub>1</sub>X<sub>2</sub>P<sub>3</sub>.

Además del reconocimiento de péptidos que contengan el motivo PxxP, recientemente se han encontrado dominios SH3 que presentan afinidad por secuencias consenso diferentes [45], en algunos casos con motivos atípicos que carecen de residuos prolina [46]. Esto ampliaría el espectro de acción de los dominios SH3, pero aún es necesario un estudio más profundo para conocer mejor las peculiaridades de estos casos.



**Figura 1.4 Alineamiento de secuencia de varios representantes de la familia de dominios SH3 con estructura resuelta por rayos-X.** En la parte superior e inferior se representa mediante un esquema a que región de la estructura secundaria corresponde cada parte de la secuencia. Los diferentes residuos están resaltado en colores dependiendo de las propiedades de los residuos, en rojo los residuos cargados, en verde los residuos cargados, en azul los aromáticos y en naranja los polares

## CAPÍTULO 1



**Figura 1.6 Representación del dominio SH3-cSrc formando complejo tipo I con el ligando VSL12 (panel izquierdo) y tipo II con el ligando APP12 (panel derecho).** A partir del PDB con código 1QWF para SH3-cSrc/VSL12 y 1QWE para SH3-cSrc/APP12. En rojo se destaca el dominio SH3-cSrc. Los ligandos se han mostrado en estructura de varilla y coloreados según las propiedades de cada residuo, en Azul los básicos, en Verde los polares y en Gris los apolares

Los dominios SH3 se caracterizan por presentar una afinidad de unión moderada por sus ligandos naturales ( $K_d=1-100\mu\text{M}$ ), permitiendo la formación de complejos transitorios, característica deseable para desempeñar su papel en los complejos mecanismos de transducción de señales. En la misma línea, también presentan una especificidad lo suficiente elevada como para discriminar entre substratos, pero no es extraño que distintos dominios reconozcan algunos motivos comunes. Por ejemplo, se ha podido comprobar cómo los ligandos VSL12 (VSLARRPLPLP) y APP12 (APPLPLPRNRPRL), descubiertos mediante expresión de fagos, interaccionan de forma exclusiva con los dominios SH3 de la familia de cSrc, mientras que no forman complejos con los de la familia de cAbl. No obstante, entre estas dos familias se han descrito casos en los que un mismo péptido puede unir a dominios de ambas con afinidades similares. Este es el caso de los ligandos naturales 3BP-1

(RAPTMPPPLPP) y 3BP-2 (PPAYPPPVP) que interactúan con el dominio SH3 de cAbl y con el dominio SH3 de Fyn (de la familia de cSrc) con unas constantes de disociación en torno a  $35\mu\text{M}$  [47]. Por otro lado, tenemos el caso de la proteína asociada a filamentos de Actina de 110 kDa (AFAP-10) que ha sido identificada como uno de los sustratos naturales del dominio SH3 de la tirosina quinasa cSrc, tanto *in vivo* como *in vitro* [48, 49], pero que no interactúa con el dominio SH3 de cYes a pesar de ser de la misma familia y poseer una altísima homología de secuencia, diferenciándose en solamente 8 residuos [50].

Es decir, encontramos algunos casos en los que existe una especificidad muy fina, y otros en los que existe una gran promiscuidad en la unión de ligandos. La pregunta es obvia: ¿cómo encaja este comportamiento en la sensible red de mecanismos que regulan los procesos celulares? Este balance entre reconocimiento específico y promiscuo podría responder a la necesidad de mantener una cierta interconectividad entre las redes de comunicación. Así, por ejemplo, es lógico pensar que ciertos estímulos requieran la modulación de dos procesos celulares diferentes bajo ciertas circunstancias, mientras que bajo otras solamente sea necesaria la modulación de uno [43]. Otra posible explicación a este comportamiento se debería a que los fenómenos de transducción de señales se encuentran normalmente muy localizados en la célula, como es el caso de las tirosina quinasas de cSrc, que al contrario que cAbl están unidas a la membrana plasmática, de forma que directamente no existiría interferencia por reconocimientos cruzados al no estar presentes estas proteínas en el mismo compartimiento celular o incluso tipo celular. Por otro lado, donde quedan más resaltados los fenómenos de solapamiento de especificidad es en los experimentos *in vitro*, donde en la mayoría de las ocasiones se exponen los dominios SH3 a secuencias que no son naturales o

## CAPÍTULO 1

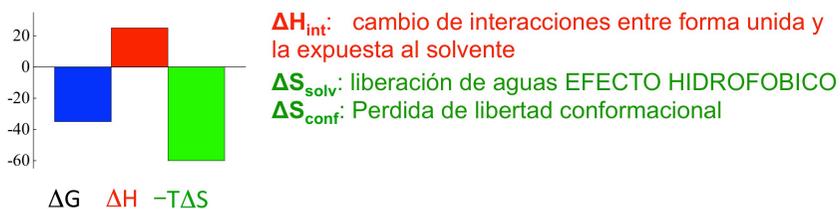
que no corresponden a proteínas del mismo organismo. Por lo tanto, el solapamiento de especificidad observado se debería a que los dominios SH3 de un organismo estarían preparados para discriminar entre secuencias que existen en su entorno, de forma que una secuencia foránea puede ser reconocida por más de un dominio, como demuestra el trabajo de M. Zarrinpar et. al. con los dominios SH3 de levadura y un ligando de origen fisiológico, Pbs2 [39]. En este estudio se observó cómo el ligando era reconocido específicamente por un solo dominio SH3 de levadura, mientras que el mismo péptido era reconocido por diversos dominios SH3 de otros organismos, lo que reforzaría la hipótesis de la localización celular para explicar la versatilidad de los dominios SH3.

En todo caso, queda patente la existencia de un complejo mecanismo de reconocimiento, contrariamente a lo que cabría esperar para una proteína con un sitio de unión tan sencillo y pequeño. En este contexto, aunque se posee mucha información sobre los dominios SH3, el mecanismo que determina la afinidad y reconocimiento específico entre dominios SH3 aún no se conoce por completo. De hecho, la mayoría de las secuencias que interactúan con dominios SH3 han sido identificadas mediante el uso de bibliotecas de péptidos, química combinatorial y expresión de fagos, entre otras técnicas, mientras que los intentos de diseño racional, salvo en contadas ocasiones, han tenido escaso éxito [47]. Por un lado sabemos que la variabilidad de las secuencias de los lazos RT y n-Src juega un papel clave [47], pero deben existir otros protagonistas implicados en este complejo mecanismo que expliquen, por ejemplo, cómo los dominios SH3 de la familia cSrc, a pesar de ser muy similares en secuencia, presentan diferentes patrones de especificidad [50].

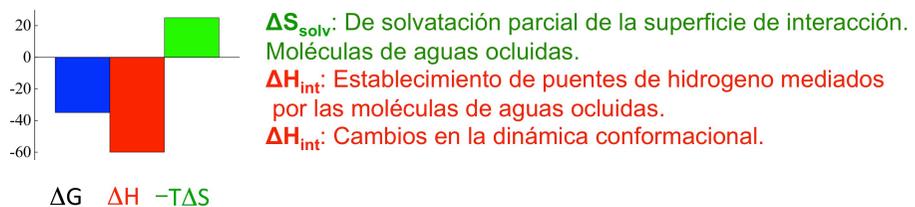
En este sentido, los diferentes estudios termodinámicos de la unión de ligandos ricos en prolina a dominios SH3 aportan información de gran

relevancia. Desde un punto de vista termodinámico, esperaríamos que este tipo de complejos, con escasas interacciones de naturaleza polar, presentara un patrón termodinámico dominado por el llamado efecto hidrofóbico, de forma que la interacción estaría dirigida por una entropía favorable y un cambio de entalpía positivo [51]. Sin embargo, todos los análisis calorimétricos llevados a cabo sobre complejos SH3-péptido revelan un patrón termodinámico totalmente diferente: una gran entalpía de unión negativa que estaría parcialmente compensada por una contribución entrópica desfavorable [52-54]. Este comportamiento no puede ser explicado solamente de acuerdo a la interacción directa entre superficies hidrofóbicas, lo que pone de manifiesto la existencia de otros mecanismos que contribuyen a la interacción de los dominios SH3 con sus ligandos. Figura 1.7

#### Huella termodinámica esperada



#### Huella termodinámica observada



**Figura 1.7 Esquema de las magnitudes esperadas y observadas para la unión de motivos ricos en prolina a dominios SH3.** Representación en forma de barras de la energía de Gibbs (Azul), la entalpía (Rojo) y la entropía (Verde). A la izquierda se remarcan los factores que contribuirían significativamente a la huella termodinámica esperada o la observada. En texto de color rojo factores que contribuyen a la entalpía y texto en color verde a la entropía

## CAPÍTULO 1

En esta dirección, y como veremos en detalle en los Capítulos 2 y 3 de esta Memoria, nuestro grupo de investigación ha estado trabajando para identificar los factores que determinan esta singular huella termodinámica y así mejorar el conocimiento que existe sobre el mecanismo de especificidad y afinidad de los dominios SH3, lo que permitiría el diseño racional de ligandos con potencial terapéutico. Uno de los factores hallados que podría estar contribuyendo a aumentar el factor entálpico sería la redistribución conformacional del estado nativo inducida por la unión del ligando. Los estudios de intercambio hidrógeno/deuterio indican que bajo condiciones nativas el dominio SH3 puede considerarse como un conjunto estadístico de estados conformacionales y presenta, además, una elevada cooperatividad estructural, lo que significa que cualquier cambio local puede transmitirse eficientemente por toda la estructura [55, 56]. Este hecho favorecería que, no sólo los residuos del sitio de unión participen en el mecanismo de reconocimiento y especificidad, sino que el dominio entero contribuyera a éste, lo que explicaría casos en los que dominios de la misma familia con una alta homología en el sitio de unión posean diferente afinidad por un mismo sustrato, como es el caso de la familia de dominios SH3 de cSrc con el ligando natural de AFAP-10, mencionado anteriormente.

Paralelamente, también en nuestro grupo de investigación se identificaron moléculas de agua ocluidas en la interfaz de unión que median las interacciones del complejo del dominio SH3 de cAbl con el péptido de diseño p41 (APSYSPPPPP). Se realizó un estudio exhaustivo a nivel estructural que mostró cómo estas moléculas de agua estarían estrechamente relacionadas con el patrón termodinámico de la unión de ligando del dominio SH3. El comportamiento de estas moléculas de agua reduciría la aportación favorable de entropía al no liberarse al solvente, pero estos efectos estarían

compensados por la formación de fuertes puentes de hidrógeno entre el dominio y el ligando, lo que contribuiría favorablemente a la entalpía de unión. [52, 53]. Esto explicaría por qué la unión del dominio SH3 de cAbl se caracteriza por presentar una entalpía muy favorable ligeramente compensada por una entropía positiva. Posteriormente, y también en nuestro grupo de investigación, se analizaron las estructuras cristalográficas de los dominios SH3 disponibles en el Protein Data Bank (PDB), lo que permitió observar cómo la presencia de moléculas de agua en el sitio de unión se repetía en todos los dominios, permitiendo identificar patrones de hidratación comunes [57]. Esto refuerza la idea de que las moléculas de agua son importantes en los mecanismos de unión y reconocimiento de secuencias en la familia de dominios SH3. Estas moléculas de agua forman parte estructural del sitio de unión y juegan un papel importante para la formación del complejo, bien mediando interacciones entre residuos, bien debilitando interacciones desfavorables, o bien rellenando espacios para mejorar los contactos entre las superficies proteicas.

### 1.3 Objetivos

En definitiva, a pesar del elevado número de estudios realizados a nivel estructural y funcional, aún no conocemos completamente los mecanismos que gobiernan la interacción de estos dominios con sus ligandos, lo que dificulta el diseño racional de nuevos ligandos de alta afinidad. A su vez, el comportamiento de los dominios SH3 juega un papel clave en los mecanismos de regulación de las proteínas de las que forman parte, como por ejemplo, cAbl y cSrc. El objetivo último que pretendemos obtener en esta Tesis es ampliar el conocimiento que se tiene del mecanismo de reconocimiento de secuencias ricas en prolina por parte de los dominios SH3, en concreto de los efectos de la

## CAPÍTULO 1

cooperatividad y del papel que juegan las moléculas de agua ocluidas. Asimismo, pretendemos estudiar el papel de las regiones no catalíticas de la quinasa de tirosina cAbl la regulación de su actividad catalítica y los mecanismos mediante los cuales esta regulación tiene lugar.

Los objetivos específicos perseguidos en este trabajo se enumeran a continuación:

- I. Desarrollar un programa que permita el seguimiento del comportamiento de las moléculas de agua de la interfaz de unión de distintos dominios SH3 durante simulaciones de dinámica molecular, para aumentar el conocimiento del papel de éstas en el fenómeno de reconocimiento y afinidad de ligando.
- II. Mediante técnicas computacionales detectar cambios cooperativos que permitan justificar la influencia en la energética de cambios en los parámetros termodinámicos de unión de posiciones alejadas del sitio de unión del dominio SH3 de cSrc.
- III. Describir el papel de las regiones no catalíticas de la quinasa de tirosina cAbl mediante el estudio de su equilibrio conformacional mediante técnicas computacionales.

---

**2 ESTUDIO DE LA DINÁMICA CONFORMACIONAL DE  
DOMINIOS SH3 Y SUS COMPLEJOS CON LIGANDOS  
PEPTÍDICOS. EFECTOS COOPERATIVOS Y  
REESTRUCTURACIÓN DE INTERACCIONES.**

CAPÍTULO 2

## 2.1 INTRODUCCIÓN

El reconocimiento de secuencias ricas en prolina por algunos dominios modulares juega un papel destacado en los procesos de reconocimiento e interacción proteína-proteína en el contexto de la señalización celular. Los dominios SH3 son los módulos de reconocimiento de secuencia ricas en prolina (MRP) más ampliamente representados en los diferentes proteomas y forman parte de diversas proteínas [38]. Actúan generalmente como sitios de anclaje para el reclutamiento de sustratos y la formación de complejos supramoleculares que conducen a la modificación enzimática de algunos de sus componentes. En ocasiones desempeñan también un papel esencial en la regulación de la actividad enzimática de las proteínas que los contienen mediante el establecimiento de interacciones intramoleculares con otros elementos de la molécula; éste es el caso de las quinasas de tirosina de las familias Src y Abl entre otras, que están implicadas en la regulación de importantes mecanismos celulares [22, 58].

Los dominios SH3 se caracterizan por presentar una afinidad de unión moderada por sus ligandos naturales ( $K_d=1-100\mu\text{M}$ ), y una especificidad lo suficientemente precisa para discriminar entre sustratos, pero a la vez permitiendo que distintos dominios reconozcan algunos motivos comunes. Estas características son necesarias para desempeñar su papel en los complejos mecanismos de transducción de señales. Por un lado, la débil afinidad permitiría la formación de complejos transitorios y, por otro, el equilibrio entre reconocimiento específico y promiscuo otorgaría plasticidad e interconectividad a las redes de interacción proteína-proteína. Como se ha descrito en la introducción de esta tesis, los parámetros termodinámicos de unión de las interacciones SH3-ligando son atípicos para lo que cabría esperar por su naturaleza principalmente hidrofóbica. La energía libre de unión en

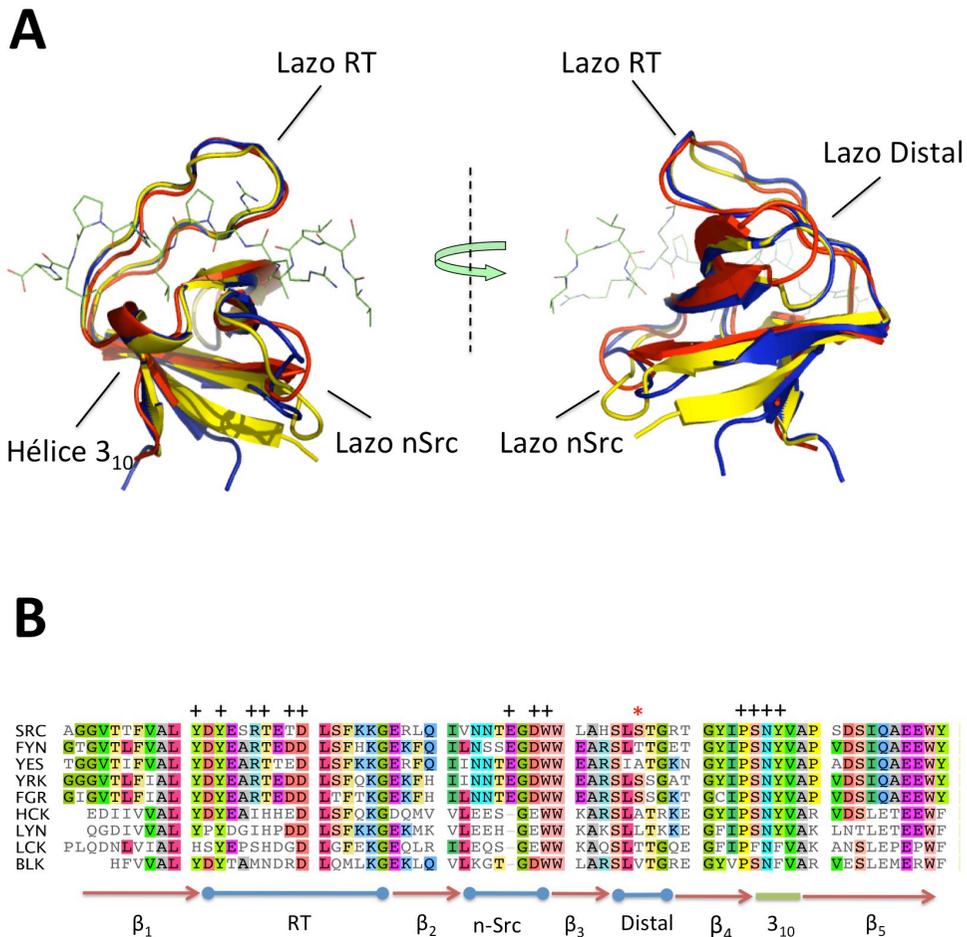
## CAPÍTULO 2

lugar de estar entrópicamente favorecida, parámetro típico del efecto hidrofóbico, se encuentra mayormente dirigida por una entalpía favorable. Por consiguiente, este carácter exotérmico de la interacción difícilmente puede justificar una interacción netamente hidrofóbica, dejando entrever que el funcionamiento de estos sencillos dominios oculta un grado de complejidad mayor, que aún no se conoce por completo y que es vital para entender y desarrollar terapias contra numerosas patologías en las que están implicados los dominios SH3[53].

La familia de quinasas Src está implicada en la regulación del desarrollo celular, activando una gran variedad de proteínas citosólicas, nucleares y de membrana mediante la fosforilación de un resto tirosina [59]. Está compuesta por nueve miembros, cSrc, cYes, Fyn, Lck, Hck, Lyn, Blk, Fgr y Yrk, los cuales comparten una arquitectura común que incluye un dominio catalítico, un dominio SH3 y otro SH2, estos dos últimos implicados en la regulación de la actividad quinasa del primero y en la incorporación de sustratos [50]. Dentro de la familia, los miembros con mayor homología de secuencia entre sus dominios SH3 son cSrc, cYes y Fyn, que comparten una identidad en la secuencia del 70% y una alta homología estructural [50]. No obstante, aunque el sitio de unión es una de las zonas más conservadas, como se aprecia en el panel B de la Figura 2.1, estos tres dominios han demostrado tener diferentes preferencias de unión tanto por ligandos peptídicos como por proteínas, tanto *in vitro* como *in vivo* [50, 60-62].

Con el objetivo de estudiar los mecanismos que determinan la afinidad y especificidad de unión en dominios SH3, el Dr. José Manuel Martín García realizó, en el seno de nuestro grupo de investigación, un extenso análisis termodinámico mediante calorimetría isotérmica de titulación (ITC) de la interacción de diversos ligandos naturales y de diseño con los dominios SH3

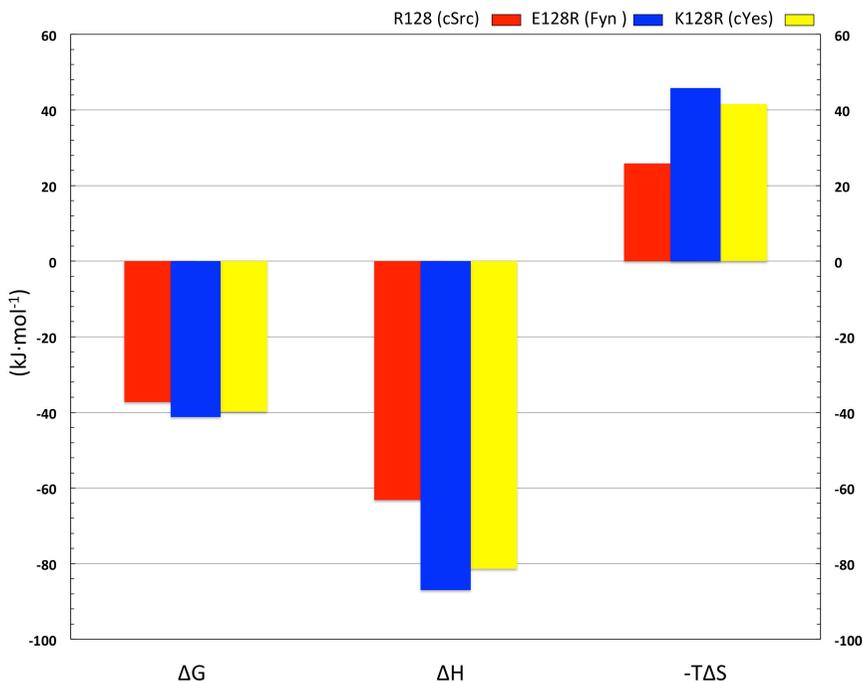
de cSrc, cYes y Fyn. Entre los péptidos estudiados se incluye el péptido con la secuencia VSLARRPLPPLP (VSL12), identificado mediante técnicas de expresión en fagos. VSL12 une a los tres dominios con orientación tipo I y hasta la fecha es uno de los ligandos que presenta mayor afinidad para cualquiera de ellos [63].



**Figura 2.1 Estructura y secuencia de los dominios SH3 de la familia de Src. Panel A:** Superposición de las estructuras del complejo c-Src-SH3/VSL12 (1qwf) y de las estructuras libres de los dominios SH3 de Fyn (1shf) y c-Yes (2hda) en representación de cintas y en colores rojo, azul y amarillo respectivamente. La estructura del ligando VSL12 se muestra en representación de varillas. **Panel B.** Alineamiento de secuencias de los dominios que componen la familia SH3 de Src. Los residuos se representan en un código de colores de acuerdo con el carácter del aminoácido. Se han resaltado con el símbolo + aquellas posiciones que forman el sitio de unión. La posición 128 ha sido resaltada con un asterisco rojo. En la parte inferior del alineamiento se representa esquemáticamente a que región del dominio pertenece cada posición

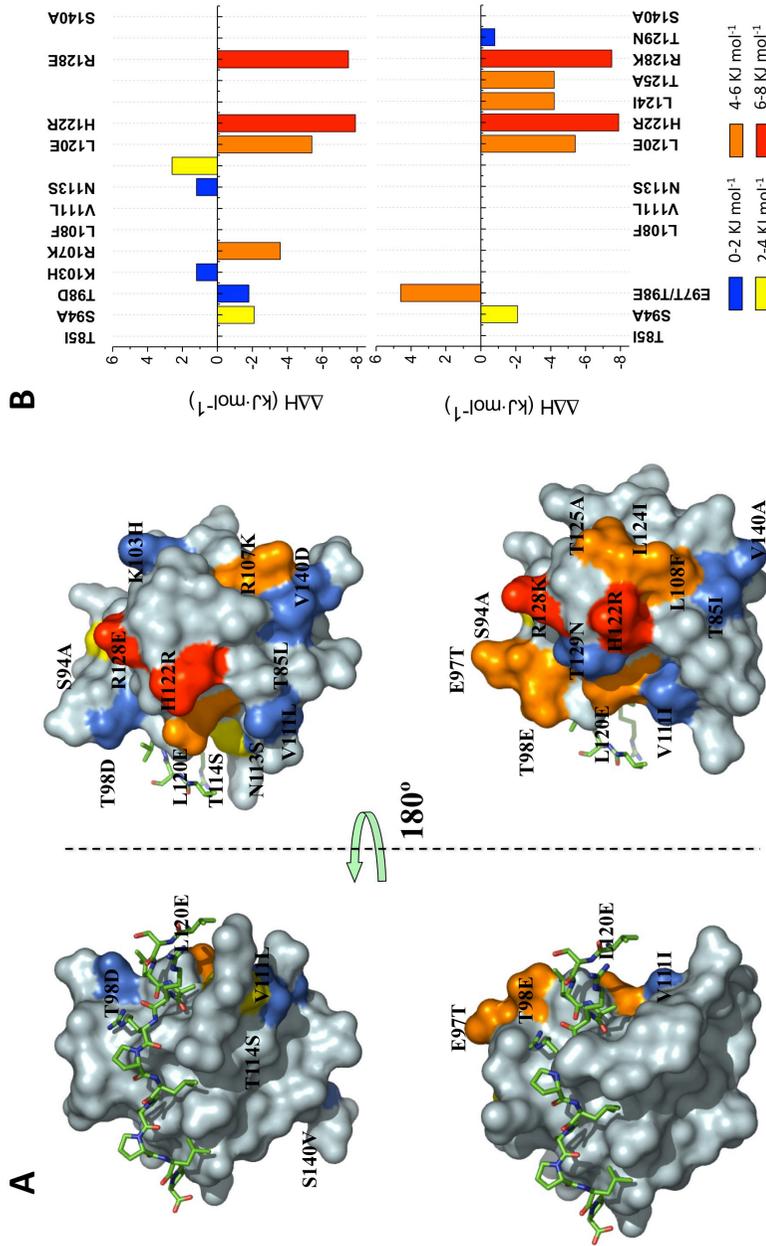
## CAPÍTULO 2

Sin embargo, el estudio termodinámico reveló una gran variabilidad en las contribuciones entálpica y entrópica para los tres dominios SH3 de la familia al unirse al péptido VSL12. Así, cSrc une a VSL12 con un cambio de entalpía muy inferior ( $20 \text{ kJ}\cdot\text{mol}^{-1}$  de diferencia) a los de Fyn y cYes (Figura 2.2). Este hecho llama la atención si tenemos en cuenta que los tres dominios sólo difieren en un residuo en la región de unión, como se puede observar en la Figura 2.1 panel B. Con la intención de localizar los residuos responsables de esta diferencia entálpica y así identificar las causas que están contribuyendo a la inesperada energética de unión característica de los dominios SH3, el Dr. José Manuel Martín García realizó un completo análisis mutacional sobre el dominio cSrc-SH3, substituyendo uno por uno aquellos residuos en los que éste difiere de cYes-SH3 y Fyn-SH3 por los respectivos de estos dominios y analizando los parámetros termodinámicos de la unión de estos mutantes al ligando VSL12. (figura 2.3 Panel B). Este estudio demostró que las mutaciones en los residuos cercanos a la zona de unión no provocan cambios significativos en los parámetros termodinámicos de unión. Por el contrario, los mayores efectos están sorprendentemente asociados a las mutaciones en el lazo distal, (H122R, R128E y R128K), llamado tradicionalmente así precisamente por ser el más distante del sitio de unión del dominio. Como se observa en el alineamiento de la Figura 2.1, la posición 122 se encuentra muy conservada dentro de la familia de Src, al contrario que la posición 128, donde encontramos una mayor variabilidad, ya que está ocupada por una lisina en cYes (R128K) y por un glutámico en Fyn (R128E). Llama especialmente la atención el comportamiento del mutante R128K, ya que la lisina comparte numerosas propiedades físico químicas con la arginina, por lo que no cabría esperar *a priori* un efecto de tanta magnitud en dicha substitución, contrariamente a lo observado en la Figura 2.3.



**Figura 2.2. Diferencias entálpicas de la unión de VSL12 a dominios SH3 de la familia Src.** Representación en forma de barras de los parámetros termodinámicos medidos por ITC a 25°C, pH 7 y tampon fosfato para la unión de VSL12 a los dominios c-Src (barras rojas), y Fyn (barras azules) y c-Yes (barras amarillas).

El cambio de entalpía que se obtiene en un experimento de ITC es una medida directa del balance neto en el número y/o la fuerza de las interacciones que se forman entre el complejo y la forma libre. Dado que tanto el sitio de unión como el ligando no han sido modificados en los complejos de VSL12 con R128K (cYes), R128E (Fyn) con respecto a R128-cSrc, la variación de entalpía observada no debería de estar relacionada con cambios en las interacciones directas establecidas en la interfaz entre el ligando y el dominio. Tampoco es de esperar efectos asociados a la oclusión diferencial de moléculas de agua ni a un comportamiento diferente del ligando para adoptar la conformación de hélice poliprolina II. Por tanto, la variación observada en los parámetros termodinámicos debe tener fundamentalmente su origen en la propagación cooperativa de los cambios locales en la posición 128.



**Figura 2.3 Estudio de Mutagénesis dirigida en cSrc y datos termodinámicos. Panel A:** superficies moleculares de los dominios Fyn-SH3 (panel superior) y c-Yes-SH3 (panel inferior) representados en la misma orientación, sobre las que se han señalado las mutaciones puntuales realizadas a partir de las diferencias con respecto al dominio c-Src-SH3 y coloreadas de acuerdo a los valores de la entalpía de unión representados en el panel B. El ligando VSL12 se ha representado como varillas en la misma orientación que aparece en el complejo con el dominio c-Src-SH3 para indicar la localización del sitio de unión. **Panel B:** Representación de los valores de entalpía de unión aparente en  $\text{kJ}\cdot\text{mol}^{-1}$  para la unión del ligando VSL12 a cada uno de los mutantes representados en el panel A

En este sentido, es importante señalar que estudios previos de resonancia magnética nuclear en los que se han analizado las propiedades dinámicas y el intercambio hidrógeno/deuterio del dominio SH3 de c-Src, tanto libre como unido al ligando de poliprolina RLP2(RALPPLPRY), indican que la unión del ligando provoca cambios significativos en la dinámica del esqueleto peptídico, en las longitudes de los puentes de hidrógeno y en los factores de protección al intercambio tanto en la vecindad del sitio de unión como en posiciones alejadas de éste, produciéndose una propagación efectiva de las interacciones del ligando a través del dominio SH3 [64, 65]. Esta cooperatividad estructural implica que cualquier cambio en las interacciones locales puede afectar al comportamiento global de la estructura y a la distribución de las poblaciones conformacionales, alterando sus propiedades al reconocer y unir ligandos [55, 56, 66]. Es, por tanto, nuestra hipótesis que las mutaciones introducidas en el lazo distal provocan una variación local de interacciones que se propagaría a lo largo del dominio y que podría estar desplazando sutilmente el equilibrio conformacional de cada mutante. Esto afectaría a la variación de entalpía de unión al producirse diferentes fenómenos de reorganización, que se manifestarían en las diferencias detectadas en los parámetros termodinámicos de la unión a VSL12.

Los experimentos de ITC proporcionan una caracterización macroscópica de la interacción, es decir, obtenemos la contribución total de todos los microestados y de los fenómenos que ocurren durante la unión [67-70]. Por tanto, mediante esta técnica no es posible elucidar con detalle el mecanismo mediante el cual el cambio de residuo en la posición 128 del dominio c-Src-SH3 contribuye a la variación de entalpía observada. En este contexto, se abordó la determinación de los cambios producidos en los factores de protección al intercambio desde el sitio de unión al lazo distal y los

## CAPÍTULO 2

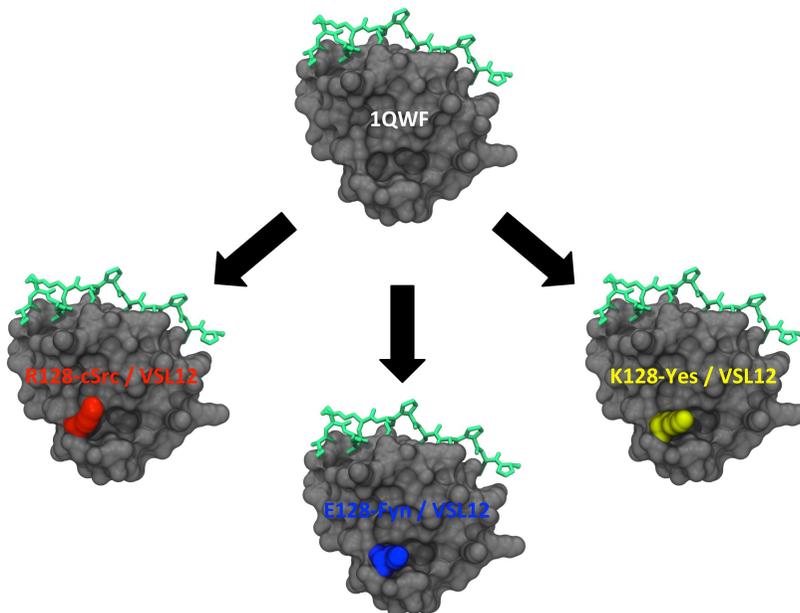
cambios en las longitudes de los puentes de hidrógeno tras la unión del ligando VSL12 mediante el estudio del intercambio hidrógeno/deuterio seguidos mediante resonancia magnética. Desgraciadamente, debido a la elevada afinidad mostrada por los diferentes complejos en las condiciones en las que se observan los mayores efectos en la energética de unión y también a la alta estabilidad del dominio, no nos ha sido posible realizar este estudio, ya que implicaría períodos de medida demasiado largos, incluso de algunos meses, como consecuencia de la baja velocidad de intercambio.

Por este motivo decidimos estudiar los efectos de las mutaciones R128K (cYes) y R128E (Fyn) en el dominio SH3 de cSrc mediante simulaciones de dinámica molecular, con la intención de explorar el sistema a nivel atómico, lo que nos permitirá evaluar las variaciones en la dinámica conformacional y hacer un seguimiento a nivel molecular de la dinámica de las interacciones diferenciales en cada mutante, con el fin de racionalizar las diferencias observadas en los parámetros energéticos [71-74].

### 2.2 RESULTADOS Y DISCUSIÓN

Con el objetivo de detectar la influencia de la mutación en la posición 128 en el dominio cSrc-SH3, se comparó el comportamiento estructural y de la red de interacciones del dominio, tanto en su forma libre como unida a VSL12, para los mutantes del dominio SH3-cSrc previamente estudiados por calorimetría R128, K128R y E128R, a lo largo de la trayectoria de dinámica molecular. Con la intención de tener presente en todo momento la procedencia del residuo introducido en la posición 128 de SH3-cSrc, se le añadió al nombre de cada constructo el dominio de origen del residuo introducido. De forma que la forma silvestre se denomina R128-cSrc, el

mutante con la lisina K128-cYes y finalmente E128-Fyn. Al no disponer de las estructuras para cada mutante, fue necesario modelar cada constructo, siguiendo el protocolo de la **sección XX** de Materiales y Métodos, usando como plantilla la estructura resuelta por resonancia magnética nuclear del dominio SH3 de cSrc unida a VSL12 (con código PDB 1QWF) [63]. Substituyendo solamente el residuo de la posición 128 se generaron los modelos R128-cSrc/VSL12, K128-cYes/VSL12 y E128-Fyn/VSL12, así como las formas sin ligando R128-cSrc, K128-cYes y E128-Fyn mediante la eliminación de la cadena correspondiente al ligando (FIGURA 2.4). Partiendo de estos modelos y siguiendo el protocolo de la **sección XX** de Materiales y Métodos, se generaron 30 ns de trayectoria de dinámica molecular para cada estructura a 298 K, y se analizaron mediante la herramienta ptraj que incorpora el paquete AMBER [75].



**Figura 2.4** Los modelos creados a partir del PDB de 1QWF. Superficies moleculares del dominio cSrc en gris, resaltando en rojo ARG en la posición 128, en azul el GLU en la posición 128 y la LYS en la posición 128. En verde y en esta resaltado el ligando VSL12.

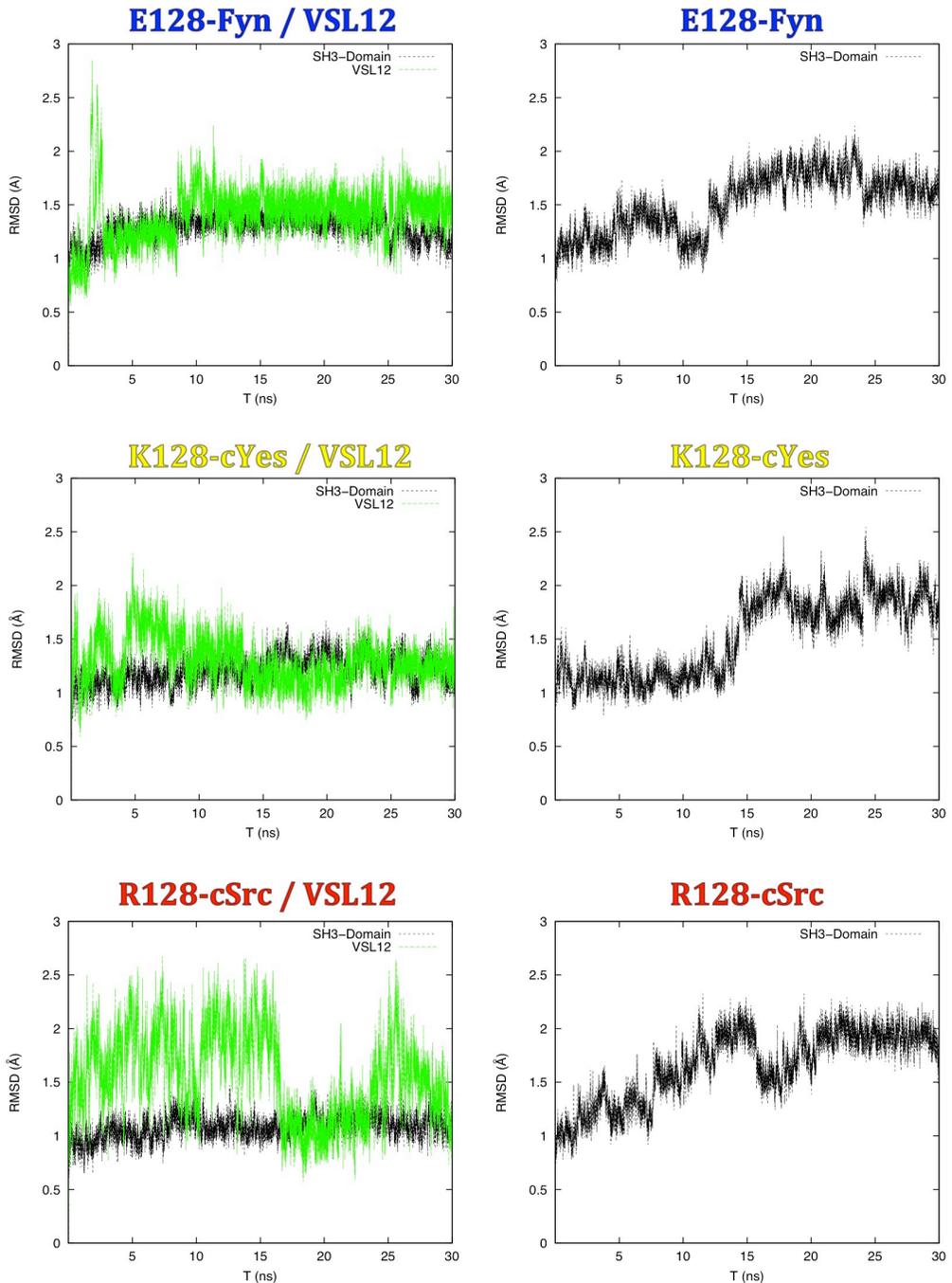
## CAPÍTULO 2

### 2.2.1 Análisis de la dinámica estructural

La desviación cuadrática media (RMSD) es la distancia promedio que existe entre los átomos después de superponer dos estructuras, lo que nos cuantifica cómo son de diferentes dos estructuras entre sí. Al representar la secuencia de valores que obtenemos al calcular el RMSD de los átomos del esqueleto peptídico para cada instante de la trayectoria, usando como referencia la primera estructura de la simulación, podemos observar cómo evoluciona la estructura a lo largo de la trayectoria, lo que permite detectar cambios conformacionales, y al mismo tiempo evaluar cómo responden las estructuras a las mutaciones introducidas [76, 77].

En la figura 2.5, se representan los valores de RMSD a lo largo de las diferentes trayectorias realizadas, representando con una línea negra los valores correspondientes al RMSD para el dominio, y con una línea verde el RMSD para el ligando VSL12. Como se puede observar, el RMSD para los complejos con VSL12 muestran una fluctuación estable del RMSD por debajo de 1 Å, de forma que podemos afirmar que las estructuras no han sufrido una desestabilización o grandes cambios conformacionales como consecuencia de las mutaciones.

Por otro lado, si comparamos los valores de RMSD entre las trayectorias con los dominios unidos a VSL12 respecto a las trayectorias con los dominios libres, observamos cómo en las últimas el RMSD es ligeramente mayor, fluctuando alrededor de los 2 Å. Estas diferencias son debidas al aumento de la libertad conformacional de los lazos RT y n-Src, como veremos con más detalle a continuación gracias al análisis del RMSF.

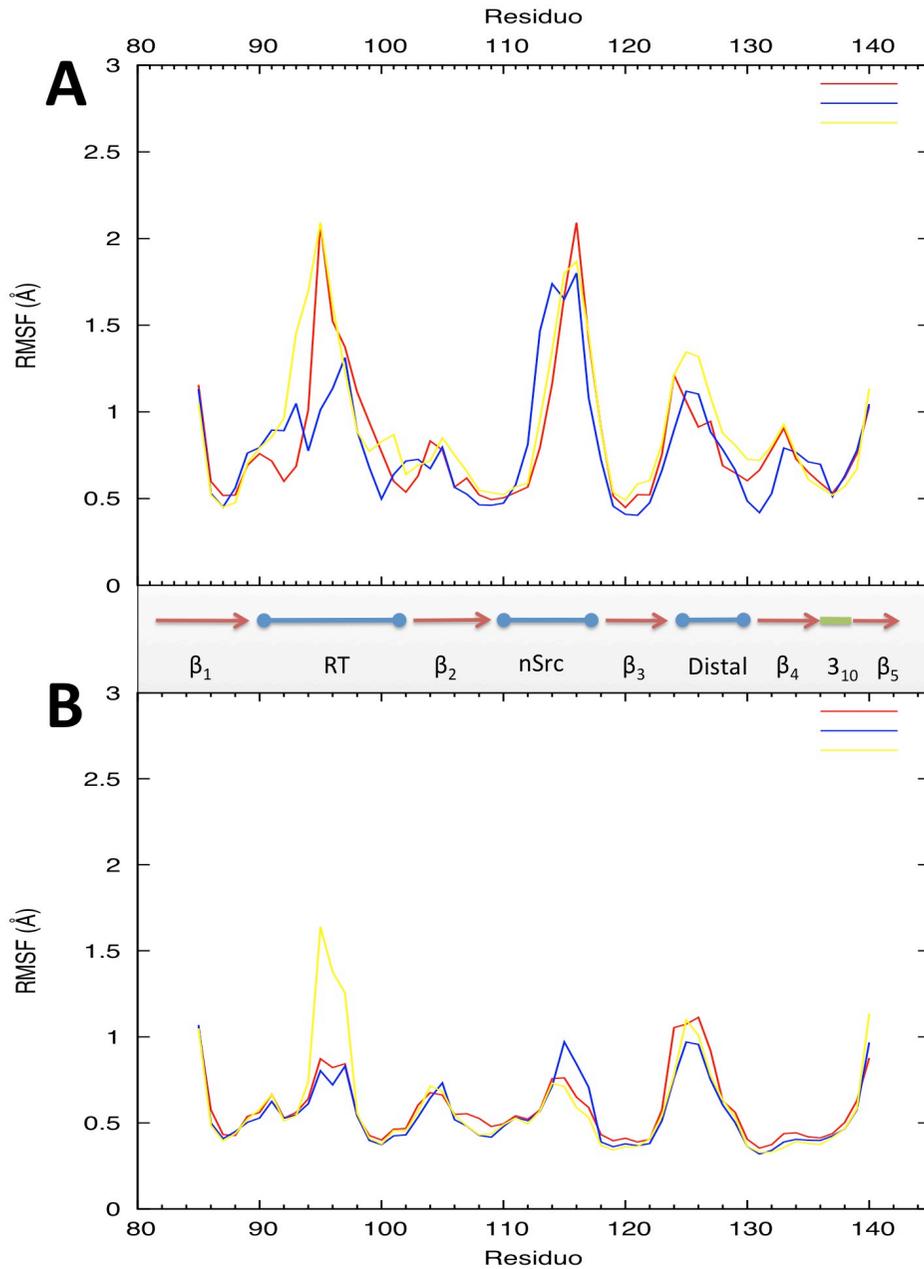


**Figura 2.5** Representación de la evolución de la desviación cuadrática media (RMSD) a lo largo de la simulación de dinámica molecular del esqueleto peptídico. En todos los casos se utilizó como referencia la primera estructura para el dominio (línea negra) y para el ligando (línea verde). En la parte superior se representan los valores registrados para las trayectorias del modelo E128-Fyn, en el panel central los valores para el modelo K128-cYes y finalmente en el panel inferior para E128-cSrc. En todos los casos, tanto para los complejos como para los modelos libres.

## CAPÍTULO 2

Finalmente, destaca el incremento y fluctuación de RMSD para el ligando VSL12 de la trayectoria R128-cSrc/VSL12. Aunque los valores de RMSD no son significativamente más altos que los registrados para el resto de complejos (aproximadamente 0.5 Å más) y no implicarían un cambio de conformación, el aumento en las fluctuaciones estructurales respecto a la conformación original podría indicar que ha ocurrido algún cambio adicional en la región del sitio de unión que no ocurre en los complejos E128-Fyn y K128-cYes,. Este cambio favorecería que el péptido, bien tenga mayor libertad conformacional, o bien se favorezca una conformación ligeramente diferente del ligando peptídico VSL12. Recordemos de nuevo que la única diferencia entre las estructuras se encuentra en la posición 128, que se encuentra a más de 10 Å del sitio de unión, con lo cual no era de esperar diferencias que afectaran directamente al ligando. Para analizar esta situación hemos de comprobar las interacciones intermoleculares.

Al calcular el RMSD en cada instante de la trayectoria usando la estructura promedio como referencia obtenemos la fluctuación cuadrática media (RMSF), que representa la oscilación de los átomos del esqueleto peptídico a lo largo de la simulación de dinámica molecular. Se deduce que su valor nos revelará qué zonas de la estructura presentan una mayor o menor movilidad, dato que nos puede ayudar a detectar variaciones en las propiedades dinámicas del esqueleto peptídico del dominio entre las distintas trayectorias [78]. La dinámica o flexibilidad de un residuo está íntimamente ligada a las relaciones que establece con su entorno, de forma que cambios en la flexibilidad nos pueden indicar cambios en la dinámica de las interacciones.



**Figura 2.6 Dinámica estructural de los dominios en complejo y libres.** Representación de la fluctuación cuadrática media (RMSF) del esqueleto peptídico para cada residuo durante las trayectorias de R128-cSrc (línea roja), K128-cYes (línea amarilla) y E128-Fyn (línea azul) para las formas libres (Panel A) y para los complejos con VSL12 (panel B). En la zona central se representa esquemáticamente la distribución de estructuras secundarias del dominio c-Src-SH3 a lo largo de la secuencia.

## CAPÍTULO 2

En el panel A de la Figura 2.6 se muestran los valores de RMSF para las trayectorias de los mutantes libres, mientras que en el panel B encontramos los valores para los mutantes unidos a VSL12. Al comparar los valores de RMSF entre los complejos y las formas libres, se observa cómo la flexibilidad de los mutantes, en especial las regiones de los lazos RT y n-Src, es mayor en las formas libres que en las formas unidas a VSL12. Este cambio en la flexibilidad ha sido previamente descrito en experimentos de resonancia magnética nuclear [65] y se debe a la liberación de las restricciones conformacionales impuestas por las interacciones establecidas entre los residuos de los lazos y los del ligando. Esta mayor flexibilidad de la región de los lazos sería la causante del aumento de RMSD en las trayectorias libres.

Por otro lado, si comparamos los valores de RMSF entre los constructos libres observamos que, sorprendentemente, las diferencias más significativas entre modelos no las encontramos en la región del lazo distal, sino que se concentran en parte del lazo RT. Durante las trayectorias de los modelos libres, el RMSF de la región del lazo RT es similar para los modelos con un residuo básico en la posición 128 (R128-cSrc y K128-cYes), en cambio la flexibilidad del lazo RT de E128-Fyn, con un resto ácido en la posición 128, se encuentra por debajo de los dos anteriores, en especial los residuos de SER94 a la SER101 con diferencias de RMSF que oscilan entre los 0.5 Å y los 1.3 Å. En el resto de posiciones la diferencia en RMSF entre los modelos es de 0.10 Å en promedio.

Al comparar los valores de RMSF obtenidos para las trayectorias de los modelos con el ligando VSL12 todas las trayectorias muestran una disminución de la flexibilidad en la región de los lazos, tal como hemos mencionado anteriormente, pero observamos que los valores de RMSF del lazo RT para el complejo de K128-cYes/VSL12, no se encuentran al nivel de

R128-cSrc y E128-Fyn. En concreto, esta diferencia de comportamiento afecta a los residuos SER94 , THR95 y ASP96 con diferencias de RMSF que oscilan entre 0.5 y 0.6 Å respecto a R128-cSrc y E128-Fyn.

En resumen, el análisis dinámico nos muestra que no existen grandes cambios conformacionales en el esqueleto peptídico de los modelos, pero sí ligeras diferencias de comportamiento dinámico del lazo RT, tanto en las formas unidas al ligando como en las libres. Sorprendentemente los cambios de RMSF no afectan a la región del lazo distal que es en la que se diferencian los modelos. Esto nos indica una transmisión de los efectos de las mutaciones en el lazo distal a través de la estructura del dominio, muy probablemente inducida por las cadenas laterales. Dado que el lazo RT está directamente implicado en la unión del ligando VSL12, será necesario comprobar si este cambio en la dinámica afecta a las interacciones intermoleculares.

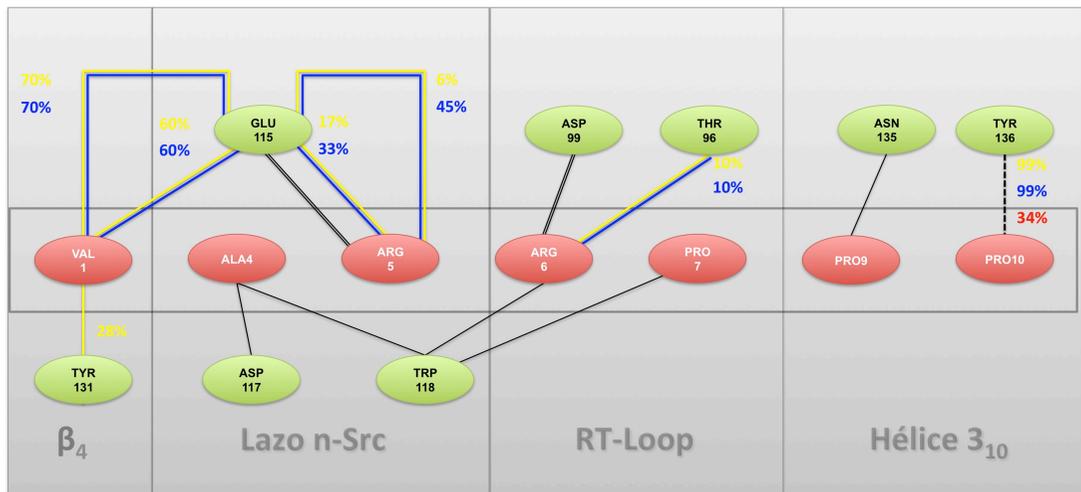
### **2.2.2 Análisis de las interacciones; enlaces de hidrógeno y puentes salinos**

Los resultados de los experimentos de ITC realizados previamente en nuestro grupo de investigación, revelaron unas diferencias muy notables, de hasta 20 kJ·mol<sup>-1</sup>, en las entalpías de unión de los complejos R128-cSrc/VSL12, E128-Fyn y K128-cYes. Dado que una entalpía de unión favorable refleja un balance positivo de interacciones establecidas entre el ligando y la proteína respecto a las que ambos establecen con el disolvente, es necesario comparar las interacciones de los modelos con y sin ligando para detectar el origen de las diferencias en entalpía observadas. Aunque a día de hoy no es posible cuantificar de forma precisa el aporte entálpico de cada interacción, sí que podemos buscar diferencias en número, tipo, frecuencia de formación según el resto que ocupe la posición 128 del lazo distal.

## CAPÍTULO 2

### 2.2.3 Exámen comparativo de las interacciones intermoleculares

Como hemos indicado anteriormente, no existe ninguna diferencia en las secuencias del ligando y del sitio de unión entre los diferentes modelos, con lo que en un principio no esperaríamos observar diferencias significativas en las interacciones que se establecen a lo largo de la trayectoria, tanto puentes de hidrógeno como puentes salinos.



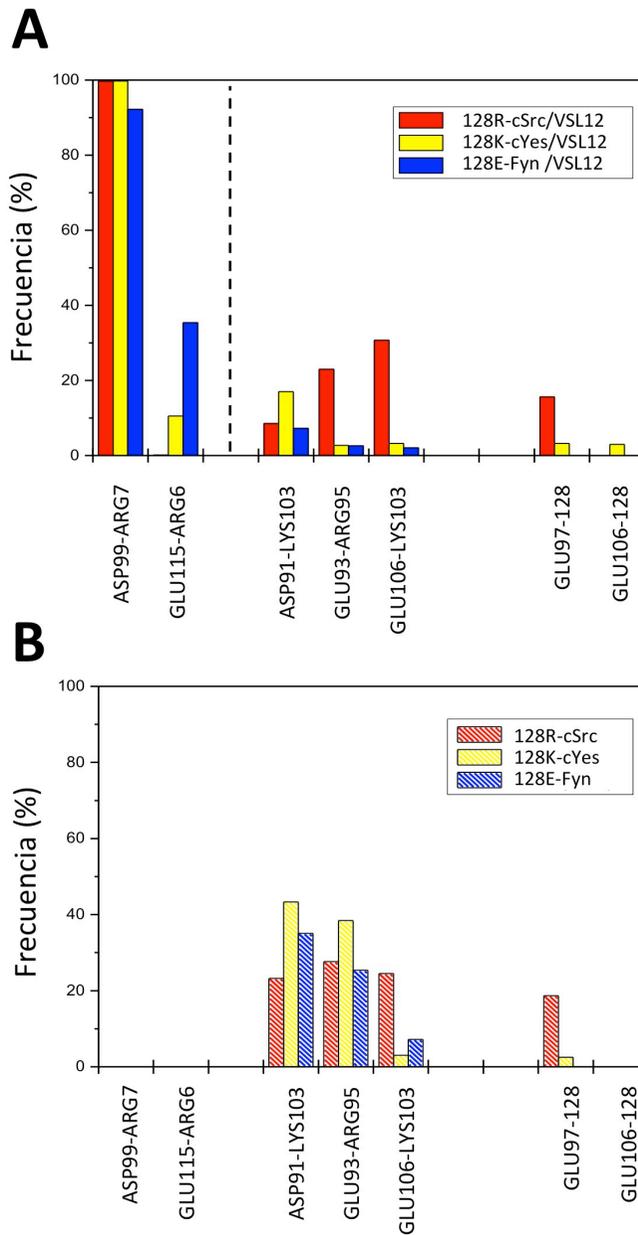
**Figura 2.6 Representación esquemática de los puentes de hidrógeno intermoleculares.** Los nodos rojos representan los residuos del ligando VSL12, mientras que los verdes los residuos del dominio. Las líneas negras continuas representan los puentes de hidrógeno que no varían entre modelos, las discontinuas los puentes de hidrógeno que varían en frecuencia entre modelos. Las líneas de color representan los puentes de hidrógeno que se establecen característicamente en cada modelo, en azul para E128-Fyn y en amarillo para K128-cYes

Mediante el análisis exhaustivo de las trayectorias de dinámica molecular se recopilaron las frecuencias de formación de los puentes de hidrógeno que se establecen entre el ligando y el dominio a lo largo de las trayectorias, considerando que se establece un puente de hidrógeno cuando el átomo donador y el aceptor se encuentran a una distancia inferior a 3.5 Å y formando un ángulo máximo de 90°. En la Figura 2.6, representamos los puentes de hidrógeno establecidos entre el dominio y el ligando que han

perdurado durante más del 5% del tiempo de simulación para las trayectorias de los tres complejos R128-cSrc/VSL12, E128-Fyn/VSL12 y K128-cYes/VSL12, donde, sorprendentemente, se pueden observar algunas diferencias significativas. Así, en la región de la hélice 3<sub>10</sub> el puente de hidrógeno entre los restos TYR136 del dominio y PRO10 del ligando se establece con una frecuencia 3 veces inferior en la trayectoria R128-cSrc/VSL12 respecto a los otros dos complejos. En el otro extremo del ligando un residuo de la lámina  $\beta$ 4 del dominio, el resto TYR131, establece un puente de hidrógeno con el resto VAL1 de VSL12 con una frecuencia del 28%, en la trayectoria del complejo K128-cYes/VSL12 pero no en las demás. En la región del lazo n-Src existen varios puentes de hidrógeno establecidos con elevada frecuencia en los complejos E128-Fyn/VSL12 y K128-cYes/VSL12 que no aparecen en la trayectoria correspondiente al complejo R128-cSrc/VSL12. En concreto, el puente de hidrógeno entre la cadena lateral del GLU115 del dominio y el esqueleto péptido de la VAL1 en el ligando, que se establece con una frecuencia de formación del 60% y 70% en K128-cYes/VSL12 y E128-Fyn respectivamente, y además, existe un segundo puente de hidrógeno, establecido entre la cadena lateral de la ARG5 del ligando y el GLU115 del dominio que se forma con una frecuencia 3 veces superior en E128-Fyn/VSL12 respecto a K128-cYes y está totalmente ausente en R128-cSrc/VSL12. Esto indica que el lazo n-Src de E128-Fyn y/o el ligando VSL12 se encuentran en una conformación que favorece la formación de este puente salino, mientras que en el complejo K128-cYes/VSL12, y especialmente en el R128-cSrc/VSL12, esta interacción no está tan favorecida. Finalmente, en la región del lazo RT se observa la formación de un puente de hidrógeno entre las posiciones THR96 y ARG6 de VSL12, que sólo se establece en los complejos E128-Fyn/VSL12 y K128-CYes/VSL12, aunque con una baja frecuencia de formación (un 10% del tiempo de simulación).

**Tabla 2.1** Detalle de las frecuencias de formación de los puentes de hidrógeno intermoleculares para cada modelo.

R128-cSrc			K128-cYes			128E-Fyn		
Aceptor	Hidrógeno	Donor	Aceptor	Hidrógeno	Donor	Aceptor	Hidrógeno	Donor
% Ocupancia			% Ocupancia			% Ocupancia		
			GLU115:OE2	VAL1@H	VAL1@N	GLU115:OE1	VAL1@H	VAL1@N
			GLU115:OE1	VAL1@H	VAL1@N	GLU115:OE2	VAL1@H	VAL1@N
			TYR131:OH	VAL1@H	VAL1@N			
ALA4@O	TRP118:H	TRP118:N	ALA4@O	TRP118:H	TRP118:N	ALA4@O	TRP118:H	TRP118:N
ALA4@O	ASP117:H	ASP117:N	ALA4@O	ASP117:H	ASP117:N	ALA4@O	ASP117:H	ASP117:N
GLU115:O	ARG5@HE	ARG5@NE	GLU115:O	ARG5@HE	ARG5@NE	GLU115:O	ARG5@HE	ARG5@NE
GLU115:O	ARG5@HH21	ARG5@NH2	GLU115:O	ARG5@HH21	ARG5@NH2	GLU115:O	ARG5@HH21	ARG5@NH2
GLU115:O	ARG5@HH11	ARG5@NH1	GLU115:O	ARG5@HH11	ARG5@NH1	GLU115:O	ARG5@HH11	ARG5@NH1
			GLU115:OE1	ARG5@HH21	ARG5@NH2	GLU115:OE1	ARG5@HH21	ARG5@NH2
			GLU115:OE2	ARG5@HH21	ARG5@NH2	GLU115:OE2	ARG5@HH21	ARG5@NH2
			GLU115:OE1	ARG5@HE	ARG5@NE	GLU115:OE1	ARG5@HE	ARG5@NE
			GLU115:OE2	ARG5@HE	ARG5@NE	GLU115:OE2	ARG5@HE	ARG5@NE
ASP99:OD1	ARG6@HH12	ARG6@NH1	ASP99:OD1	ARG6@HH12	ARG6@NH1	ASP99:OD1	ARG6@HH12	ARG6@NH1
ASP99:OD2	ARG6@HH22	ARG6@NH2	ASP99:OD2	ARG6@HH22	ARG6@NH2	ASP99:OD2	ARG6@HH22	ARG6@NH2
ASP99:OD2	ARG6@HH12	ARG6@NH1	ASP99:OD2	ARG6@HH12	ARG6@NH1	ASP99:OD2	ARG6@HH12	ARG6@NH1
ASP99:OD1	ARG6@HH22	ARG6@NH2	ASP99:OD1	ARG6@HH22	ARG6@NH2	ASP99:OD1	ARG6@HH22	ARG6@NH2
ARG6@O	TRP118:HE1	TRP118:NE1	ARG6@O	TRP118:HE1	TRP118:NE1	ARG6@O	TRP118:HE1	TRP118:NE1
			THR96:OG1	ARG6@HH12	ARG6@NH1	THR96:OG1	ARG6@HH12	ARG6@NH1
PRO7@O	TRP118:HE1	TRP118:NE1	PRO7@O	TRP118:HE1	TRP118:NE1	PRO7@O	TRP118:HE1	TRP118:NE1
PRO9@O	ASN135:HD21	ASN135:ND2	PRO9@O	ASN135:HD21	ASN135:ND2	PRO9@O	ASN135:HD21	ASN135:ND2
PRO10@O	TYR136:HH	TYR136:OH	PRO10@O	TYR136:HH	TYR136:OH	PRO10@O	TYR136:HH	TYR136:OH



**Figura 2.7** Frecuencias de formación de los puentes salinos durante las trayectorias de los mutantes E128-Fyn, K128-cYes y R128-cSrc libres y unidos a VSL12. Considerando la frecuencia como número de imágenes donde se observa formado el puente dividido por el número total de imágenes de la dinámica y multiplicado 100. Cada imagen de simulación corresponde a 1ps. Se ha considerado que el puente salino estaba formado cuando el centro de masas del oxígeno del residuo ácido estaba localizado a menos de 4 Å del centro de masas del nitrógeno del residuo básico. En el Panel A se representan los mutantes unidos a VSL12 y en el Panel B los libres. La línea discontinua distingue los puentes salinos que se producen entre el ligando y el dominio del resto

## CAPÍTULO 2

En la Figura 2.7 hemos representado la frecuencia de formación de los puentes salinos que se forman durante las simulaciones de los complejos con VSL12 (Panel A) y de los dominios libres (Panel B). Consideramos que se establece un puente salino siempre que el centro de masas de la cadena lateral de un resto ácido se encuentra a menos de 4 Å del centro de masas de la cadena lateral de un resto básico [79]. Si nos centramos en los puentes salinos establecidos entre el ligando y el dominio (panel A) destaca, por su constante frecuencia de formación, el puente salino entre el resto ASP99 (situado en el lazo RT) y la ARG6 del ligando VSL12 que se encuentra casi permanentemente formado (frecuencia de formación próxima al 100%) en los tres modelos. Este puente salino ha sido previamente descrito en la literatura y, como se describió en la Introducción, la posición de la arginina con respecto al motivo canónico de unión determina la orientación tipo I o II de los péptidos ricos en prolina. Por otro lado, se observa en los complejos E128-Fyn/VSL12 y K128-cYes/VSL12 la formación de un segundo puente salino entre el ligando y el dominio, en concreto entre el resto ARG5 de VSL12 y el GLU115 del dominio (lazo n-Src). Este puente salino, ausente en R128-cSrc/VSL12, se establece con una frecuencia 4 veces mayor en E128-Fyn/VSL12 que en K128-cYes/VSL12 y facilitaría la formación de los puentes de hidrógeno, descritos en la sección anterior, entre el dominio y el ligando.

Es decir, que aunque no existe ningún cambio en la secuencia del ligando ni en el sitio de unión del dominio para cada modelo, sí que detectamos que K128-cYes y E128-Fyn establecen un mayor número de puentes de hidrógeno con el ligando que R128-cSrc. El establecimiento de puentes de hidrógeno u otras interacciones polares, aunque supone una ganancia entálpica, suele estar acompañado de una penalización entrópica debido a la reducción de la libertad conformacional del sistema [80, 81]. Este

hecho estaría en consonancia con los parámetros termodinámicos observados para K128-cYes y E128-Fyn, que, a pesar de mostrar una mayor entalpía al unir VSL12 que R128-cSrc, también presentan una penalización entrópica mayor (FIGURA 2.2). En cualquier caso, los cambios en el patrón de interacciones intermoleculares detectados, son difíciles de cuantificar en términos energéticos, por lo que es difícil establecer el grado en que son los responsables del considerable cambio entálpico medido mediante ITC.

Además, recordemos que en la sección anterior hemos descrito cómo el cambio de comportamiento dinámico afecta principalmente al lazo RT, aunque las variaciones en las interacciones intermoleculares se reparten a lo largo de todo el sitio de unión, siendo especialmente llamativas en la región del lazo n-Src. Estos cambios son difícilmente justificables en términos de interacciones locales, lo que indica que las mutaciones en la posición 128 tienen un efecto global sobre el dominio, que puede llegar incluso a afectar a los residuos que componen el sitio de unión del dominio SH3.

#### **2.2.4 Análisis comparativo de las interacciones intramoleculares**

La presencia de un puente salino entre los restos ARG128 y GLU96 del lazo RT en el dominio SH3 de alfa espectrina (SH3-spc) sugiere que esta misma interacción podría estar estableciéndose en el dominio SH3 de cSrc, influyendo en la dinámica conformacional y en las interacciones intermoleculares e intramoleculares del lazo RT, y pudiendo ser, en consecuencia, una de las causas implicadas en el cambio de comportamiento dinámico entre los distintos mutantes descrito en la sección anterior. Por un lado, podemos agrupar los modelos correspondientes a R128-cSrc y K128-cYes, donde la posición 128 es ocupada por una carga positiva, mientras que, por otro lado, estaría el complejo E128-Fyn, donde se ha añadido una carga

## CAPÍTULO 2

negativa. *A priori*, la formación del puente salino observado en el dominio SH3 de alfa espectrina sería posible en los dos primeros complejos pero no en el tercero. Por otro lado, la presencia de una carga negativa en E128-Fyn podría influir en otros puentes salinos del resto del dominio, alterando la flexibilidad del esqueleto peptídico e influyendo en el equilibrio conformacional [82]. Para evaluar estas posibilidades, es necesario analizar y comparar todos los puentes salinos que se establecen intramolecularmente para cada complejo.

Siguiendo el mismo criterio que definimos en la sección anterior, en la Figura 2.7 se representan las frecuencias de formación de todos los puentes salinos intermoleculares, tanto para los complejos (Panel A), como para los dominios libres (Panel B). En el mutante R128-cSrc, observamos que el puente salino equivalente al descrito para el dominio SH3 de espectrina se establece entre las posiciones ARG128 y GLU97 durante un 20% del tiempo de simulación, tanto para el complejo con VSL12 como para el dominio libre. Sorprendentemente, aún compartiendo la misma carga, este mismo puente salino se constituye en K128-cYes sólo durante un 7%, tanto en el complejo como en el dominio libre.

El resto de los puentes salinos intramoleculares están fundamentalmente localizados en el lazo RT, (GLU106-LYS103, ASP91-LYS103, GLU93-ARG95). No obstante, al contrario que para E128-Fyn y K128-cYes, en R128-cSrc la frecuencia de formación de estos puentes salinos no varía entre el complejo y el dominio libre. En los complejos K128-cYes/VSL12 y E128-Fyn/VSL12 la frecuencia de formación de los puentes salinos (ASP91-LYS103 y GLU93-ARG95) disminuye hasta valores inferiores al 10%. Por otro lado, el puente salino GLU106-LYS103 se establece con una frecuencia entre el 5% y el 10% en K128-cYes y E128-Fyn, mientras que en R128-cSrc se

establece un 40% y un 30%, en el complejo y en la forma libre, respectivamente.

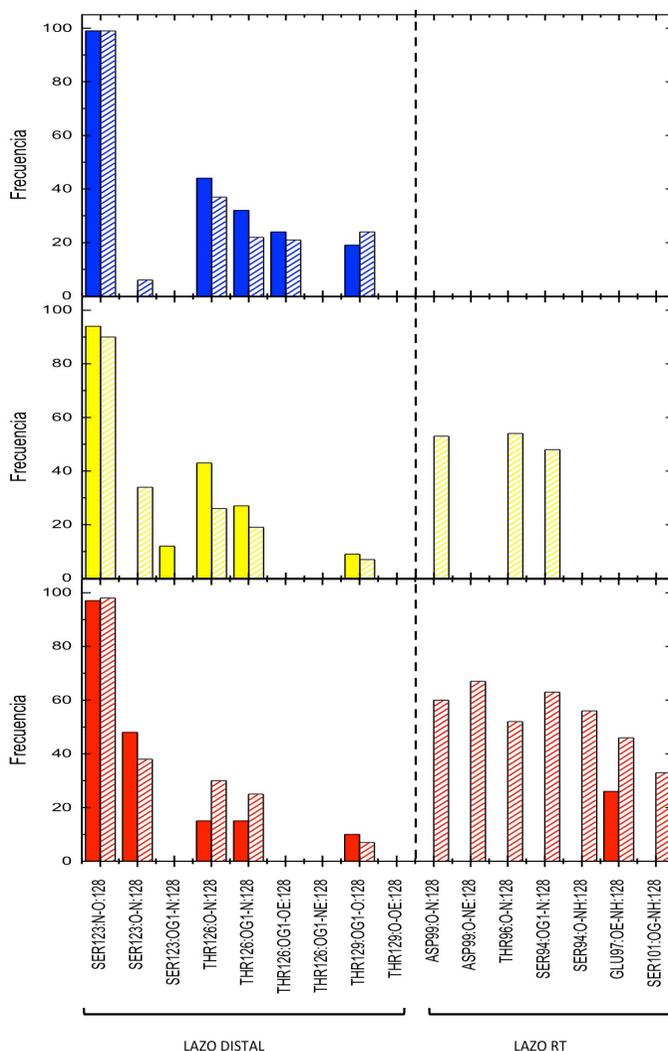
Al igual que sucedía con los puentes de hidrógeno, los puentes salinos muestran cómo los modelos de K128-cYes y E128-Fyn tienen un comportamiento similar entre sí, mientras que R128-cSrc muestra una configuración propia. Es decir, sorprendentemente, K128-cYes y R128-cSrc, aún teniendo ambos un residuo básico cargado positivamente en la posición 128, difieren en cuanto a la distribución y comportamiento de los puentes salinos. Estas diferencias de comportamiento en el lazo RT podrían estar originadas por la formación del puente salino entre la ARG128 y el GLU97 (situado en el lazo RT), el puente salino equivalente al descrito para el dominio SH3 de alfa espectrina, que solamente se establecería con frecuencia inferior al 5% en K128-cYes/VSL12, mientras que en R128-cSrc/VSL12 con una frecuencia de formación del 20%. Esta diferencia entre ambos restos básicos se debería a que la arginina, además de tener una mayor longitud, tiende a establecer un mayor número de interacciones electrostáticas y con mayor fuerza que la lisina [83], lo que explicaría por qué K128-cYes posee un comportamiento más próximo a E128-Fyn que R128-cSrc.

#### **2.2.4.1 Estudio de los puentes de hidrógeno de la posición 128.**

En definitiva, la única diferencia entre los modelos reside en la posición 128, de forma que todos los cambios detectados, es decir, el cambio del comportamiento dinámico del lazo RT y el mayor número de puentes de hidrógeno intermoleculares que se establecen en E128-Fyn y K128-cYes, deben tener su origen en la naturaleza del residuo que ocupa esa posición. En la Figura 2.8 hemos representado la frecuencia de formación de los puentes de hidrógeno que forma el residuo en la posición 128 a lo largo de las

## CAPÍTULO 2

trayectorias, tanto para los modelos unidos a VSL12 (barras sólidas), como para los libres (barras con trama). Siguiendo el mismo criterio usado para los puentes de hidrógeno intermoleculares, se recopilaron las frecuencias de formación de los puentes de hidrógeno.



**Figura 2.8 Frecuencia de formación para los puentes de hidrógeno que forma el residuo en la posición 128 en las diferentes trayectorias.** En rojo las trayectorias del dominio c-Src-SH3, en amarillo las del mutante R128K y en azul las del mutante R128E. Las barras en color sólido representan las trayectorias de los dominios en presencia de VSL12, y las rayadas las trayectorias en ausencia del ligando. La línea discontinua a través de los tres paneles separa las interacciones que se producen con residuos del lazo distal de las que se producen con los del lazo RT

En el panel superior se recopilan los puentes de hidrógeno para E128-Fyn, donde observamos que los puentes de hidrógeno establecidos por el glutámico se limitan a los residuos vecinos del lazo distal (SER123, THR126 y THR129) y que aparentemente no se ven afectados por la presencia o ausencia del ligando. En el segundo panel tenemos los puentes de hidrógeno que establece la lisina en la posición 128 (K128-cYes). En la trayectoria para el dominio libre se observa una situación muy similar a la descrita para E128-Fyn, aunque la lisina en la posición 128 establece algunos puentes de hidrógeno adicionales con los restos SER94, el ASP99 y la THR96 situados en el lazo RT. En el complejo K128-cYes/VSL12 todos estos puentes de hidrógeno con los restos del lazo RT dejan de establecerse y la lisina en la posición 128 no establece puentes de hidrógeno con el GLU97, ni en el complejo ni en la forma libre.

En el panel inferior de la Figura 2.8 se representan las frecuencias de los puentes de hidrógeno establecidos por la arginina en la posición 128 (R128-cSrc). En la forma libre observamos que, al igual que la lisina de K128-cYes, la arginina establece puentes de hidrógeno con los restos SER94, ASP99 y la THR96 del lazo RT, pero además forma también interacciones con la SER101 y el GLU97. Este último puente de hidrógeno es el único de los que establece la arginina con el lazo RT que se mantiene en la forma unida al ligando VSL12, aunque con menor frecuencia. Además, recordemos que los restos implicados en este puente de hidrógeno son los que establecen el puente salino que hemos descrito en la sección anterior y que sería el equivalente al observado para el dominio SH3 de alfa espectrina.

En resumen, en los constructos libres los enlaces de hidrógeno entre el lazo distal y el lazo RT, muy frecuentes en R128-cSrc, se ven significativamente reducidos en el modelo K128-cYes y desaparecen por

## CAPÍTULO 2

completo en E128-Fyn. Estas interacciones entre el lazo distal y el lazo RT en los modelos con un residuo básico en la posición 128 explicaría el mayor RMSF en esta región observado en la Sección 2.2.1. En cambio, en los complejos E128-Fyn/VSL12 y K128-cYes/VSL12 no existe ningún tipo de interacción entre el lazo distal y el lazo RT, mientras que en R128-cSrc las interacciones electrostáticas no se rompen del todo. Este hecho debe estar relacionado con que en este último complejo, R128-cSrc, no se forme el puente salino entre el resto GLU115 del lazo n-Src y la ARG5 del ligando.

### **2.2.4.2 Mapa de los puentes de hidrógeno del dominio SH3. Conectando la posición 128 con el sitio de unión**

Con la intención de observar el impacto global de los cambios locales descritos anteriormente y de encontrar cómo están conectados los cambios de la posición 128 con los observados en el sitio de unión, se recogieron todos los puentes de hidrógeno intramoleculares que se establecen durante las trayectorias para confeccionar un mapa que nos permitiera detectar cambios en la organización de la red de puentes de hidrógeno del dominio. Estos datos revelaron la existencia de un gran número de puentes de hidrógeno que se formaban en los complejos K128-cYes/VSL12 y E128-Fyn/VSL12, que estaban ausentes en el complejo R128-cSrc/VSL12 y viceversa. También se observó un cambio en las frecuencias de formación de numerosos puentes de hidrógeno, diferenciándose principalmente en el caso del complejo R128-cSrc/VSL12. Con el objetivo de representar lo más claramente posible estas diferencias, se muestran sólo los puentes de hidrógeno que presentan una diferencia significativa en la frecuencia de formación sobre un mapa esquemático del dominio SH3. En la Figura 2.9 observamos cómo el complejo R128-cSrc/VSL12 presenta un patrón propio de enlaces de hidrógeno, especialmente en la configuración de los lazos distal y RT. Esta

reconfiguración del lazo RT respecto a E128-Fyn/VSL12 y K128-cYes/VSL12 está inducida por medio de las interacciones que establecen ARG128 y GLU97, afectando a la configuración de casi todos los residuos que conforman el lazo RT, la misma región en la que observamos los cambios de comportamiento dinámico en la Sección 2.2.1. Además, podemos observar cómo el efecto de la reconfiguración en los puentes de hidrógeno se extiende por todo el dominio, llegando a afectar algunos puentes de hidrógeno de los residuos GLU115 y ASP117 del lazo n-Src, o a la TYR136 de la hélice 3<sub>10</sub>, todos ellos implicados en interacciones directas con el ligando. Esta reconfiguración de las interacciones del lazo RT, puede resultar, además, en una acomodación diferencial del péptido VSL12, favoreciendo la formación del puente salino entre GLU115 y la ARG5 en los complejos K128-cYes/VSL12 y E128-Fyn/VSL12.

En definitiva, los resultados del análisis de dinámica molecular confirman que, en efecto, como habíamos propuesto a partir de los resultados experimentales, las mutaciones en la posición 128 del lazo distal (R128E y R128K) inducen una reorganización de los puentes de hidrógeno que se propaga a través del dominio y que llega incluso a afectar a las interacciones que establece el dominio con el ligando. Es razonable considerar que dicha reorganización contribuya de forma importante a las notables diferencias en la energética de unión (en torno a los 20 kJ·mol<sup>-1</sup> en entalpía entre los complejos c-Src, Fyn y c-Yes). Esto indica que el reconocimiento de ligandos por parte de los dominios SH3 esconde una complejidad mayor que la descrita tradicionalmente, donde solamente se tenía en consideración la naturaleza hidrofóbica de la superficie de reconocimiento, jugando un papel determinante factores como el equilibrio conformacional del ligando [REF] y del propio dominio SH3, así como la reorganización inducida en las

## CAPÍTULO 2

distribuciones conformacionales de ambos como consecuencia de la interacción.

Los fenómenos de cooperatividad y restructuración de interacciones responderían a la necesidad de estos dominios modulares de tener un amplio rango de especificidad manteniendo una baja variabilidad estructural. Además, como se describirá en el Capítulo 4 de esta Memoria, en el contexto de las tirosinas quinasas completas existe una alta cooperatividad, ya no solo intradominio sino también interdominio, necesaria para su correcto funcionamiento de la proteína. De esta forma, es razonable pensar que los procesos de reconocimiento de sustrato, fundamentales para los mecanismos de activación de las quinasas con dominios modulares, no estén regulados solamente por los residuos presentes en la superficie de interacción, existiendo una influencia considerable de otras regiones del dominio. Estos resultados ponen en evidencia, por tanto, la necesidad del estudio profundo, no sólo de la superficie de unión, sino del conjunto del sistema para llegar a esclarecer el funcionamiento de estos dominios y así conseguir diseñar ligandos que posean una alta afinidad y especificidad y que abran las puertas al desarrollo de fármacos efectivos para el tratamiento de las patologías en las que estos dominios están implicados

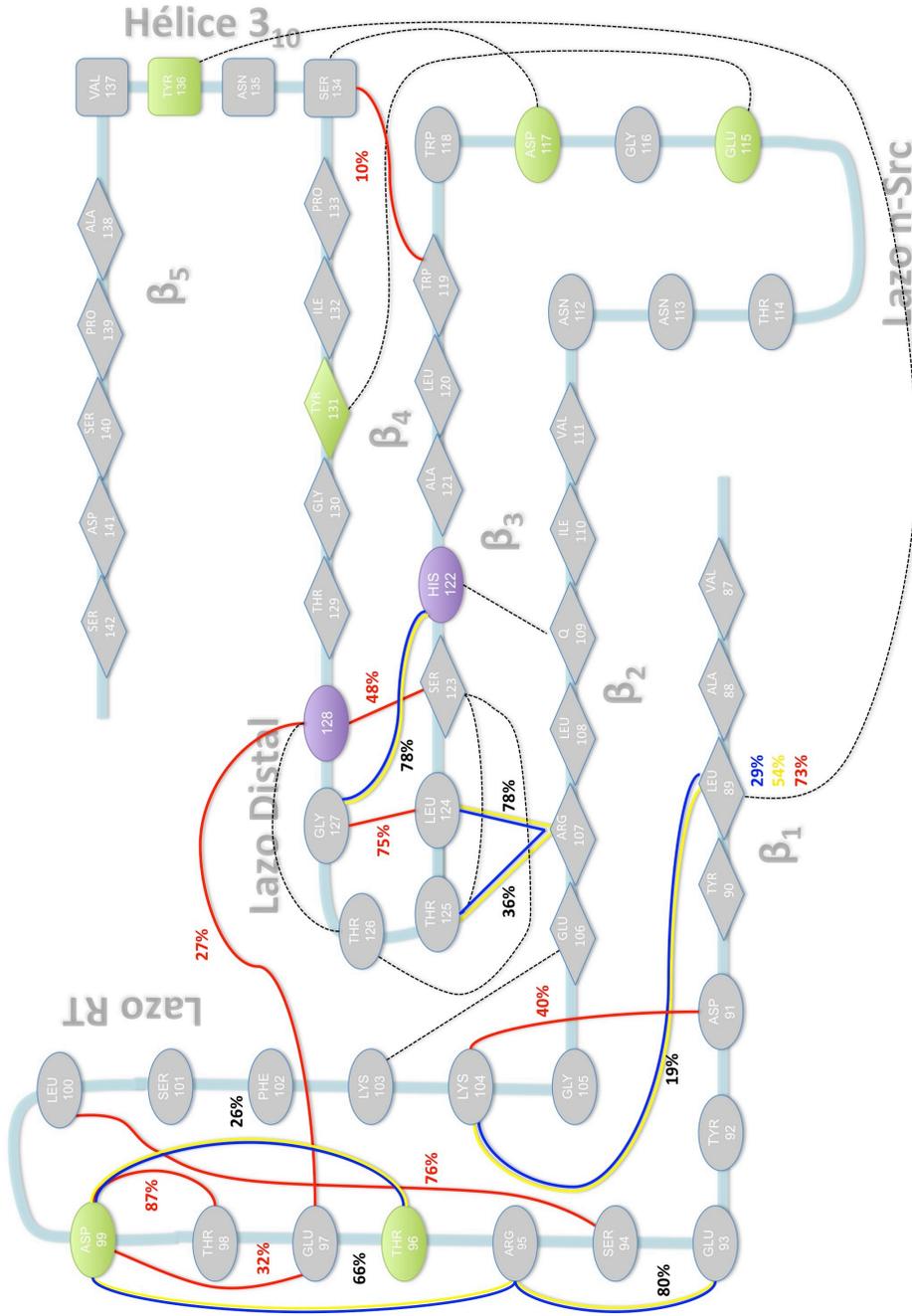


Figura 9. Representación de las diferencias entre puentes de hidrógeno que aparecen intradominio en las trayectorias de los tres mutantes de c-Src-SH3 unidos a VSL12. Cada nodo representa un residuo y cada línea que los une un puente de hidrógeno que se establece entre dos de esos residuos. Sólomente se representaron las diferencias más significativas de trayectorias. La línea negra punteada es para los puentes de hidrógeno presentes en las tres trayectorias pero que su frecuencia de formación es diferente en mas de un 20%. Se resaltaron los puentes de hidrógeno que solo están presentes en una de las trayectorias coloreándolas en amarillo, azul y rojo, para K128-Yes, E128-Fyn y R128-cSrc respectivamente. Resaltados en verde tenemos los residuos que interactúan con el ligando VSL12. En púrpura están resaltadas las posiciones 128 y 122.

### 2.3 CONCLUSIONES

Del análisis termodinámico, estructural y dinámico llevado a cabo en este capítulo se extraen las siguientes conclusiones:

- 1- Mediante las simulaciones de dinámica molecular de los modelos de R128-cSrc, K128-cYes y E128-Fyn libres y unidos a VSL12, hemos podido observar cómo R128c-Src muestra una red de interacciones propia, tanto intramoleculares como intermoleculares, en concordancia con los resultados termodinámicos.
- 2- Mediante el análisis detallado de las interacciones que se establecen durante las diferentes trayectorias hemos podido comprobar que las diferencias entre modelos tienen su origen principalmente en el lazo distal, más concretamente la posición 128, y se extienden a todo el dominio SH3 llegando a afectar a los restos que componen el sitio de unión que se encuentra a 10 Å del lazo distal.



---

**3 ZAHORÍ: UN ALGORITMO PARA EL ANÁLISIS  
AUTOMATIZADO DE HIDRATACIÓN INTERFACIALES  
MEDIANTE DINÁMICA MOLECULAR**



## CAPÍTULO 3

### 3.1 INTRODUCCIÓN

El agua y sus propiedades únicas juegan un papel importante en todos los procesos biológicos, no sólo como medio donde se producen procesos esenciales para la vida sino también como elemento activo en todas las reacciones bioquímicas, siendo fundamental, por ejemplo, para el plegamiento de las proteínas, las reacciones enzimáticas y la formación de complejos macromoleculares [84-88]. En las últimas décadas, la presencia de moléculas de agua en las interfaces de unión biomoleculares ha planteado múltiples interrogantes, motivando numerosos estudios para conocer con mayor detalle el papel que juegan en el proceso de reconocimiento molecular [89-91], especialmente en las interacciones transitorias y en el reconocimiento promiscuo, fenómenos típicos de la transducción de señales [92]. Los abundantes datos teóricos y experimentales indican que las moléculas de agua ocluidas en las interfaces de unión actuarían como adaptadores, rellenando espacios en la superficie de interacción y optimizando las interacciones de van der Waals y de puentes de hidrógeno, lo que contribuirá favorablemente a la entalpía de unión y permitirá cierta variabilidad en los sustratos para la misma interfaz.

Como hemos mencionado en la Introducción de esta Tesis, debido a la naturaleza principalmente hidrofóbica de los residuos implicados en la unión de dominios SH3 a motivos ricos en prolina, la energía libre de Gibbs de unión debería estar favorecida principalmente por una entropía favorable. Sin embargo, los parámetros termodinámicos medidos por calorimetría isotérmica de titulación indican todo lo contrario: la energía libre de Gibbs se caracteriza por una entalpía favorable y compensada por una ligera penalización entrópica.

Los ligandos canónicos de los dominios SH3 presentan un elevado número de restos prolina y, por lo tanto, se encuentran preformando la conformación PPII para acomodarse al sitio de unión [51], de forma que el origen de la penalización entrópica debe estar relacionado con otros procesos. La entalpía favorable nos indica que, en el proceso de unión, el balance neto de interacciones de las formas libres del dominio y del ligando respecto al complejo es positivo, de forma que además de la necesidad de adoptar la hélice PPII la unión de motivos ricos en prolina debe implicar la formación de nuevas interacciones. Este hecho deja entrever la existencia de un segundo mecanismo en el proceso de reconocimiento de ligandos poliprolina por dominios SH3 que sería el responsable del marcado carácter entálpico de la unión y de la penalización entrópica.

Inicialmente, con la intención de esclarecer la atípica huella termodinámica, los estudios se enfocaron en la dinámica conformacional del dominio, pero en el seno de nuestro grupo de investigación se propuso que, además de cambios en la dinámica conformacional, la desolvatación parcial y la mediación de interacciones por moléculas de agua ocluidas en el sitio de unión podrían contribuir a la huella termodinámica. Así, desde un punto de vista termodinámico, la desolvatación incompleta de la superficie de interacción ligando-proteína implica una penalización entrópica que compensa parcialmente los efectos entálpicos favorables, de modo que la influencia de estas moléculas de agua en la energía de Gibbs debería ser modesta [93].

Por consiguiente, el segundo nivel en el mecanismo de unión de los dominios SH3 consiste en una red de puentes de hidrógeno intermoleculares mediados por moléculas de agua, implicadas en la modulación del reconocimiento de ligandos peptídicos [94].

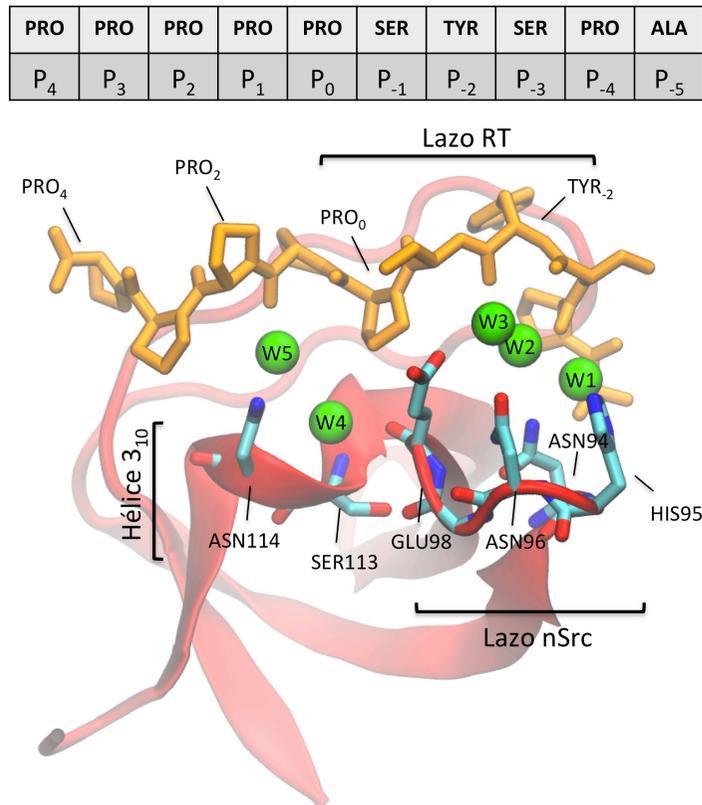
## CAPÍTULO 3

### 3.1.1 Papel energético de las moléculas de agua en el sitio de unión del dominio SH3-cAbl

En este contexto, en un trabajo realizado en nuestro laboratorio se describió por primera vez la presencia de moléculas de agua enterradas en la interfaz de unión del dominio SH3 de cAbl unido al péptido de diseño p41 (APSYSPPPPP). Mediante el análisis de la estructura cristalográfica del dominio SH3 de cAbl unido a p41 (1BBZ), se localizaron 5 moléculas de agua en la interfaz de unión con unos factores  $\beta$  equivalentes a los de los átomos del dominio. Estas moléculas estaban repartidas entre varias cavidades localizadas en las regiones de la hélice  $3_{10}$  y del lazo n-Src, con una exposición al solvente menor de  $10 \text{ \AA}^2$  y mediando interacciones entre el ligando y el dominio Figura 3.1. La geometría y composición de estos sitios de hidratación es ideal para acomodar moléculas de agua, aislándolas parcialmente del solvente y favoreciendo que las moléculas de agua se encuentren estableciendo puentes de hidrógeno con residuos del dominio y del ligando.

Las posiciones caracterizadas en este estudio se denominaron como sitios de hidratación W4 y W5 para los situados en la región de la hélice  $3_{10}$  y W1, W2 y W3 para las situadas en el lazo n-Src. La molécula de agua del sitio W4 está mediando puentes de hidrógeno con los átomos de los esqueletos peptídicos de los restos SER113 y ASN114 de la hélice  $3_{10}$  y del residuo GLU98 del lazo n-Src. Por otro lado, la molécula de agua que se encuentra en la posición W5, está mediando un puente de hidrógeno entre la cadena lateral del residuo ASN114 y el oxígeno carbonilo de la prolina en posición P<sub>1</sub> del ligando. Figura 3.1. Los sitios de hidratación del lazo n-Src del complejo SH3-cAbl/p41, W1, W2 y W3, son adyacentes entre sí, y están mediando puentes de hidrógeno entre los residuos GLU98, ASN96, ASN94 y HIS95 del dominio y las posiciones P<sub>-2</sub> a P<sub>-4</sub> del ligando. La configuración de esta zona es más próxima

a la de un túnel que a la de una cavidad, de forma que las moléculas de agua suelen pasar de un sitio a otro con frecuencia. Este fenómeno se observó en el análisis de las simulaciones de dinámica molecular para el complejo SH3-cAbl/p41 [94].



**Figura 3.1 Representación del dominio de SH3-cAbl formando complejo con el ligando p41 (pdb con código: 1BBZ).** En rojo se representan las estructuras secundarias del dominio SH3-cAbl. Los restos del implicados en la formación de puentes de hidrógeno con moléculas de agua se han representado en color fucsia y en estructura de varillas. El péptido p41 se muestra en color naranja. Las moléculas de agua enterradas en la interfase de unión se indican con número del 1 al 5 como esferas de color verde. En la parte superior se representa la secuencia del ligando p41 e incluye el índice de posición que ocupan. La numeración de los restos del dominio SH3 es la correspondiente a la isoforma A de la proteína cAbl.

Mediante un completo análisis termodinámico, estructural y computacional, se confirmó que las posiciones donde se encontraban dichas moléculas de agua estaban ocupadas por moléculas de agua con tiempos de

## CAPÍTULO 3

residencia anormalmente altos. Por otro lado, se confirmó el papel relevante de estas moléculas de agua interfaciales en la energética de unión del ligando de alta afinidad p41. Mediante la disrupción de posiciones clave para el sitio de hidratación W5 a través de las mutaciones(N114A y N114Q se consiguió alterar la entalpía de unión, observándose además que dicho cambio estaba correlacionado con una alteración del comportamiento de las moléculas de agua a lo largo de las simulaciones de dinámica molecular respecto a lo observado de las simulaciones del complejo silvestre [94]. Estas observaciones corroboran la hipótesis inicialmente propuesta, según la cual las interacciones mediadas por moléculas de agua interfaciales desempeñan un papel clave en la determinación del patrón termodinámico de unión de p41 a SH3-cAbl [51]. En definitiva, podemos concluir que la unión de un ligando al dominio SH3 de cAbl está dirigida por un doble mecanismo en el que se combinan las interacciones hidrofóbicas canónicas en dominios SH3 con una amplia red de enlaces hidrógeno mediados por moléculas de agua [53, 94]. Esta perspectiva abre una nueva dimensión en la comprensión de la afinidad y especificidad de unión en estos sistemas y, obviamente, en el diseño, optimización e identificación de ligandos.

### **3.1.2 Huella termodinámica de los péptidos de diseño para SH3-cAbl.**

#### **Análisis de los posibles cambios en el patrón de hidratación**

De nuevo, en nuestro grupo de investigación, se inició el estudio mediante calorimetría isotérmica de titulación (ITC) de los parámetros termodinámicos de la unión de los diferentes péptidos derivados del proceso de diseño de p41 [53]. El péptido p41 fue diseñado a partir del péptido natural 3BP1 (APTMPPPLPP), sustituyendo posiciones de forma secuencial, de tal

forma que se ensayaron varios péptidos en el proceso hasta lograr el de mayor afinidad y especificidad: el péptido denominado p41 (APSYSPPPPP) [47].

Entre todos los péptidos ensayados, llamaron especialmente la atención los parámetros termodinámicos de los péptidos denominados p0, p7 y p17 respecto a los obtenidos para los péptidos p40 y p41. Como podemos observar en la Tabla 3.1, los péptidos p0, p7 y p17 sólo difieren entre sí en dos posiciones, P<sub>-1</sub> y la P<sub>2</sub>. Dichas posiciones se encuentran muy expuestas al solvente, de forma que no establecen interacciones directas con el dominio SH3-cAbl, Figura 3.4. Pero los datos termodinámicos obtenidos mediante el análisis de ITC revelaron un panorama totalmente distinto a lo esperado, encontrándonos ante una situación donde los resultados de estos péptidos son difíciles de justificar si consideramos el sitio de reconocimiento del dominio SH3 como una simple superficie hidrofóbica.

**Tabla 3.1** Comparación de las secuencias de los péptidos p41, p40, p17, p7 y p0 referenciado la posición que ocupan en la interfaz de unión.

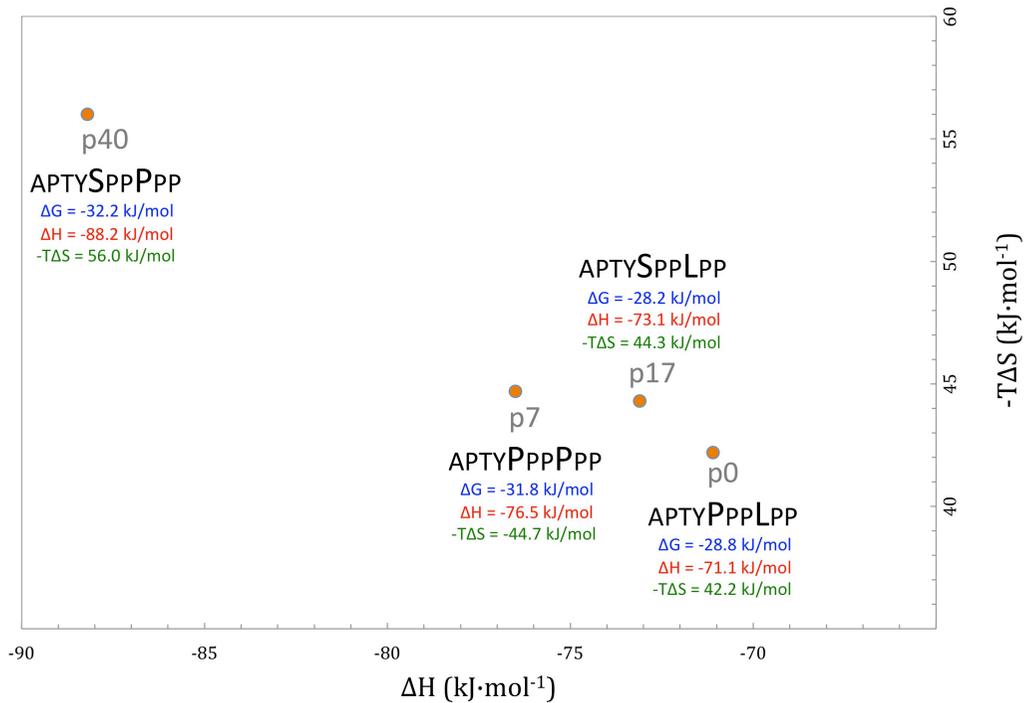
<b>Péptido</b>	<b>P<sub>-5</sub></b>	<b>P<sub>-4</sub></b>	<b>P<sub>-3</sub></b>	<b>P<sub>-2</sub></b>	<b>P<sub>-1</sub>*</b>	<b>P<sub>0</sub></b>	<b>P<sub>1</sub></b>	<b>P<sub>2</sub>*</b>	<b>P<sub>3</sub></b>	<b>P<sub>4</sub></b>
<b>P41</b>	ALA	PRO	SER	TYR	<b>SER</b>	PRO	PRO	<b>PRO</b>	PRO	PRO
<b>P40</b>	ALA	PRO	THR	TYR	<b>SER</b>	PRO	PRO	<b>PRO</b>	PRO	PRO
<b>P17</b>	ALA	PRO	THR	TYR	<b>SER</b>	PRO	PRO	<b>LEU</b>	PRO	PRO
<b>P7</b>	ALA	PRO	THR	TYR	<b>PRO</b>	PRO	PRO	<b>PRO</b>	PRO	PRO
<b>P0</b>	ALA	PRO	THR	TYR	<b>PRO</b>	PRO	PRO	<b>LEU</b>	PRO	PRO

Si examinamos con detalle la Tabla 3.1, comprobamos que los péptidos p0 y p7 se diferencian únicamente en la posición P<sub>2</sub>, donde se ha substituido la leucina del ligando p0 por una prolina, de forma que el péptido p7 está compuesto por una mayoría de residuos prolina, especialmente concentrados en la región del péptido que se acomoda en los bolsillos hidrofóbicos del dominio SH3. Teóricamente lo que cabría esperar es una disminución en la penalización entrópica, ya que las numerosas prolinas deben rigidificar el

## CAPÍTULO 3

péptido y aumentar la tendencia del ligando a adoptar la hélice PPII, conformación necesaria para que el ligando interaccione con el dominio SH3-cAbl como hemos comentado previamente. Asimismo, la entalpía debería verse poco afectada por este cambio, ya que no se espera que la sustitución en esta posición promueva la pérdida o la formación de ninguna nueva interacción. Sin embargo, en la Figura 3.2 observamos un efecto totalmente opuesto: un incremento favorable de la entalpía de  $5,4 \text{ kJ}\cdot\text{mol}^{-1}$  y de la penalización entrópica en  $2,5 \text{ kJ}\cdot\text{mol}^{-1}$  respecto a los valores que presenta p0. Este mismo efecto se reproduce entre los péptidos p17 y p40,. Ambos péptidos también se diferencian únicamente en la posición P<sub>2</sub>, una leucina en p17 y una prolina en p40, y como se aprecia en la Figura 3.2, la introducción de la prolina resulta en un cambio de los parámetros termodinámicos en el mismo sentido que el caso anterior, pero en una magnitud mayor, aumentando la entalpía en  $15,1 \text{ kJ}\cdot\text{mol}^{-1}$  y la penalización entrópica en  $11,7 \text{ kJ}\cdot\text{mol}^{-1}$ .

Seguidamente comparemos las secuencias de p0 y p17: observamos que ambos péptidos solamente se diferencian en la posición P<sub>-1</sub>, donde el péptido p17 tiene una serina y p0 una prolina. Como apreciamos en la Figura 3.2, la sustitución de la prolina en la posición P<sub>-1</sub> provoca un aumento en la entalpía favorable de  $2 \text{ kJ}\cdot\text{mol}^{-1}$  y de la penalización entrópica en  $2,1 \text{ kJ}\cdot\text{mol}^{-1}$ . Ahora bien, si comparamos p7 con p40 observaremos que este mismo cambio, es decir, la sustitución de una prolina en posición P<sub>-1</sub> por una serina, supone un aumento substancialmente mayor y favorable en entalpía de  $12 \text{ kJ}\cdot\text{mol}^{-1}$  y una penalización entrópica, de  $11,2 \text{ kJ}\cdot\text{mol}^{-1}$  respectivamente. Estos resultados indican que las sustituciones de las posiciones P<sub>-1</sub> y P<sub>2</sub> parecen tener un mayor efecto en p40. Como consecuencia, se propuso que existe una sinergia entre la presencia de una serina en posición P<sub>-1</sub> y de una prolina en la posición P<sub>2</sub>.



**Figura 3.2** Representación de la entalpía aparente frente a la entropía obtenida de los experimentos de calorimetría de titulación isotérmica para cada ligando [53]. En azul se resalta la energía libre de Gibbs, en rojo la entalpía aparente y en verde la entropía.

En la descripción clásica del sitio de unión de los dominios SH3, la sustitución de restos alifáticos por prolina en estas posiciones, que como hemos mencionado se encuentran muy expuestas al solvente, debería resultar en una contribución entrópica más favorable a la unión al estabilizar la conformación de hélice PPII del ligando libre. Una posible explicación de las discrepancias entre los parámetros termodinámicos esperados y los observados para estos péptidos podría estar relacionada con una disrupción o alteración de la red de puentes de hidrógeno mediados por moléculas de agua descritos previamente. De esta forma, se postuló que las modificaciones en las posiciones P<sub>-1</sub> y P<sub>2</sub>, provocarían cambios en el comportamiento de las

## CAPÍTULO 3

moléculas de agua de los sitios de hidratación del complejo que se traducirían en cambios de los parámetros termodinámicos de difícil explicación [53].

### 3.1.3 Patrón de sitios de hidratación en la familia de dominios SH3

Posteriormente, y también en nuestro grupo de investigación, con el objetivo de certificar si el doble mecanismo observado para el complejo de SH3-cAbl/p41 era una característica única de este sistema o, por el contrario, era extensible a otros dominios SH3, se recopilaron y analizaron las estructuras cristalográficas disponibles en el PDB de dominios SH3. Se recopilaron todas las estructuras con una resolución superior a los 2 Å y se estudiaron las posiciones de las moléculas de agua cristalográficas enterradas en el sitio de unión[57]. La colección de 98 estructuras resultante de este estudio permitió confirmar la presencia de moléculas de agua cristalográficas enterradas en el sitio de unión en todas las estructuras de dominios SH3 analizadas. Además se observó cómo la localización de algunas moléculas de agua estaba conservada en numerosas estructuras, incluyendo algunas de las aguas descritas para el complejo SH3-cAbl/p41. La observación detallada de los sitios de hidratación de la colección de estructuras insinuaba una clasificación de éstos según su posición [57].

#### 3.1.3.1 Sitios de hidratación en la hélice 3<sub>10</sub>

Como se ha mencionado en la Introducción de esta Tesis, la región 3<sub>10</sub>, al contrario que los lazos RT y n-Src, es una región con un alto grado de conservación de secuencia dentro de la familia de dominios SH3 (PSNY), Figura 3.3. En esta zona encontramos dos sitios de hidratación que están ampliamente representados, aunque cada uno muestra sus propias peculiaridades.

El sitio de hidratación W4 fue el más ampliamente observado, estando ocupado por una molécula de agua en un 77% de las estructuras de la colección. Este sitio de hidratación se encuentra formado por los átomos de los esqueletos peptídicos de los restos SER113 y el ASN114 de la hélice 3<sub>10</sub> y de un residuo del lazo n-Src, normalmente un glutámico o un aspártico (GLU98 para el complejo SH3-cAbl/p41), dejando una cavidad donde existen hasta 4 átomos del esqueleto peptídico del dominio con potencial para formar puentes de hidrógeno con la molécula de agua. Las moléculas de agua que se acomodan en este sitio de hidratación se caracterizan por tiempos de residencia muy altos durante las simulaciones de dinámica molecular (>1 ns). Además, los factores  $\beta$  para las moléculas de agua encontradas en esta posición presentan valores similares a los correspondientes a las zonas más estables del dominio. Es decir, que las moléculas de agua presentes en el sitio de hidratación W4 se encuentran fuertemente unidas a la estructura, lo que nos indica que esta molécula de agua puede ser tratada como parte estructural del dominio y debería ser tomada en consideración en los procesos de diseño racional de péptidos para los dominios SH3.

La ausencia de moléculas de agua en el sitio de hidratación W4 en el 23% de las estructuras de la colección generalmente se encuentra asociada a secuencias y cambios estructurales anómalos del dominio y del ligando, por ejemplo, cambios en la secuencia PNSY de la hélice 3<sub>10</sub> que generarían problemas estéricos que dificultarían la acomodación de moléculas de agua. Igualmente, pueden ocurrir modificaciones de la secuencia del lazo n-Src que variarían su conformación, cambiando la geometría del sitio de hidratación y disminuyendo su capacidad para coordinar la molécula de agua [94].

Por otro lado, el sitio de hidratación W5 se encuentra mediando un puente de hidrógeno entre el dominio y el ligando. En concreto, las moléculas

## CAPÍTULO 3

de agua en esta posición se hallan coordinadas por la cadena lateral de ASN114 y el oxígeno carbonílico del residuo en la posición P<sub>1</sub> del ligando. Este sitio de hidratación, que se encuentra ocupado en un 30% de las estructuras de complejos, no se observa en los dominios libres, por lo que la presencia del ligando es determinante para su ocupación. La molécula de agua se encuentra localizada entre la cadena lateral del resto de la posición 114 de la hélice 3<sub>10</sub> del dominio y el esqueleto peptídico del resto de la posición P<sub>1</sub> del ligando, Figura 3.3, estableciendo puentes de hidrógeno con ambos. Estos dos residuos generalmente se encuentran demasiado alejados para establecer interacciones directas, por lo que la presencia de una molécula de agua en el sitio W5 es muy dependiente de cómo se acomode el ligando en el sitio de unión, y, en consecuencia, de la secuencia de dicho ligando y del residuo que exista en la posición 114 del dominio. [94].

### 3.1.3.2 Sitios de hidratación en los lazos RT y n-Src

Como se observa en la Figura 3.3, en la región de los lazos la mayor dispersión en la posición de las moléculas de agua, nos indica que los sitios de hidratación presentan una mayor flexibilidad, lo que está en concordancia con la mayor variabilidad conformacional y de secuencia de estas regiones. De esta manera, los factores que determinan la presencia de moléculas de agua en estos sitios estarán íntimamente correlacionados con cambios estructurales en los lazos y con la secuencia del ligando. Esta variabilidad de patrones y la mayor flexibilidad de estas posiciones indican que, *a priori*, estos sitios de hidratación pueden jugar un papel importante en la determinación de la especificidad de unión.



## CAPÍTULO 3

En general las moléculas de agua detectadas en el lazo RT median puentes de hidrógeno entre restos cargados del dominio y del ligando. Esto explicaría la ausencia de moléculas de agua en el lazo RT del complejo SH3-cAbl/p41. En cambio, el lazo n-Src es una de las regiones con mayor variabilidad de secuencia y longitud, de forma que las posiciones de las moléculas de agua en esta región se encuentran muy poco conservadas y muy condicionadas a la presencia del ligando.

No obstante, el análisis de la colección de estructuras cristalográficas permitió detectar una molécula de agua en el lazo n-Src bastante conservada, presente en un 60% de las estructuras e independientemente de la presencia o no de ligando. Esta molécula de agua está situada por debajo de los sitios de hidratación W1, W2 y W3 y se encuentra completamente enterrada; aún cuando no parece establecer contacto directo con el ligando, establece una media de 4 puentes de hidrógeno con átomos del esqueleto peptídico del lazo n-Src. Su ausencia está muy correlacionada con la longitud del lazo n-Src, estando presente en el 85% de las estructuras con un lazo n-Src con una longitud comprendida entre los 5 y 8 restos, mientras que no existe en estructuras con lazos más largos o más cortos, como es el caso de SH3-cAbl. Las moléculas de agua en este sitio de hidratación presentan unos valores de factor  $\beta$  equivalentes a los de los átomos del dominio, lo que nos indicaría que podría tratarse de una molécula de agua estructural, al igual que W4 [57]. A este sitio de hidratación nos referiremos en esta Tesis como sitio de hidratación WA-nSrc (Figura 3.3).

### 3.1.4 Estudio y seguimiento de las moléculas de agua en los sitios de hidratación de los dominios SH3

Gracias a toda la información recopilada por nuestro grupo de investigación podemos afirmar que la presencia de moléculas de agua en la interfase de unión no es una característica única de SH3-cAbl/p41. Además, si tenemos en cuenta que todos los dominios SH3 comparten la peculiar huella termodinámica, el mecanismo de interacción dual descrito para el complejo SH3-cAbl/p41 no debe ser una excepción.

Por otro lado, el análisis de la colección de estructuras nos permite esbozar una primera clasificación de los sitios de hidratación, agrupándolos en tres tipos: i) **sitios de hidratación estructurales**, altamente conservados, con moléculas de agua firmemente unidas a la estructura del dominio y que no parecen interactuar directamente con el ligando, como por ejemplo W4 y WA-nSrc; ii) un segundo tipo de sitios de hidratación, algo menos conservados y formados por residuos del dominio y el ligando, donde se alojan moléculas de agua que median puentes de hidrógeno entre ambos, siendo necesaria la presencia del ligando para fijar las moléculas de agua, que podríamos denominar **sitios de hidratación optimizadores** (un claro ejemplo de este tipo de sitios es el sitio de hidratación W5); iii) todos aquellos que presentan una baja conservación y generalmente moléculas de agua mucho más expuestas al solvente, agrupados principalmente en la región de especificidad de los dominios SH3, que denominaremos **sitios de hidratación de especificidad**, como por ejemplo W1, W2 y W3

La existencia de diferentes tipos de sitios de hidratación, sumado a la presencia de numerosos patrones de hidratación entre los distintos dominios, sugiere que no todas las moléculas de agua interfaciales juegan el mismo papel ni contribuyen por igual al mecanismo de unión. Es imprescindible conocer

## CAPÍTULO 3

estos detalles si queremos diseñar ligandos de alta afinidad y especificidad. En este sentido, existen dos estrategias principales para aprovechar las moléculas de agua enterradas en la interfaz de unión para aumentar la afinidad del ligando: o bien modificar el ligando para que expulse la molécula de agua, o bien intentar establecer nuevas interacciones del ligando con la molécula de agua. Sin embargo, para poder aprovechar correctamente la red de las moléculas de agua, no basta con conocer la posición en la que se encontró una molécula de agua enterrada. Así, por ejemplo, expulsar moléculas de agua firmemente unidas al dominio podría dar lugar a una gran penalización entálpica al romperse interacciones muy favorables, igualando o superando a la ganancia entrópica de la liberación de la molécula al solvente, lo que podría suponer una pérdida de la afinidad.

Tanto el coste entrópico como la ganancia entálpica de enterrar una molécula de agua son parámetros muy variables y de difícil predicción, que dependen de diferentes factores como el grado de inmovilización de la molécula de agua en el sitio de hidratación o el número y la fuerza de las interacciones que establece. De esta forma, poseer el máximo posible de información sobre las propiedades de las moléculas de agua alojadas en los sitios de hidratación es importante, ya que nos da pistas de qué estrategia seguir para aumentar la afinidad de un ligando.

Los sitios de hidratación descritos en las Secciones anteriores fueron detectados mediante el análisis de las estructuras cristalográficas. La conformación de los sitios de hidratación permite que las moléculas de agua queden inmovilizadas en un mínimo energético, de forma que pueden ser detectadas en los mapas de densidad electrónica al resolver la estructura por cristalografía de rayos X. Aunque el análisis de las estructuras cristalográficas aporta una valiosa información para entender los factores que determinan la

presencia de estos sitios de hidratación, es una información incompleta. Por un lado es difícil inferir a partir de la información cristalográfica las propiedades dinámicas de las moléculas de agua, algo esencial para poder racionalizar su participación en el mecanismo de especificidad y afinidad [95]. Además, las diferentes resoluciones a las que están resueltas las estructuras cristalográficas y la disparidad de criterios a la hora de anotar las moléculas de agua [96, 97], dificultan la identificación de los factores que determinan la ocupación de estos sitios de hidratación, de forma que es necesario complementar la información cristalográfica con datos de diferentes técnicas para entender el papel de estos sitios de hidratación, entre las que destacan la dispersión de neutrones [98, 99] y la espectroscopia de infrarrojo [100, 101], entre otras [102-104]. Recientemente las simulaciones de dinámica molecular han demostrado un enorme potencial en este campo, ya que proporcionan información inaccesible desde otras perspectivas [105, 106], demostrando en numerosos casos su valía en el estudio del papel de las moléculas de agua en la función de las macromoléculas [86, 107, 108].

Con el objetivo de ampliar nuestro conocimiento de los factores que determinan la existencia de estos sitios de hidratación dentro de la familia de dominios SH3 y el papel que tienen en el mecanismo de especificidad, es necesario desarrollar un protocolo para detectar los sitios de hidratación mas allá de la información cristalográfica disponible, y conocer las propiedades dinámicas de las moléculas de agua que se alojan en los sitios de hidratación. En el marco de esta Tesis, hemos abordado esto último. Nos hemos centrado en primer lugar en el desarrollo y validación de un algoritmo que permita conocer y registrar las propiedades de los sitios de hidratación detectados por cristalografía. Y una vez este puesto apunto, utilizar esta herramienta para abordar un estudio comparativo, incluyendo los casos anteriormente

## CAPÍTULO 3

expuestos en la Sección 3.1. La comparación del comportamiento de las moléculas de agua entre los diferentes péptidos de diseño con el objetivo de confirmar si existe un cambio en los sitios de hidratación que sea el responsable de los cambios en el patrón termodinámico. En segundo el algoritmo nos permitirá comparar la configuración de sitios de hidratación entre dos representantes de diferentes subfamilias de dominios SH3 para determinar si podíamos establecer patrones generales de estos sitios de hidratación. De este modo, abrimos el campo al desarrollo futuro de un método automático de análisis de estructuras para la detección de potenciales sitios de hidratación. Permitiendo su clasificación y descripción para un potencial uso en el diseño de ligandos peptídicos aprovechando completamente este doble mecanismo de unión.

### 3.2 RESULTADOS Y DISCUSIÓN

Aunque las simulaciones de dinámica molecular son una técnica muy valiosa para estudiar el comportamiento de las moléculas de agua, la mayoría de las herramientas disponibles para el análisis de trayectorias no están orientadas a obtener tiempos de residencia para moléculas de agua individuales de sitios concretos de la proteína. El análisis que se realizó para describir el comportamiento de las moléculas de agua del complejo SH3-cAbl/p41 fue parcialmente llevado a cabo mediante el análisis visual de cada paso de la trayectoria. Pero, debido a lo tediosa que puede ser esta tarea (en una simulación estándar de 10 ns de un dominio SH3 existen alrededor de 6000 moléculas de agua y de unos 5000 pasos de trayectoria) se propuso el desarrollo de un algoritmo para seguir y anotar de forma automática el comportamiento de las moléculas de agua en los sitios de hidratación, necesitando un mínimo de capacidad computacional. De esta forma, nuestros

algoritmo tendrá el potencial de analizar y comparar entre sí un gran número de simulaciones entre miembros de la familia SH3. En esta Tesis describimos, a modo de ejemplo, el estudio del patrón de moléculas de agua en los diferentes complejos del dominio SH3 de cAbl con los péptidos de diseño, así como la comparación de los patrones de hidratación entre las dominios de cSrc y cAbl.

### **3.2.1 ZAHORÍ: un algoritmo para el seguimiento de moléculas de agua en trayectorias de dinámica molecular**

El programa de seguimiento de aguas, ZAHORÍ, se desarrolló en el lenguaje de programación python. Este lenguaje de programación se encuentra ampliamente usado en biología estructural [109], habiendo demostrado su potencia y versatilidad para desarrollar de forma rápida y eficiente herramientas científicas [110-115]. Con el objetivo de validar los resultados del programa, se usaron como referencia los valores obtenidos de las dinámicas del complejo SH3-cAbl/p41 [94].

#### **3.2.1.1 Descripción del funcionamiento de ZAHORÍ**

ZAHORÍ está dividido en dos bloques: en el primero el programa adquiere toda la información relevante de la trayectoria para el análisis de los distintos sitios de hidratación bajo estudio y la almacena en una base de datos implementada mediante SQLite 3.1. Esto permite el acceso a la información de una forma rápida y ordenada en el segundo bloque, en el que se obtienen los resultados para cada sitio de hidratación.

Con la intención de poder utilizar el programa independientemente del software que se use para la generación de las simulaciones, en el primer bloque se convierte la trayectoria a un formato estándar. El formato

## CAPÍTULO 3

seleccionado por ser el más versátil es el fichero de coordenadas PDB. Para cada uno de estos ficheros, correspondiente a un paso de la trayectoria, se genera una lista con todas las moléculas de agua que están formando un puente de hidrógeno con alguno de los átomos que definen los sitios de hidratación. Para ello, ZAHORÍ hace uso del programa externo HBplus [116], que analiza el fichero PDB e identifica los puentes de hidrógeno que se están formando, utilizando como umbral una distancia máxima entre el donador y el aceptor de 3,5 Å y un ángulo máximo de 90° entre el hidrógeno de los grupos donador y aceptor.

En el segundo bloque, el programa analiza la evolución cada sitio de hidratación a lo largo de la trayectoria, anotando los intervalos de tiempo en los que las moléculas de agua se encuentran ocupando el sitio, de acuerdo con los criterios que se describen en la siguiente Sección. Una vez el bloque ha determinado todos los intervalos para un sitio de hidratación, podemos obtener como resultado el tiempo máximo de ocupancia, el tiempo medio de ocupancia, la frecuencia de ocupancia y el total de moléculas de agua que ocupan cada sitio de hidratación durante la simulación. Aunque estos parámetros están más o menos relacionados entre sí, describen diferentes características del sitio de hidratación. Así, el tiempo máximo de ocupancia corresponde al mayor intervalo en el que una molécula de agua ha permanecido en el sitio de hidratación; algunos autores relacionan la magnitud de este tiempo con la barrera energética que debe superar una molécula de agua para abandonar el sitio de hidratación. A modo de ejemplo, podemos considerar el comportamiento de las moléculas de agua que se pueden encontrar enterradas en la estructura de las proteínas: éstas presentan tiempos máximos de ocupación en la escala de los nanosegundos, ya que se encuentran muy aisladas del solvente y deben existir grandes cambios

conformacionales para que puedan ser liberadas. Paralelamente, una vez que la molécula de agua abandona esta posición aislada puede resultar muy difícil que otra molécula de agua vuelva a ocupar el sitio de hidratación, presentando por tanto frecuencias de ocupación inferiores a un 20%. La frecuencia de ocupación dependería de la barrera energética para transferir una molécula del solvente al sitio de hidratación [107].

Para el cálculo de todos estos parámetros se descartaron todas aquellas moléculas de agua con intervalos de residencia inferiores a los 6 ps. La eliminación de estas moléculas de agua (que podemos definir como ultra rápidas) responde a que, en ausencia de otro cosolvente, el agua simplemente tiende a rellenar el espacio disponible, de forma que al no tomar en consideración estos cortos intervalos nos aseguramos que la frecuencia de ocupancia refleja la frecuencia con la que el sitio está ocupado por moléculas de agua no oportunistas, es decir, unidas mediante interacciones específicas [107].

### **3.2.1.2 Definición de ocupancia y validación del algoritmo**

Un primer paso crítico en el análisis de la hidratación a lo largo de una trayectoria de dinámica molecular es la definición del término ocupancia. Como ya hemos comentado, se decidió mantener el programa dentro de la máxima simplicidad posible, de forma que en un principio consideramos simplemente que una molécula de agua está ocupando un sitio de hidratación si ésta se encuentra formando al menos una de las interacciones cristalográficas, permitiendo que este criterio fuera infringido por periodos iguales o inferiores a 2 ps. Este criterio resultó ser insuficiente para la determinación automática e inequívoca de las moléculas de agua durante las trayectorias, ya que cada sitio de hidratación está formado por varios

### CAPÍTULO 3

donadores y aceptores de hidrógeno para los que es posible encontrar varias moléculas de agua que estén en disposición de establecer al menos una de las varias interacciones cristalográficas. Una posible forma para mejorar el cribado, sería escoger la molécula de agua con mayor número de interacciones pero, evidentemente, este criterio se enfrentaría a numerosos casos en los que existen varias moléculas de agua con el mismo número de interacciones. Otro criterio posible sería usar estrictamente la posición geométrica, es decir, establecer que una molécula se encuentra ocupando el sitio de unión si está en la misma posición que la molécula de agua cristalográfica.

La determinación de la posición geométrica de la molécula de agua demostró ser clave para discriminar si ésta se encuentra o no en el sitio de hidratación. Para conocer si la molécula de agua se encontraba en la misma posición que la cristalográfica, se decidió usar como referencia la distancia al centro de masas de los átomos que forman el sitio de hidratación. Se determinó que los átomos que forman el sitio de hidratación son todos aquellos átomos donadores o aceptores de hidrógeno ¿expuestos¿ al solvente que estén a una distancia menor de 4 Å de la molécula de agua cristalográfica. De esta forma, consideraremos que una molécula de agua se encuentra en un sitio de hidratación siempre que se encuentre por debajo de una distancia límite del centro de masas del sitio de hidratación.

El uso exclusivo del criterio geométrico, aunque mejoró mucho la detección inequívoca de las moléculas de agua, no evitó seguir sufriendo problemas de detección de falsos positivos y falsos negativos. La flexibilidad y disposición espacial de los átomos seleccionados para algunos sitios de hidratación hacen que el centro de masas quede, en ocasiones puntuales, muy enterrado en la proteína, alejándose mucho de la molécula de agua que ocupa el sitio de hidratación, lo que nos llevó a incrementar la distancia límite por

encima de los 4 Å para evitar que el sitio de hidratación se considerara desocupado. No obstante, dicho incremento provocaba la pérdida de resolución, detectando numerosos falsos positivos, es decir, permitía la selección de moléculas que no se encontraban realmente en el sitio de hidratación. A la vista de estos resultados, se decidió usar la combinación de la información geométrica con la de los puentes de hidrógeno cristalográficos, determinando que la molécula de agua mejor situada geométricamente y con un mayor número de interacciones era la que se consideraba que estaba ocupando el sitio de hidratación, lo que permitió rebajar el límite de distancia hasta 3,5 Å y detectar de forma correcta las moléculas de agua de cada sitio de hidratación.

**Tabla 3.2** Comparativa de los resultados obtenidos del análisis semi-manual con los obtenidos por la primera versión de ZAHORÍ para las dinámicas del complejo SH3-cAbl/p41. El tiempo máximo de residencia ( $T_{max}$ ), el Tiempo medio de residencia ( $T_{medio}$ ), número de moléculas de agua durante la trayectoria y la frecuencia de tiempo de la trayectoria que el sitio de hidratación se encuentra ocupado por una molécula de agua.

Sitio de hidratación	$T_{max}$ (ps)	$T_{medio}$ (ps)	Nº de moléculas de agua	Frecuencia de Ocupancia
<b>W1</b> Manual	<b>340</b>		<b>92</b>	<b>89,4</b>
W1 ZAHORÍ	228	26,29	96	60,74
<b>W2</b> Manual	<b>1584</b>		<b>59</b>	<b>97,2</b>
W2 ZAHORÍ	626	84,23	62	91,81
<b>W3</b> Manual	<b>82</b>		<b>25</b>	<b>7,2</b>
W3 ZAHORÍ	744	118,3	53	94,64
<b>W4</b> Manual	<b>2712</b>		<b>12</b>	<b>97,6</b>
W4 ZAHORÍ	2682	344,14	12	96,36
<b>W5</b> Manual	<b>1214</b>		<b>46</b>	<b>79,8</b>
W5 ZAHORÍ	1310	81,00	95	86,68

Al comprobar los datos generados por esta primera versión del algoritmo con los resultados obtenidos de forma no automática en el trabajo con SH3-cAbl (Tabla 3.2) observamos que los resultados son similares en

## CAPÍTULO 3

ambos casos. No obstante, éste no fue el caso para el resto de sitios de hidratación.

**Tabla 3.4** Comparativa de los resultados obtenidos del análisis semi-manual con los obtenidos por la segunda versión de ZAHORÍ para las dinámicas del complejo SH3-cAbl/p41. El tiempo máximo de residencia ( $T_{max}$ ), el Tiempo medio de residencia ( $T_{medio}$ ), número de moléculas de agua durante la trayectoria y la frecuencia de tiempo de la trayectoria que el sitio de hidratación se encuentra ocupado por una molécula de agua.

Sitio de hidratación	$T_{max}$ (ps)	$T_{medio}$ (ps)	Nº de moléculas de agua	Frecuencia de Ocupancia
<b>W1</b> Manual	<b>340</b>		<b>92</b>	<b>89,4</b>
W1 Zahori	228	26,29	96	60,74
<b>W2</b> Manual	<b>1584</b>		<b>59</b>	<b>97,2</b>
W2 Zahori	626	84,23	62	91,81
<b>W3</b> Manual	<b>82</b>		<b>25</b>	<b>7,2</b>
W3 Zahori	744	118,3	53	94,64
<b>W4</b> Manual	<b>2712</b>		<b>12</b>	<b>97,6</b>
W4 Zahori	2682	344,14	12	96,36
<b>W5</b> Manual	<b>1214</b>		<b>46</b>	<b>79,8</b>
W5 Zahori	1302	108,03	44	70,22

El análisis de la trayectoria para el sitio de hidratación W5 permitió detectar que la sobre-estimación en la frecuencia de ocupancia se debía a que, en ciertos momentos, el oxígeno carboxilo de la prolina en la posición P<sub>-1</sub> y el nitrógeno delta de la cadena lateral de ASN114 se encontraban formando un puente de hidrógeno entre ellos, de forma que no era posible que una molécula de agua estuviera mediando un puente de hidrógeno entre el ligando y el dominio. No obstante, aún en estos casos, el programa sí detectaba moléculas de agua en este sitio de hidratación. Para solventar este problema se actualizó el algoritmo añadiendo un paso de control adicional. Así, antes de comprobar si existen moléculas de agua en los sitios de hidratación formados por solamente un donador y un aceptor, se comprueba si éstos se encuentran formando un puente de hidrógeno, y en caso afirmativo se considera el sitio de

hidratación como no disponible y, por lo tanto, no ocupado por ninguna molécula de agua.

Como se observa en la tabla, 3.4 estas precauciones mejoraron la predicción para la posición W5, acercando los resultados del algoritmo automático a los datos obtenidos del análisis convencional. Sin embargo, aún no teníamos unos buenos resultados para las moléculas de agua del lazo n-*Src*, W1, W2 y W3. Como podemos ver en la Figura 3.1, estos sitios de hidratación se encuentran muy contiguos, asemejándose más a un canal con numerosos grupos polares disponibles para establecer puentes de hidrógeno, estando los sitios de hidratación menos definidos que los de la hélice 3<sub>10</sub>. Esto, sumado a la flexibilidad de la región del lazo n-*Src*, facilita que las moléculas de agua fluyan a través de los sitios de hidratación, existiendo incluso momentos en los que la región es ocupada por más de 3 moléculas de agua. En estos casos, frecuentemente, una o varias moléculas de agua pueden cumplir los criterios de selección para más de un sitio de hidratación, al estar ocupando posiciones intermedias entre dos de estos sitios durante periodos de tiempo relativamente largos. Por tanto, una misma molécula de agua podría ser seleccionada para dos sitios de hidratación distintos. Con el objetivo de evitar esta redundancia se modificó el algoritmo añadiendo un nuevo criterio, de manera que no se permite que una molécula de agua pueda ser seleccionada dos veces para distintos sitios de hidratación. Así, una vez una molécula de agua ha sido identificada como ocupante de un sitio de hidratación para un intervalo de tiempo, no está disponible para ser asignada a otro sitio de hidratación durante el mismo periodo de tiempo.

Por otro lado en la Tabla 3.5, se observa que esta modificación del algoritmo provoca una enorme disminución de los valores de ocupancia en las posiciones del lazo n-*Src* respecto a los de la versión anterior del programa,

## CAPÍTULO 3

pero al mismo tiempo siguen siendo diferentes a los datos de ocupancia recopilados manualmente. Esta divergencia se debe al comportamiento y disposición singular de los sitios de hidratación del lazo n-Src, ya que, como vemos, esta modificación no ha afectado a los datos de los sitios de la hélice 3<sub>10</sub>. De forma que aunque los resultados para los sitios de hidratación W1, W2 y W3 no sean óptimos, se trata de una situación extremadamente compleja; en cualquier caso, nuestro objetivo inicial que era la caracterización de sitios de hidratación bien definidos y estables está conseguido. Por lo tanto, estamos en situación de aplicar este algoritmo a algunos sistemas de interés, como validación y como paso previo a su incorporación a algoritmos más complejos de identificación de sitios de hidratación, que se abordará en un futuro.

**Tabla 3.5** Comparativa de los resultados obtenidos por el análisis semi-manual con los obtenidos por la última versión de ZAHORÍ para las dinámicas del complejo SH3-cAbl/p41. El tiempo máximo de residencia (*T*<sub>max</sub>), el Tiempo medio de residencia (*T*<sub>medio</sub>), número de moléculas de agua durante la trayectoria y la frecuencia de tiempo de la trayectoria que el sitio de hidratación se encuentra ocupado por una molécula de agua

Sitio de hidratación	<i>T</i> <sub>max</sub> (ps)	<i>T</i> <sub>medio</sub> (ps)	Nº de moléculas de agua	Frecuencia de Ocupancia
<b>W1<sub>Manual</sub></b>	<b>340</b>		<b>92</b>	<b>89,4</b>
W1 Zahori	158	20,96	104	55,34
<b>W2<sub>Manual</sub></b>	<b>1584</b>		<b>59</b>	<b>97,2</b>
W2 Zahori	106	18,82	46	26,92
<b>W3<sub>Manual</sub></b>	<b>82</b>		<b>25</b>	<b>7,2</b>
W3 Zahori	182	30,48	69	55,18
<b>W4<sub>Manual</sub></b>	<b>2712</b>		<b>12</b>	<b>97,6</b>
W4 Zahori	2682	344,14	12	96,36
<b>W5<sub>Manual</sub></b>	<b>1214</b>		<b>46</b>	<b>79,8</b>
W5 Zahori	1302	108,03	44	70,22

### 3.2.2 Análisis de las diferencias en entalpía entre los ligandos de diseño del dominio SH3 de cAbl

Como hemos mencionado en la Sección 3.1.2, el péptido de mayor afinidad para SH3-cAbl, denominado p41, fue diseñado usando como plantilla el péptido natural 3BP1. En el proceso de diseño se substituyeron diferentes posiciones de forma secuencial con la intención de ir mejorando la afinidad, generando numerosos péptidos hasta lograr los de mayor afinidad p41 y p40. Estos, aún estando compuestos por un número anormalmente bajo de grupos polares, mostraron una elevada y favorable entalpía de unión.

Recordemos también que el análisis detallado de los parámetros termodinámicos de cada péptido producido en el proceso de diseño [53], reveló un comportamiento interesante al comparar los parámetros termodinámicos de los péptidos p0, p7, p17 y p40 y que no correspondía a las propiedades individuales de los restos en los que se diferenciaban y que parecían tener un efecto cooperativo [53]. Estos péptidos, como podemos observar con detalle en la Tabla 3.6, difieren solamente en dos posiciones, P<sub>-1</sub> y P<sub>2</sub>. Estas posiciones se encuentran muy expuestas al solvente (Figura 3.4), con lo cual, *a priori*, no cabría esperar que los cambios entálpicos entre los péptidos estén correlacionados con el establecimiento de nuevas interacciones con el dominio. Como sabemos gracias a los trabajos realizados en nuestro grupo de investigación, la peculiar huella termodinámica de los dominios SH3, puede ser parcialmente causada por la existencia de un doble mecanismo de unión, de forma que, además de las interacciones hidrofóbicas, existiría una red de puentes de hidrógeno mediados por moléculas de agua ocluidas en la interfaz de unión que provocaría el cambio en el balance de los parámetros termodinámicos de la unión, de modo que ésta está dirigida entálpicamente y ligeramente penalizada entrópicamente. Esto condujo a postular que las

## CAPÍTULO 3

diferencias observadas en los parámetros termodinámicos entre los péptidos p0, p7, p17 y p40 podrían tener su origen en una alteración en este segundo nivel del mecanismo de interacción, ya que los cambios de entalpía no pueden ser explicados por si solos si tenemos en cuenta solamente las propiedades individuales de los residuos substituidos. De esta forma, los cambios en las posiciones P<sub>-1</sub> y P<sub>2</sub> estarían alterando la red de puentes de hidrógeno que componen el mecanismo secundario de unión de los dominios SH3.

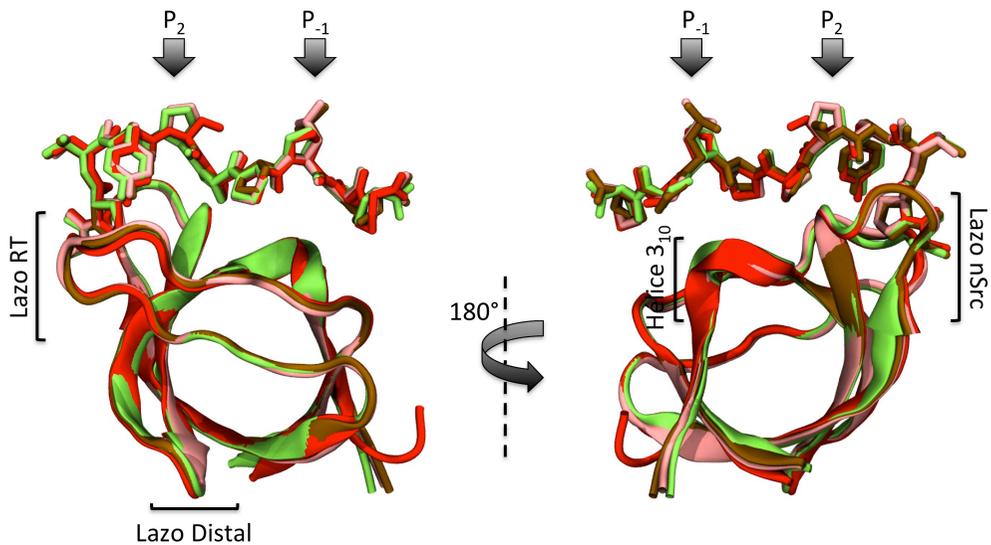
**Tabla 3.6** Parámetros termodinámicos obtenidos mediante ITC para los diferentes ligandos derivados de p41 en su unión al dominio SH3-cAbl. Se resaltan en Naranja las posiciones en los que los péptidos difieren en secuencia.

Péptido	Secuencia	$\Delta G_{app}(\text{kJ mol}^{-1})$	$\Delta H_{app}(\text{kJ mol}^{-1})$	$-T\Delta S_{app}(\text{kJ mol}^{-1})$
P41	APSYSPPPPP	-32,2	-91,7	59,5
P40	APTYSPPPPP	-32,2	-88,2	56
P17	APTYSPPLPP	-28,8	-73,1	44,3
P7	APTYPPPPPP	-31,8	-76,5	44,7
P0	APTYPPPLPP	-28,8	-71,1	42,2

Con el objetivo de determinar el origen de estas diferencias entálpicas entre los dominios, se inició un estudio computacional en el que se analizaron, tanto la red de interacciones intermoleculares, como la red de sitios de hidratación detectados para SH3-cAbl/p41. Recordemos que la fuente de estos cambios debe residir en la interfaz de unión y en el ligando, ya que el dominio es el mismo para todos los complejos.

Como observamos en la Tabla 3.6, el péptido p41 se diferencia de p40 en la posición P<sub>-3</sub>, pero la diferencia en los parámetros termodinámicos entre p40 y p41 no es muy grande ( $\Delta\Delta H = 3,5 \text{ kJmol}^{-1}$  y  $-T\Delta\Delta S = -3,5$ ). Debido que hasta la fecha no disponemos información cristalográfica del complejo SH3-cAbl/p40, usaremos el complejo SH3-cAbl/p41 para detectar el origen de las

peculiares y atípicas diferencias entálpicas originadas en las posiciones, P<sub>-1</sub> y P<sub>2</sub>.



**Figura 3.4.** Superposición de las estructuras cristalográficas de los complejos SH3-cAbl/p41 (Rojo), SH3-cAbl/p17 (marrón), SH3-cAbl/p7 (verde) y SH3-cAbl/p0 (rosa). El dominio se encuentra representado resaltando las estructuras secundarias, mientras que los ligandos peptídicos se encuentran representados en estructura de varillas. Las posiciones en las que los ligandos son diferentes en secuencia han sido resaltadas con un flecha negra.

Para abordar este análisis, se generaron 5 trayectorias independientes de 12 ns de duración para cada complejo, siguiendo el protocolo descrito en la Sección 5.2 de Materiales y Métodos . Estas trayectorias fueron posteriormente analizadas mediante el programa ZAHORÍ, descrito en la Sección anterior, haciendo también uso de las herramientas del paquete de simulaciones de dinámica molecular de AMBER para el análisis de los puentes de hidrógeno intermoleculares. Como estructura de partida para el complejo

## CAPÍTULO 3

SH3-cAbl/p41 se usó la estructura cristalográfica con PDB 1BBZ, mientras que para el resto de complejos se usaron las estructuras cristalográficas resueltas por la Dra. Ana Camara Artigas en colaboración con nuestro grupo de investigación. Anteriormente al uso de las estructuras cristalográficas, este mismo estudio se inició mediante la generación de modelos a partir de la información cristalográfica de 1BBZ. En general, los resultados obtenidos con los modelos fueron muy similares a los obtenidos a partir de las estructuras cristalográficas, lo que es un indicativo de que, al menos en este sistema, el uso de modelos es una metodología que no desvirtúa significativamente los resultados.

Como observamos en la Figura 3.4, al superponer las estructuras cristalográficas para cada complejo no podemos apreciar grandes diferencias ni en el dominio ni en la conformación del péptido. Esto confirma que las diferencias energéticas no están asociadas a cambios en el modo de unión general del péptido. Por otro lado, al analizar las moléculas de agua cristalográficas de los complejos SH3-cAbl/p0, SH3-cAbl/p7 encontramos que los sitios de hidratación de la hélice  $3_{10}$  (W4 y W5) aparecen ocupados por una molécula de agua. Esto no es así, para el sitio del lazo n-Src W3 donde no encontramos ninguna estructura con el sitio ocupado. Excepcionalmente para la estructura del complejo SH3-cAbl/p17, no encontramos ninguna molécula de agua ocupando los sitios descritos anteriormente.

A pesar de que las estructuras cristalográficas no presentan grandes diferencias entre sí, inicialmente decidimos realizar un análisis de la dinámica estructural de los péptidos a lo largo de las simulaciones de dinámica molecular. Y así, evaluar los aspectos conformacionales de los péptidos en los distintos complejos. En la Figura 3.5 se representa la frecuencia de los valores que adquieren los ángulos diedros del esqueleto peptídico ( $\phi$  y  $\psi$ ) para

cada complejo a lo largo de las simulaciones de dinámica molecular. Al examinar los valores, observamos que, en la región del péptido que se acomoda en los bolsillos hidrofóbicos del dominio SH3, la introducción de una leucina en P<sub>1</sub> (p0 y p17) conduce a una mayor dispersión de los valores respecto a los que tiene una prolina en dicha posición (p7 y p41). Esto se debería a que la leucina no proporciona la misma tendencia a adoptar la conformación PPII en el péptido, pero este leve aumento en la flexibilidad no repercute en el comportamiento de las posiciones vecinas. Algo parecido sucede al introducir una serina en la posición P<sub>2</sub>, que conduce a un ligero aumento de la deriva del ángulo, pero que igualmente no afecta a las posiciones colindantes. Es decir, el análisis de los aspectos conformacionales de los péptidos no proporciona ningún indicio que permita racionalizar el comportamiento termodinámico observado, ya que la sustitución de restos de serina y leucina por prolina reduce los grados libertad conformacional del ligando peptídico, lo que debería conducir a una mayor afinidad de unión por motivos entrópicos, no entálpicos. Estos resultados confirman que el origen de las diferencias en la termodinámica de unión debe ser otro.

En este punto, procedimos al análisis de los patrones de hidratación y las propiedades de los distintos sitios en cada uno de los complejos. En la Tabla 3.7 se recogen los resultados obtenidos de la aplicación del programa ZAHORÍ al análisis de las distintas simulaciones. Para un mismo complejo, los resultados para cada réplica fueron promediados, exceptuando el tiempo máximo de ocupancia, para el que se anotó el tiempo máximo observado de entre las 5 replicas. En general, el análisis se centró en los cinco sitios de hidratación encontrados y descritos para el complejo SH3-cAbl/p41. Conviene señalar que *a priori* no existen diferencias en el ligando que puedan promover un cambio en el patrón de hidratación descrito para SH3-cAbl/p41, ya que las

## CAPÍTULO 3

posiciones P<sub>1</sub> y P<sub>2</sub> no están directamente implicadas en ningún sitio de hidratación.

Al comparar los valores obtenidos para los sitios de hidratación de la región de la hélice 3<sub>10</sub>, observamos que los valores de frecuencia de ocupancia y tiempo medio de ocupancia para el péptido p7 en el sitio de hidratación W5 son 3 veces menores que los obtenidos para el resto de péptidos. Esto es debido, como veremos en la Tabla 3.8, a que se establece un puente de hidrógeno directo entre la cadena lateral de la ASN114 de la hélice 3<sub>10</sub> con el oxígeno del esqueleto peptídico de la prolina del péptido en la posición P<sub>1</sub>, lo que evita que se acomode una molécula de agua en dicho sitio de hidratación. Esto podría ser un indicador de que el péptido p7 está acomodándose de forma diferente los demás. Recordemos que el péptido p7 está compuesto mayoritariamente por prolinas (APTYPPPPPP), de forma que la sucesión de prolinas favorece que el ligando adopte una conformación de hélice PPII muy rígida, lo que podría ser el origen de este cambio.

En los valores recopilados para el sitio de hidratación W4 llama la atención cómo en el péptido p0 los tiempos de residencia y el tiempo medio de ocupancia son ligeramente más altos que para el resto de péptidos. Recordemos que las moléculas de agua del sitio de hidratación W4 no establecen interacciones directas con el ligando, de forma que sorprende encontrar un tiempo de residencia 3 veces mayor al resto de complejos.

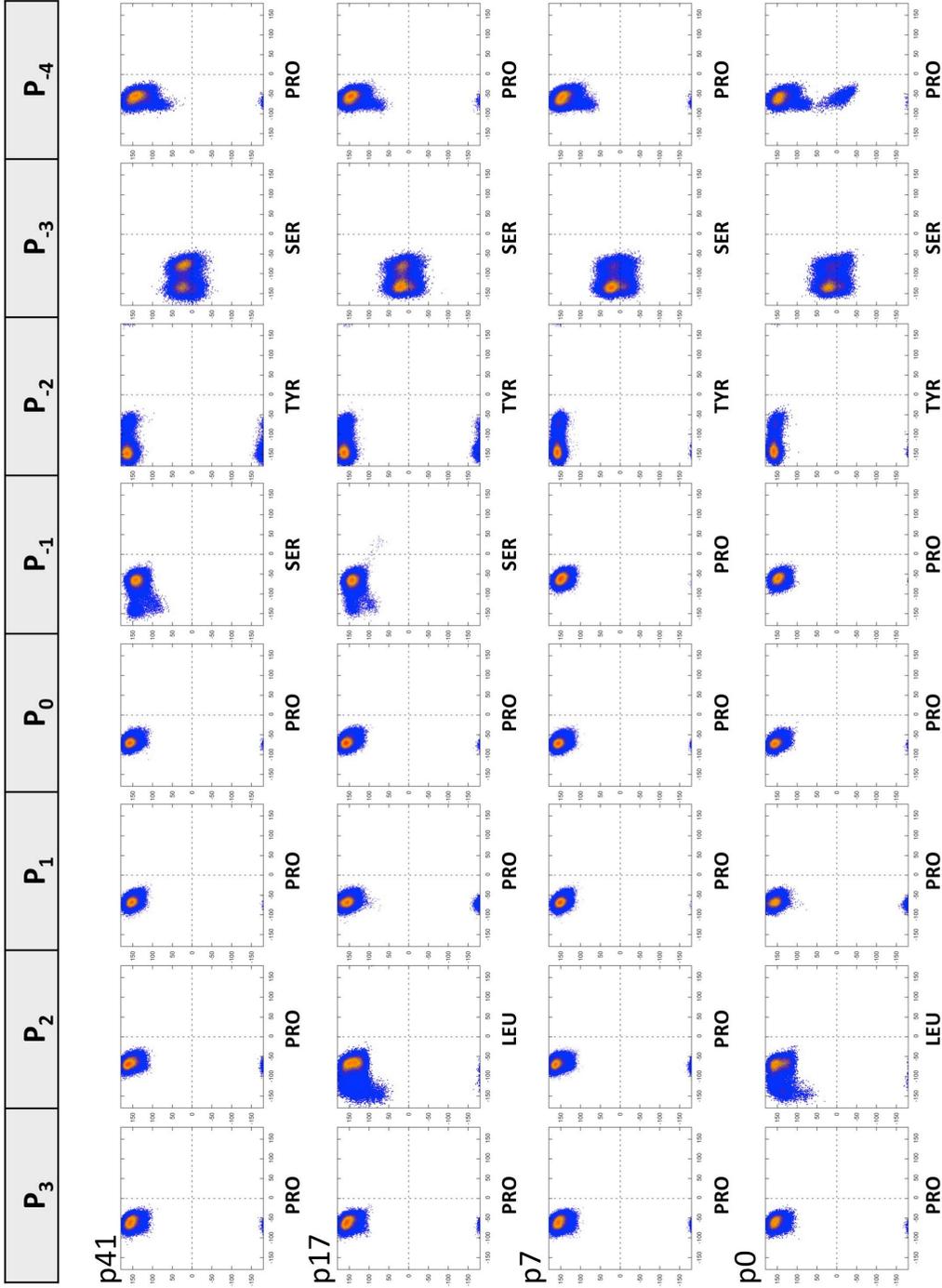


Figura 3.5 Distribución de los valores de los ángulos diedros ( $\phi$  y  $\psi$ ) del esqueleto peptídico para las posiciones del ligando a lo largo de las trayectorias. En la parte superior se indica la posición que ocupa cada residuo. En la distribución de ángulos diedros para cada residuo se representan en azul aquellas posiciones cuyo valor es adoptado entre 1 a 10 veces a lo largo de la trayectoria, en naranja de 10 a 200 veces y en rojo más de 200.

## CAPÍTULO 3

**Tabla 3.7** Comparativa de los resultados datos obtenidos con ZAHORÍ para los diferentes complejos de dominios SH3-cAbl con péptidos de diseño. El tiempo máximo de residencia (Tmax), el Tiempo medio de residencia (Tmedio), numero de moléculas de agua durante la trayectoria y la frecuencia de tiempo de la trayectoria que el sitio de hidratación se encuentra ocupado por una molécula de agua

Sitio de hidratación	Péptido	Tmax (ps)	Tmedio (ps)	Nº de moléculas de agua	Frecuencia de Ocupancia
W1	p41	164	16,24	181,6	44,61
	p17	94	15,80	167,2	38,34
	p7	78	16,62	134	39,51
	p0	82	15,53	190,2	40,55
W2	p41	686	24,52	81,6	42,14
	p17	206	19,18	93,6	40,54
	p7	244	22,37	69,8	41,98
	p0	196	23,72	120	54,37
W3	p41	432	34,16	94,2	56,10
	p17	280	32,30	88,8	49,99
	p7	356	34,84	57,8	46,85
	p0	322	31,78	91,6	45,03
W4	p41	2350	106,85	41,6	88,83
	p17	2106	273,72	25	95,00
	p7	3382	240,40	14	93,95
	p0	6004	572,20	12,6	97,27
W5	p41	432	66,62	58,6	53,42
	p17	714	85,97	69	66,81
	p7	272	50,48	31,6	21,67
	p0	582	83,80	59,6	61,71

Finalmente, al observar los resultados para los sitios de hidratación en la región del lazo n-Src (W1, W2, y W3), advertimos cómo los tiempos máximos de residencia de las moléculas de agua para el péptido p41 son de 2 hasta 3 veces más altos que para el resto de péptidos de diseño, aunque si comparamos la frecuencia de ocupancia no observamos diferencias significativas. Esto estaría indicando una mayor estabilización de las moléculas de agua del lazo n-Src en p41, mientras que el remplazo de moléculas de agua es el mismo para el resto de los complejos. Es necesario recordar que, dadas las características de los sitios de hidratación del lazo n-Src, más próximo a un canal que a una cavidad, las moléculas de agua son difíciles de seguir por el programa ZAHORÍ, con lo cual es probable que

cambios en la frecuencia de ocupancia de esta zona puedan quedar ligeramente enmascarados.

Para tener una información más completa, es necesario el estudio de las interacciones que establece el ligando con el dominio. En la Tabla 3.8 recopilamos los puentes de hidrógeno que se forman entre el dominio y el ligando durante un periodo superior a un 5% de la duración de la trayectoria, donde se aprecia cómo cada péptido muestra un patrón sustancialmente diferente en la distribución y frecuencia de los puentes de hidrógeno.

**Tabla 3.8** Puentes de hidrógeno establecidos entre el ligando y el dominio para cada complejo SH3-ligando. El tiempo de formación se obtiene como la media de las réplicas.

	Ligando	Dominio	P41	P17	P7	P0
<b>P-5</b>						
<b>P-4</b>	PRO:O	ASN94:ND2	36,16±28,3	23,06±24,7	42,67±20,9	21,52±22,2
<b>P-3</b>	SER:OG	ASP77:OD		19,60±17,7		
	SER:N	ASP77:OD	6,54±4,7	19,50±17,2		11,53±5,2
<b>P-2</b>	TYR:OH	SER75:OG	10,68±6,6	11,42±3,4	6,90±4,3	11,79±9,4
	TYR:O	TRP99:NE1	95,74±2,8	94,81±3,2	93,93±2,1	91,88±2,7
	TYR:OH	ASP77:OD			16,28±6,5	16,03±5,7
<b>P-1*</b>	SER:OG\PRO:O	GLU98:OE	42,34±10,2	44,02±16,4		
<b>P0</b>						
<b>P1</b>	PRO:O	TYR115:OH	47,79±16,8	67,94±20,9	24,24±12,6	47,97±23,7
	PRO:O	ASN114:ND2	36,27±9,8	22,21±21,8	63,76±12,1	28,19±17,7
<b>P2*</b>	PRO:O\LEU:O	TYR115:OH	51,87±14,9	29,71±21,1	75,84±12,5	44,31±17,3
<b>P3</b>						
<b>P4</b>						

Cuando la posición P<sub>-1</sub> está ocupada por una serina, como es el caso de los péptidos p41 y p17, la cadena lateral de este resto establece un puente de hidrógeno con la cadena lateral del GLU98, con una frecuencia similar en ambos complejos (>40%). Este puente de hidrógeno no se observa en la estructura cristalográfica, de forma que es una interacción que no

## CAPÍTULO 3

esperábamos encontrar. En los péptidos en los que existe una prolina en la posición  $P_{-1}$  no encontramos ninguna interacción, ya que la cadena lateral de la prolina no tiene capacidad para establecer puentes de hidrógeno, al contrario que la serina. No obstante, la presencia de la prolina en esta posición parece favorecer que la tirosina en la posición  $P_{-2}$  establezca un puente de hidrógeno con la cadena lateral del ASP77, con la misma frecuencia en SH3-cAbl/p7 y SH3-cAbl/p0 (~16%). Es decir, que la naturaleza del resto de la posición  $P_{-1}$  tiene un claro impacto en la red de interacciones que establece el ligando con el dominio.

Al comparar los péptidos que tienen una prolina en  $P_2$  (p41 y p7) con los péptidos con una leucina (p0 y p17) observamos que, aunque se conservan los mismos puentes de hidrógeno en la posición  $P_2$  y en la colindante  $P_1$ , las frecuencias de formación varían entre ellos, aunque es difícil correlacionar estas variaciones con la presencia de una leucina o una prolina en la posición  $P_2$ .

Es necesario destacar que es muy difícil evaluar la aportación energética de cada puente de hidrógeno, con lo que no podemos saber en qué grado contribuyen estos cambios a la entalpía de unión, y quizás sería necesario un exámen mucho mas exhaustivo para comparar el patrón de enlaces de hidrógeno entre péptidos. Pero queda patente que la modificación en estas posiciones,  $P_{-1}$  y  $P_2$ , afecta al comportamiento global del péptido. Algo que sorprende si tenemos en cuenta el análisis de los ángulos diedros del esqueleto peptídico, por lo que el cambio de flexibilidad podría traducirse en un sutil cambio en la acomodación del péptido, algo que sí facilitaría la reconfiguración de los puentes de hidrógeno intermoleculares y las interacciones con las moléculas de agua de la región del lazo n-Src.

Por otro lado, el patrón propio de los puentes de hidrógeno de cada complejo podría estar afectando a las moléculas de agua que se acomodan en el lazo n-Src. No hay que olvidar que las cadenas laterales de los residuos GLU98 y ASN94, restos que interaccionan de forma característica con cada ligando, forman parte de los sitios de hidratación de la región del lazo n-Src (W1, W2 y W3), de forma que no sería extraño que los cambios de frecuencia observados en los puentes de hidrógeno repercutieran en el comportamiento de las moléculas de agua de estos sitios, como sucede en el sitio W4 donde el establecimiento directo entre ASN114 y la prolina en la posición P<sub>2</sub> desplaza la molécula de agua en el complejo SH3-cAbl/p7.

En resumen, la mayor contribución entálpica a la unión, así como la mayor penalización entrópica que presenta el complejo SH3-cAbl/p41 respecto al resto de complejos (SH3-cAbl/p0, SH3-cAbl/p7 y SH3-cAbl/p17) podría estar originada por un aumento en la fijación de las moléculas de agua en los sitios de hidratación W1, W2 y W3. Como sugiere por un lado el aumento de los tiempos máximos de ocupancia del complejo SH3-cAbl/p41 respecto al resto de complejos, y por otro lado la información cristalográfica, donde se observa una mayor dificultad para detectar una molécula en el sitio W3 y en algunos casos también W2 y W1. Este efecto se debería a la variación en la calidad de los puentes de hidrógeno establecidos entre las moléculas de agua y los restos que forman los sitios de hidratación, provocada por ligeras diferencias en el patrón de interacciones entre el dominio y cada ligando peptídico.

## CAPÍTULO 3

### 3.2.3 Aplicación del algoritmo ZAHORÍ al estudio de patrones de hidratación en distintas subfamilias de dominios SH3

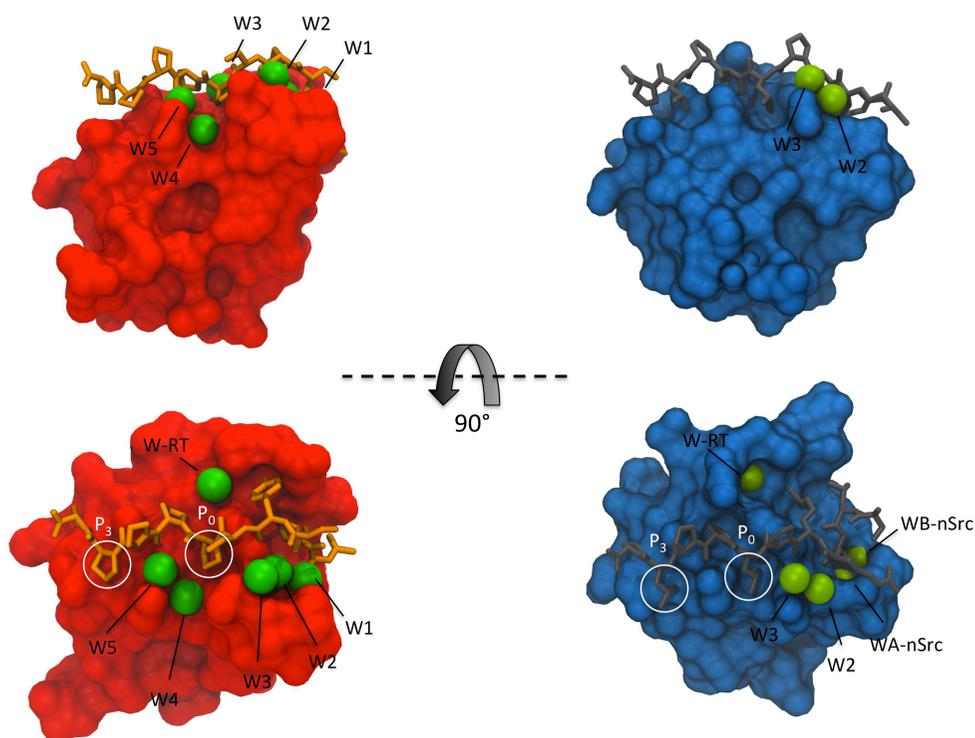
Los dominios SH3 de la subfamilia de Src son los que tienen una mayor homología con los de Abl. No obstante, el análisis de la colección de estructuras de los dominios SH3 desveló una baja coincidencia en la posición de las moléculas de agua de los dominios Src respecto al de Abl. Mientras el sitio de hidratación estructural de la hélice  $3_{10}$ , definido como W4, aparece comúnmente ocupado por una molécula de agua en Src, el resto (W1, W2, W3 y W5) aparecen con una baja frecuencia en las estructuras cristalográficas. En cambio, se encontraron otras moléculas de agua altamente conservadas que, no obstante, están ausentes en SH3-cAbl/p41. Este es el caso, por ejemplo, del sitio de hidratación estructural definido en la Sección 3.1.3.2 como WA-Src. Estas diferencias en la conservación de las moléculas de agua entre SH3-Fyn y SH3-cAbl, apuntaría a la existencia de varios patrones de sitios de hidratación que podrían ser propios para cada subfamilia de dominios SH3.

Con el objetivo de realizar un estudio comparativo de los sitios de hidratación dentro de las subfamilias de SH3, se tomaron como punto de partida los complejos de SH3-cAbl/p41 y SH3-Fyn/VSL12. El ligando p41 (APSYPPPPP), además de ser el que tiene una mayor afinidad por el dominio SH3 de Abl ( $2.1 \mu\text{M}$ ), tiene una muy baja afinidad por el dominio SH3-Fyn ( $630 \mu\text{M}$ ). De igual forma, el ligando VSL12 (VSLARRPLPLP) es el que presenta la mayor afinidad de unión por dominios SH3 de la familia Src, concretamente cSrc, cYes y Fyn ( $0.29 \mu\text{M}$ ), y muy baja por el dominio SH3-cAbl ( $2000 \mu\text{M}$ ). Ambos ligandos unen con una afinidad dirigida principalmente por una entalpía favorable, pero el ligando peptídico p41 al contrario que VSL12 y la mayoría de ligandos péptidos de los dominios SH3, no posee en la secuencia ningún resto cargado cuya posición, como mencionamos en la Introducción de

esta Tesis, determina el sentido en el que se unirá el ligando al dominio (tipo I o tipo II). Aun así, ambos péptidos unen a sus respectivos dominios con una alta especificidad y con una alta entalpia de unión,  $-91,7 \text{ kJ}\cdot\text{mol}^{-1}$  para p41 y  $-63,2 \text{ kJ}\cdot\text{mol}^{-1}$  para VSL12. Todo esto convierte a estos dos sistemas en casos ideales para el estudio comparativo de los sitios de hidratación entre subfamilias de dominios SH3.

Para el estudio comparativo se utilizó la estructura de alta resolución del complejo SH3-Fyn/VSL12, resuelta por la Dra. Ana Cámara Artigas en colaboración con nuestro grupo de investigación. Siguiendo el mismo criterio que se usó para identificar las moléculas de agua interfaciales de SH3-cAbl/p41, se identificaron 3 moléculas de agua dentro de los parámetros de búsqueda, en posiciones totalmente diferentes a SH3-cAbl/p41, dos en la región del lazo de n-Src, denominadas como WA-nSrc y WB-nSrc, así como una en el lazo RT, denominada como W-RT Figura 3.6. Como podemos observar en la Tabla 3.10 estas moléculas de agua identificadas en el sitio de unión del dominio SH3-Fyn, presentan características, valores de ASA y factores  $\beta$  en el mismo intervalo que los de las moléculas de agua del complejo SH3-cAbl/p41.

## CAPÍTULO 3



**Figura 3.6** Representación de la superficies moleculares para los dominios SH3-Fyn (azul) y SH3-cAbl (rojo) con sus respectivos ligandos, VSL12 (gris) y p41 (naranja), representados en estructura de varillas. En ambos casos se resaltan con un círculo blanco las posiciones definidas mediante el índice P<sub>3</sub> y P<sub>0</sub>. En verde se representan las moléculas de agua cristalográficas que han sido seleccionadas para el estudio comparativo y el análisis mediante el algoritmo ZAHORÍ.

**Tabla 3.9.** Comparación de ligando p41 y VSL12, resaltado la posición que ocupa en la interfaz de unión de los dominio SH3.

	P <sub>4</sub>	P <sub>3</sub>	P <sub>2</sub>	P <sub>1</sub>	P <sub>0</sub>	P <sub>-1</sub>	P <sub>-2</sub>	P <sub>-3</sub>	P <sub>-4</sub>	P <sub>-5</sub>	P <sub>-6</sub>	P <sub>-7</sub>
<b>P41</b>	PRO	PRO	PRO	PRO	PRO	SER	TYR	SER	PRO	ALA		
<b>VSL12</b>	PRO	LEU	PRO	PRO	LEU	PRO	ARG	ARG	ALA	LEU	SER	VAL

Sorprendentemente, en la estructura del complejo SH3-Fyn/VSL12 no se encontró ninguna molécula de agua ocupando el sitio de hidratación estructural de la región de la hélice 3<sub>10</sub>, el W4, así como tampoco en el W5. En la región de la hélice 3<sub>10</sub>, como se observa claramente en la Figura 3.6, la acomodación del péptido en SH3-Fyn/VSL12 hace óptimo el puente de hidrógeno entre el aspártico de la hélice 3<sub>10</sub> y la prolina en la posición P<sub>1</sub> del

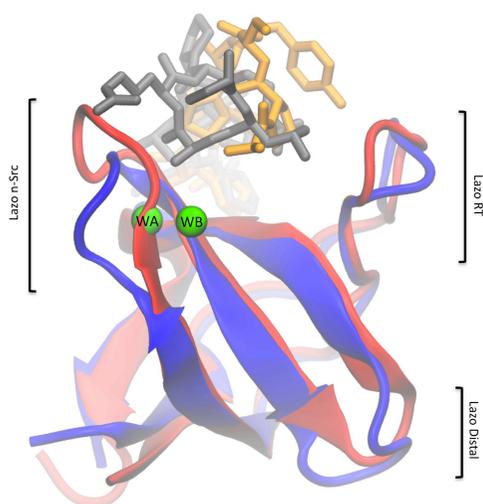
ligando, de forma que no existe espacio para acomodar una molécula de agua en W5. Además, la cadena lateral de la leucina P<sub>0</sub> de VSL12 rellena totalmente la cavidad del sitio de hidratación W4. Esta es una conformación excepcional, ya que no es común que los bolsillos hidrofóbicos del dominio SH3 estén ocupados por residuos que no sean prolina. Además, la causa general de no encontrar una molécula en el sitio de hidratación W4 en las estructuras anteriormente analizadas se debía a conformaciones anómalas del dominio o a secuencias raras de la hélice 3<sub>10</sub>, siendo la primera vez que se describe el desplazamiento de esta molécula de agua del sitio de hidratación estructural W4 por parte del ligando. Recordemos que las moléculas de agua alojadas en esta posición se encuentran fuertemente unidas al dominio, de forma que este desplazamiento podría tener un impacto directo en los parámetros termodinámicos de unión del complejo SH3-Fyn/VSL12.

**Tabla 3.10** Las moléculas de agua enterradas en el interfaz para los dominios SH3 de Fyn y cAbl en complejo con sus péptidos de alta afinidad VSL2 y p41 respectivamente. Para cada molécula e agua se indica en que región del dominio SH3 se encuentra, que sitio de hidratación ocupa, el número identificativo dentro del fichero PDB, el factor beta normalizado y Área expuesta al disolvente (ASAS).

Región	Posición	SH3-cAbl/p41			SH3-Fyn/VSL12		
		PDB	Bfactor norm	ASAS Å <sup>2</sup>	PDB	Bfactor norm	ASAS Å <sup>2</sup>
n-Src	W1	1015	0,50	5,13	-	-	-
	W2	1064	2,30	0,14	14	1,67	15,6
	W3	2103	3,20	5,55	41	2,12	18,77
	WA-nSrc	-	-	-	50	0,10	0
	WB-nSrc	-	-	-	25	2,82	1,86
3 <sub>10</sub>	W4	1082	0,75	3,03	-	-	-
	W5	1105	0,69	2,65	-	-	-
RT	W-RT	2041	3,16	14,44	35	0,01	1,41

## CAPÍTULO 3

En la región del lazo n-Src ambos complejos se diferencian radicalmente. Como describimos en la Sección 3.1.3.2, las moléculas de agua alojadas en el sitio de hidratación WA-nSrc, aparecen frecuentemente en la colección de estructuras de dominios SH3 (>60%). Esta alta frecuencia de conservación y la carencia de interacciones directas con el ligando nos indican que este sitio de hidratación entraría dentro del grupo denominado como sitios de hidratación estructurales. Por otro lado, la posición definida como WB-nSrc, aparece poco conservada en la colección de estructuras (~12%), tanto para los dominios libres como para los complejos. Su moderada conservación y su proximidad al ligando nos indican que la molécula de agua se encuentra alojada en un potencial sitio de hidratación de especificidad.



**Figura 3.7 Superposición de las estructuras de los complejos SH3-cAbl/p41 y SH3-Fyn/VSL12.** El dominio SH3-cAbl es representado en rojo y su ligando, p41, en estructura de varillas y color naranja. El dominio SH3-Fyn está representado en azul y su ligando, el péptido VSL12, en estructura de varillas y color gris. En verde se encuentran destacadas las moléculas de agua WA-nSrc y WB-nSrc

Si observamos con detalle la región del lazo n-Src en la superposición de las dos estructuras cristalográficas, SH3-Fyn/VSL12 y SH3-cAbl/p41, Figura 3.7, y advertimos que la ausencia de los sitios de hidratación en la región del lazo n-Src en SH3-cAbl/p41 se justifica claramente por un cambio en la configuración longitud del lazo n-Src, que, a su vez, favorece la acomodación de moléculas de agua en los sitios de hidratación W1, W2 y W3. Esto se corresponde con lo descrito en estudios previos realizados en nuestro grupo de investigación, ya descrito en la Sección 3.1.3. Además, el análisis de ambas estructuras permitió identificar en la estructura de SH3-Fyn/VSL12 dos moléculas de agua en posiciones homólogas a W2 y W3 de SH3-cAbl/p41. No obstante las moléculas de agua del complejo con Fyn presentan valores de exposición al solvente y de factor  $\beta$  por encima del criterio de selección, como podemos apreciar en la Tabla 3.10. No obstante, se decidió incluirlas en el análisis ya que están en posiciones equivalentes a las de los sitios W2 y W3 del complejo SH3-cAbl/p41.

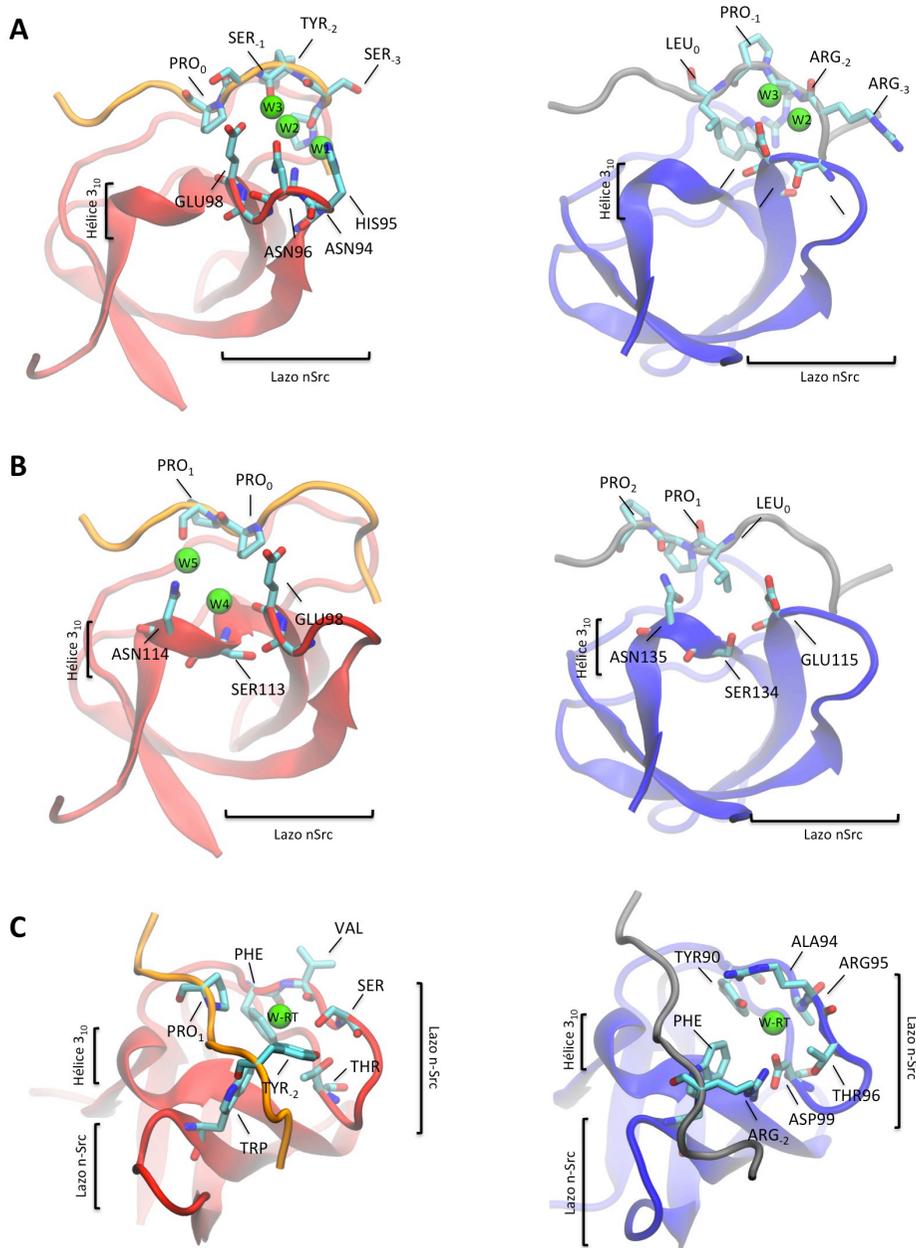
Finalmente, la molécula de agua encontrada en el complejo SH3-Fyn/VSL12 en la región del lazo RT, W-RT, se encuentra próxima a una de las argininas del ligando, Figura 3.8 En la colección de estructuras encontramos una molécula de agua ocupando una posición equivalente a W-RT en un 32% de las estructuras libres y en un 25% de las estructuras unidas a un ligando peptídico. Como comentamos en la Introducción de este Capítulo, el análisis de la colección de estructuras permitió inferir que la presencia de moléculas de agua en esta región estaba altamente correlacionado con la presencia de restos cargados. En el ligando peptídico p41 no hay ningún resto cargado y no encontramos ninguna molécula de agua en la región del lazo RT del complejo SH3-cAbl/p41 que cumpla los criterios de selección, pero la superposición de las estructuras cristalográficas de ambos complejos permitió localizar una

## CAPÍTULO 3

molécula de agua en el complejo de SH3-cAbl/p41 (la número 2041) que se encuentra en el lazo RT en una posición muy parecida a la que encontramos en SH3-Fyn/VSL12. Esta molécula de agua, como apreciamos en la Tabla 3.10, presenta unos valores de exposición al solvente y de factor  $\beta$  por encima del criterio de selección, pero la incluiremos en el estudio para comprobar los resultados entre ambos complejos.

Las estructuras cristalográficas aportan una valiosa información, pero es necesario comprobar el comportamiento dinámico de las moléculas de agua en todos los sitios de hidratación detectados para los dos complejos para tener una visión completa del sistema. La información dinámica nos permitirá confirmar si realmente los sitios de hidratación de la hélice  $3_{10}$  no se encuentran ocupados en el complejo SH3-Fyn/VSL12, comprobar la naturaleza de los potenciales sitios de hidratación identificados en la estructura cristalográfica de SH3-Fyn/VSL12, WA-nSrc, WB-nSrc y W-RT y, finalmente, evaluar si los sitios de hidratación W2 y W3 son ocupados con moléculas de agua con propiedades similares a los que se obtienen del complejo SH3-cAbl/p41.

Para comparar las propiedades dinámicas de los sitios de hidratación encontrados en el complejo SH3-cAbl/p41 respecto a los del complejo SH3-Fyn/VSL12, se realizaron 5 dinámicas independientes de cada complejo, durante un tiempo de simulación de 12 ns cada una, lo que hace un total de 60 ns por complejo, y se analizaron las propiedades de las moléculas de agua de cada posición mediante el programa ZAHORÍ, descrito en la Sección 3.2.1.



**Figura 3.8 Comparativa en detalle de las moléculas de agua del complejo SH3-cAbl/p41 (rojo) y SH3-Fyn/VSL12 (Azul).** Los residuos de cada sitio de hidratación han sido representados en estructura de varillas para cada complejo. En el Panel A se comparan los sitios de hidratación W1, W2 y W3 del lazo n-Src. En el Panel B se comparan los sitios de hidratación de la región de la hélice 3<sub>10</sub>. En el panel C se comparan los sitios de hidratación del lazo RT y el W-RT

## CAPÍTULO 3

**Tabla 3.11** Valores obtenidos del análisis con ZAHORÍ de las diferentes trayectorias obtenidas para complejos SH3-ligando. El tiempo máximo de residencia ( $T_{max}$ ), el Tiempo medio de residencia ( $T_{medio}$ ), numero de moléculas de agua durante la trayectoria y la frecuencia de tiempo de la trayectoria que el sitio de hidratación se encuentra ocupado por una molécula de agua

Región	Posición	SH3-cAbl/p41			SH3-Fyn/VSL12		
		$T_{max}$ (ps)	$T_{medio}$ (ps)	$T_{occ}$ (%)	$T_{max}$ (ps)	$T_{medio}$ (ps)	$T_{occ}$ (%)
n-Src	W1	164	16,24±28	44,62±10	-	-	-
	W2	686	24,53±15,7	42,14±0,9	150	17,24±7	19,80±23,1
	W3	432	34,17±8,5	56,10±15,1	206	22,93±2,4	62,31±1,4
	WA	-	-	-	7544	1918,02±1397	98,90±0,5
	WB	-	-	-	694	43,20±11,8	77,92±10,4
3 <sub>10</sub>	W4	2350	106,85±387	88,84±3,9	108	6,51±14,8	0,59±3,7
	W5	432	66,63±15,7	53,43±11,6	316	16,27±16,5	2,38±1
RT	W-RT	196	24,35±8,5	9,64±5,2	4016	257,00±119,9	67,21±30,6

En la Tabla 3.11 se muestran los resultados obtenidos del promedio de los valores obtenidos para todas las réplicas. Al comparar los datos obtenidos para los sitios de hidratación de la hélice 3<sub>10</sub> observamos que, tanto W4 como W5 apenas se encuentran ocupados en el complejo SH3-Fyn/VSL12. Esto es debido a que el péptido VSL12 se encuentra firmemente encajado en el interfaz de SH3-Fyn a lo largo de la trayectoria, no dejando estos sitios disponibles para ser ocupados por moléculas de agua. Esto contrasta notablemente con el comportamiento de las moléculas de agua alojadas en las mismas posiciones en el complejo SH3-cAbl/p41, donde las moléculas de agua tienen frecuencias de ocupancia de entre el 70% y el 90% del tiempo de simulación y tiempos de residencia extremadamente altos. Con el objetivo de poder explicar estas enormes diferencias se analizaron los puentes de hidrógeno que se forman entre el dominio y el ligando en cada complejo. Los resultados de este análisis se muestran en la Tabla 3.13. Como se observa, en el complejo SH3-Fyn/VSL12 la cadena lateral de la asparraguina de la hélice 3<sub>10</sub> (PSNY) se encuentra cerca de un 100% del tiempo de simulación formando un puente de hidrógeno con el oxígeno carbonilo de la prolina de la posición P<sub>0</sub> del ligando, imposibilitando que una molécula de agua entre en el sitio de

hidratación W5. Recordemos que la molécula alojada en este sitio se aloja entre ambos residuos y establece puentes de hidrógeno con ambos.

En cambio, la ausencia de moléculas de agua en el sitio de hidratación estructural W4 parece causal, principalmente por el impedimento estérico de la cadena lateral de la leucina del ligando en posición P<sub>0</sub>. Para comprobar si la leucina P<sub>0</sub> del ligando era la responsable de expulsar la molécula de agua alojada en el sitio de hidratación estructural W4, se creó un modelo a partir de la estructura cristalográfica del complejo SH3-Fyn/VSL12, siguiendo el protocolo descrito en la Sección 5.4 de Materiales y Métodos, en el que la leucina P<sub>0</sub> del péptido fue substituida por una prolina. Este modelo fue utilizado para generar 5 nuevas trayectorias de dinámica molecular independientes a 298 K durante 12 ns cada una, que se analizaron de nuevo con ZAHORÍ. Como se observa en la Tabla 3.12, la mutación del ligando no produce una variación significativa en los datos para el sitio W5, lo que nos indica que la expulsión de la molécula de agua del sitio de hidratación W5 está más relacionada con la acomodación general del péptido que con la naturaleza particular de un resto específico. No obstante, sorprendentemente, los valores de tiempo de residencia y frecuencia de ocupancia para el sitio de hidratación estructural W4 aunque experimentan un aumento significativo como consecuencia de la eliminación de la cadena lateral de leucina, siguen mostrando valores para estos parámetros muy por debajo de lo obtenido en el complejo SH3-cAbl/p41. Esto sugiere que la leucina del ligando no es la única responsable de este importante cambio.

## CAPÍTULO 3

**Tabla X.12** Valores obtenidos del análisis con ZAHORÍ durante la trayectoria del complejo SH3-Fyn/VSL12 comparados por SH3/VSL12\_L6P. El tiempo máximo de residencia ( $T_{max}$ ), el Tiempo medio de residencia ( $T_{medio}$ ), número de moléculas de agua durante la trayectoria y la frecuencia de tiempo de la trayectoria que el sitio de hidratación se encuentra ocupado por una molécula de agua

Región	Posición	SH3-Fyn/VSL12			SH3-Fyn/VSL12_L6P		
		$T_{max}$ (ps)	$T_{medio}$ (ps)	$T_{occ}$ (%)	$T_{max}$ (ps)	$T_{medio}$ (ps)	$T_{occ}$ (%)
3 <sub>10</sub>	W4	108	6,51	0,59	610	90,85	16,82
	W5	316	16,27	2,38	298	16,89	4,25

Recordemos que no existe ninguna diferencia en la secuencia de este bolsillo de la región de la hélice 3<sub>10</sub> entre ambos dominios, y que en la colección de estructuras el sitio de hidratación W4 está ocupado por una molécula de agua en un muy elevado porcentaje de las estructuras (más del 70%), de forma que llama mucho la atención que en el complejo SH3-Fyn/VSL12\_L6P, en el que no existe el impedimento de la leucina ocupando el sitio de hidratación, no se obtengan valores de residencia y ocupancia en el mismo orden que los obtenidos para el complejo SH3-cAbl/p41. Con el objetivo de comprobar si este comportamiento respondía a un cambio conformacional del dominio se analizó el área accesible al solvente de los residuos que forman el sitio de hidratación W4 (GLU98, PRO112, ASN114 y SER115) mediante el programa NACCESS [117], usando los parámetros por defecto, a lo largo de las trayectorias de SH3-cAbl/p41 y SH3-Fyn/VSL12\_L6P. Sorprendentemente, el valor promedio registrado para este último fue de  $46,63 \pm 11,4 \text{ \AA}^2$ , mientras que para el complejo SH3-cAbl/p41 la media casi duplicaba la anterior,  $78,59 \pm 13,5 \text{ \AA}^2$ . Esto indica que la conformación del sitio de hidratación para el complejo de SH3-cAbl es más abierta, permitiendo que se acomode una molécula de agua, mientras que en SH3-Fyn este se encuentra más cerrado. Esto estaría relacionado con la excepcional forma que tiene el péptido de acomodarse en el interfaz de SH3-Fyn. Recordemos que VSL12 acomoda leucinas en las posición P<sub>3</sub> y P<sub>0</sub>, algo totalmente anómalo, en lugar de

prolinas. Al mismo tiempo la colación del ligando sobre el lazo n-Src, varia, resultando también atípica, provocando un cambio conformacional en el dominio que sería el principal causante de esta modificación del sitio de hidratación W4.

En la región del lazo n-Src, tanto el sitio de hidratación WA-nSrc como el WB-nSrc observados en los complejos de SH3-Fyn/VSL12 no están nunca ocupados en el complejo SH3-cAbl/p41. En este último complejo no hay espacio para acomodar moléculas de agua en esa posición, ya que, como hemos indicado anteriormente, las diferencias de longitud y secuencia del lazo n-Src conjuntamente con la secuencia del péptido imposibilitan la acomodación de moléculas de agua en la región donde están localizados los sitios de hidratación WA-nSrc y WB-nSrc. Por el contrario, en SH3-Fyn/VSL12 ambos sitios de hidratación están ocupados por moléculas de agua el 98% (WA-nSrc) y el 77% (WB-nSrc) del tiempo de simulación. Los resultados obtenidos para las moléculas de agua que se alojan en el sitio de hidratación WA-nSrc de SH3-Fyn/VSL12 están en el mismo orden de magnitud que los que se registran para el sitio W4 del complejo SH3-cAbl/p41, con tiempos máximos de residencia bastante grandes, por encima del nanosegundo, y tiempos de ocupancia por encima del 80%. Esto, sumado al alto grado de conservación de las moléculas de agua en esta posición en la colección de estructuras de dominios SH3, presentes hasta en un 67% de las estructuras de la colección libres y un 35% de las estructuras unidas a ligando, y al hecho de que éstas no se encuentran interaccionando directamente con el ligando, nos indicara que el sitio WA-nSrc pertenece a los sitios de hidratación que hemos denominado como estructurales.

En cuanto a la molécula WB-nSrc, los tiempos de residencia registrados para el complejo SH3-Fyn/VSL12 son del orden de los  $10^3$  ps y se encuentra

## CAPÍTULO 3

interaccionando con la VAL1 del ligando. Como hemos comprobado en la colección de estructuras, esta posición se encuentra poco conservada, apareciendo alrededor de un 12%, tanto para las estructuras de complejos como las de los dominios libres. De forma que las características de este sitio de hidratación se encuentran entre las que definimos como sitios de especificidad.

**Tabla 3.13** Puentes de hidrógeno establecidos entre el dominio SH3 y el ligando para cada complejo. Los tiempos son el tanto por cierto del tiempo que está formado el puente de hidrógeno respecto al tiempo total de la trayectoria.

Sitio	SH3-cAbl/p41			SH3-Fyn/VSL12		
	Ligando	Dominio	T (%)	Ligando	Dominio	T (%)
<b>P<sub>-5</sub></b>						
<b>P<sub>-4</sub></b>	PRO:O	ASN94:ND2	36,16±28,2	ALA:O	TRP118:N	84,18±4,3
				ALA:O	ASP117:N	38,33±11,5
<b>P<sub>-3</sub></b>	SER:N	ASP77:OD	6,54±4,7	ARG:NE	GLU115:O	26,47±9,4
<b>P<sub>-2</sub></b>	TYR:OH TYR:O	SER75:OG TRP99:NE1	10,68±6,6 95,74±2,8	ARG:NH	ASP99:ODX	78,7±12,1
				ARG:NH	ASP99:ODX	82,5±20,3
				ARG:NH	THR96:OG	9,48±12,3
				ARG:O	TRP118:NE	31,07±19,6
<b>P<sub>-1</sub></b>	SER:OG	GLU98:OE	42,34±10,2	PRO:O	TRP118:NE	75,13±15,7
<b>P<sub>0</sub></b>						
<b>P<sub>1</sub></b>	PRO:O	TYR115:OH	47,79±16,8	PRO:O	ASN135:ND	90,29±9,1
	PRO:O	ASN114:ND 2	36,27±9,8			
<b>P<sub>2</sub></b>	PRO:O	TYR115:OH	51,87±14,9	PRO:O	TYR136:OH	17,72±35,1
<b>P<sub>3</sub></b>						
<b>P<sub>4</sub></b>						

También se observa cómo los sitios de hidratación W2 y W3 de SH3-Fyn/VSL12, presentan unos valores de tiempo de residencia y frecuencia de ocupancia más bajos que los de SH3-cAbl/p41, en especial el W2, con una frecuencia de ocupancia por debajo del 20% y un tiempo máximo de residencia 6 veces menor que el que obtenemos en SH3-cAbl/p41. El sitio de hidratación W3 en el complejo Fyn presenta una mayor frecuencia de

ocupancia que W2, pero un tiempo máximo de residencia de solamente 206 ps. Esto se debe a que este sitio de hidratación se encuentra más expuesto en SH3-Fyn/VSL12 que en SH3-cAbl/p41. Todo esto, sumado a que no se encuentran moléculas de agua en esta posición para el resto de estructuras cristalográficas de dominios SH3 de la subfamilia Src, nos indica que los sitios definidos como W2 y W3 tienen menor protagonismo en el mecanismo de unión en el complejo SH3-Fyn/VSL12 que en el SH3-cAbl/p41. Además, si observamos las estructuras cristalográficas de ambos complejos en las Figura 3.6 y 3.8, podremos apreciar cómo el ligando VSL12 en el complejo con SH3-Fyn está más desplazado sobre el lazo n-Src que en el ligando p41 en el complejo con SH3-cAbl. Asimismo, si comparamos los puentes de hidrógeno que establece el ligando con el dominio en ambos complejos a lo largo de las trayectorias, Tabla 3.13, queda patente un aumento de las interacciones directas entre el lazo n-Src y el ligando VSL12 con respecto al complejo SH3-cAbl/p41, en el que, como podemos apreciar con detalle en la Figura 3.8, donde se comparan con detalle la región del lazo n-Src de ambos complejos, estas interacciones están mediadas principalmente por las moléculas de agua alojadas en los sitios de hidratación W1, W2 y W3.

Finalmente, en la región del lazo RT, los datos para el sitio de hidratación W-RT dejan patente que las moléculas de agua que se alojan en esta posición, para el complejo SH3-Fyn/VSL12 permanecen fuertemente unidas, estando caracterizadas por tiempos de residencia por encima del nanosegundo y una frecuencia de ocupancia cercana al 70%. Esto contrasta claramente con lo descrito para el complejo SH3-cAbl/p41, donde registramos tiempos muy inferiores (>200ps) y una frecuencia de ocupancia que se encuentra por debajo del 10%. Esto corrobora y refuerza las observaciones realizadas en los estudios previos con la base de datos estructural, donde se

### CAPÍTULO 3

asoció la presencia de las moléculas de agua del lazo RT con la mediación de interacciones entre restos cargados del ligando y el dominio. Como podemos apreciar con más detalle en la Figura 3.8, en el complejo SH3-cAbl/p41 no encontramos ningún resto cargado en la secuencia del péptido y del lazo RT, por lo que no se llegan a establecer este tipo de interacciones, al contrario que lo sucede con el ligando VSL12, con varios restos cargados, entre ellos la arginina en posición P-2 que establece interacciones con el lazo RT y con las moléculas de agua que se alojan en el sitio de unión W-RT. Atendiendo a estas observaciones, el sitio de hidratación W-RT de SH3-Fyn/VSL12 podría clasificarse dentro de la categoría que hemos definido como sitios de hidratación de especificidad, al presentar tiempos de residencia muy altos, una conservación moderada y cuyas moléculas de agua están mediando interacciones entre el ligando y el dominio.

En resumen, el complejo SH3-Fyn/VSL12 no tiene el mismo patrón de sitios de hidratación que el SH3-cAbl/p41. Los factores que determinan estas diferencias son numerosos: desde la secuencia y estructura del propio dominio, como en el caso de los sitios WA-nSrc y WB-nSrc, hasta la secuencia del péptido y las interacciones intermoleculares que éste establezca con el dominio (W5, W-RT, W1, W2 y W3) pasando por posibles cambios conformacionales inducidos por la formación del complejo (W4). Además, aunque ambos complejos presentan una energía de Gibbs de unión muy exotérmica dirigida por una fuerte entalpía favorable, la magnitud de la contribución de las moléculas de agua al carácter exotérmico de la unión no parece ser igual para ambos complejos. En SH3-Fyn/VSL12 hemos identificado un mayor número de interacciones directas entre el ligando y el dominio. En la Tabla 3.13 se presentan los puentes de hidrógeno intermoleculares obtenidos para cada complejo y las frecuencias de ocupancia

de dichos puentes de hidrógeno. Estos datos indican que, efectivamente, en el complejo SH3-Fyn/VSL12 se están estableciendo un mayor número de puentes de hidrógeno intermoleculares, especialmente concentrados en la región de especificidad definida por los lazos RT y n-Src. El lazo n-Src, se encuentra en una configuración más cerrada sobre el ligando en SH3-Fyn/VSL12, estableciendo más interacciones directas, mientras que en el complejo SH3-cAbl/p41 el lazo n-Src establecería estas interacciones gracias a la mediación de las moléculas de agua de los sitios de hidratación W1, W2 y W3, sitios de hidratación que, o bien no se encuentran ocupados en SH3-Fyn/VSL12, o bien presentan valores muy inferiores de residencia y ocupancia con respecto al complejo SH3-cAbl/p41. Es interesante señalar que el ligando VSL12 fue identificado a partir de librerías de expresión en fagos, por lo que su secuencia está fuertemente optimizada para unir a los dominios de Src con una alta afinidad, por lo que es lógico pensar las interacciones directas entre el ligando y el dominio puedan estar más optimizadas en este complejo.

### 3.3 CONCLUSIONES

Del análisis dinámico de las moléculas de agua enterradas en la interfase de unión en los dominios SH3 se desprenden las siguientes conclusiones:

1. Hemos desarrollado un algoritmo para el análisis del comportamiento de los sitios de hidratación de forma automática y eficiente. Este algoritmo ha sido validado para su aplicación al estudio de hidratación interfacial en dominios SH3, abriendo la puerta al análisis a gran escala sitios de hidratación en proteínas.

## CAPÍTULO 3

2. La aplicación de este algoritmo nos ha permitido obtener información dinámica de los sitios de hidratación en dominios SH3, no accesible a partir de las estructuras cristalográficas, de gran relevancia para la adecuada comprensión de las interacciones de dominios SH3 con sus ligandos.
3. Hemos establecido la existencia de distintos patrones de hidratación en la familia de dominios SH3, no sólo relacionados con la región de especificidad sino también asociados a regiones conservadas, como la hélice 3<sub>10</sub>, que están determinados por la conformación global que adopta el ligando en el sitio de unión.
4. Hemos comprobado que, a pesar de la universalidad del doble mecanismo de unión en dominios SH3, puede existir una gran variabilidad en cuanto a la magnitud de la contribución de las moléculas de agua interfaciales a la energética de unión

---

**4 ANÁLISIS DE LA DISTRIBUCIÓN CONFORMACIONAL DEL  
TÁNDEM SH32 DE cABL Y SU MODULACIÓN POR  
INTERACCIONES INTRAMOLECULARES**

CAPÍTULO 4

## 4.1 Regulación de la quinasa de tirosina cAbl

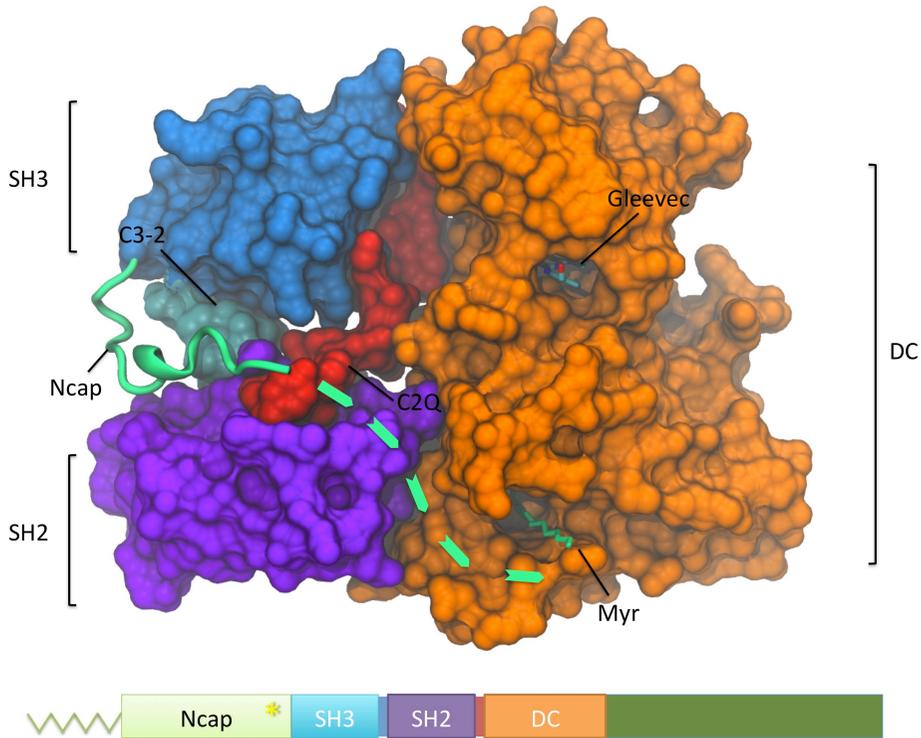
Las quinasas forman parte de la maquinaria de señalización celular, siendo las proteínas encargadas de transferir un grupo fosfato de una molécula de ATP a un residuo de una proteína diana específica, modificando su comportamiento, conformación o localización. Estos enzimas han evolucionado para desarrollar su cometido como interruptores moleculares con la mayor eficiencia posible, lo que los convierte en unas maquinarias moleculares muy complejas [118].

Las quinasas de tirosina son el grupo más diverso y amplio dentro de las quinasas, estando implicadas en todos los procesos celulares. En consecuencia, muchas patologías como el cáncer o algunas enfermedades neurodegenerativas están relacionadas con su desregulación de quinasas de tirosina, lo que las convierte en dianas atractivas para el desarrollo de nuevas terapias [119]. La quinasa de tirosina cAbl fue una de las primeras en identificarse, fue caracterizada a partir de su forma viral y ha resultado ser un importante proto-oncogén [120-123]. Gracias a décadas de trabajos sobre este sistema, sabemos que cAbl es de las quinasas de tirosina más ampliamente conservadas. Se encuentra presente en todos los metazoos, donde está implicada en los procesos de reorganización del citoesqueleto y reparación del ADN [124-128], jugando un papel vital en el desarrollo de los organismos donde se encuentra, como lo demuestra la baja viabilidad de los individuos defectivos en el gen que codifica para cAbl [129, 130].

Todos los miembros de la familia de cAbl comparten la misma distribución de dominios a lo largo de la secuencia: un extremo N-terminal o Ncap, carente de dominios o estructura, seguido por un dominio de reconocimiento de secuencias ricas en prolina (SH3), un dominio de

## CAPÍTULO 4

reconocimiento de fosfotirosina (SH2), que se conecta a través de una larga secuencia al dominio catalítico (SH1) y, finalmente, un extenso extremo C-terminal donde encontramos varios dominios de unión al citoesqueleto, secuencias señal para la nucleación, secuencias ricas en poliprolina y otras de reconocimiento de ADN[21, 131]. Figura 4.1



**Figura 4.1 Estructura de la quinasa de tirosina cAbl (2FO0).** La superficie está coloreada para resaltar las diferentes regiones que la componen: en azul está representado el dominio SH3, en púrpura el dominio SH2, en naranja se resalta el dominio catalítico, en rojo la región conectora C2Q, en cian la región conectora C3-2, en verde se resalta la región Ncap y, finalmente, el ácido miristoílico se destaca en línea de puntos. Se ha simulado la región de Ncap que no ha sido resuelta en la estructura cristalográfica. En la cavidad catalítica se encuentra el inhibidor Gleevec. En la parte inferior se representa un esquema de la organización de los dominios de cAbl, con la misma coloración que la estructura. Con un asterisco se resalta la fosfoserina en posición de la 69 y en verde se destaca la región C-terminal, donde encontramos dominios de unión al citoesqueleto, secuencias señal para la nucleación y otras de reconocimiento de ADN.

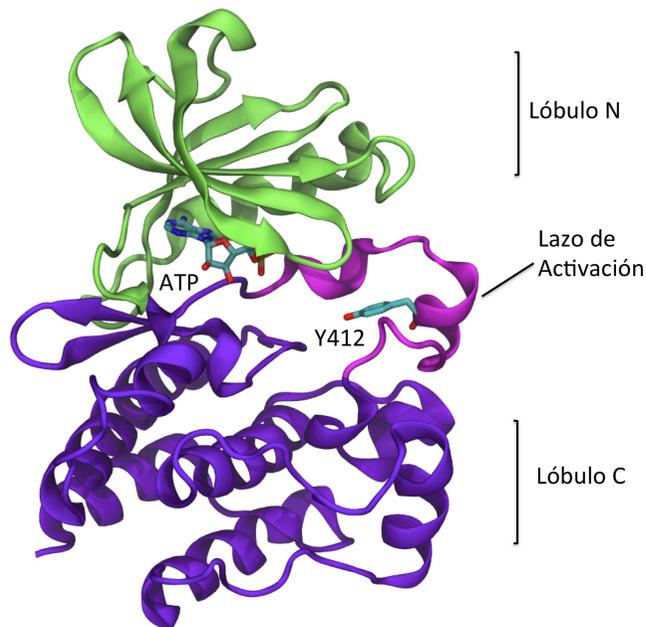
La naturaleza modular de cAbl responde a un complejo mecanismo de funcionamiento, el cual aún no ha sido completamente detallado. La actividad enzimática de cAbl está regulada por un gran número de mecanismos, como son la fosforilación de algunos de sus restos de tirosina, sus interacciones intramoleculares, su interacción con una gran gama de otras proteínas e, incluso, su propia localización celular [132]. Esto sugiere que el funcionamiento de la actividad de cAbl depende de que todos los elementos de la proteína, aunque de distinta naturaleza, colaboren entre sí para llevar a cabo su función. Esta quinasa de tirosina no es un mero interruptor con dos estados, activo e inactivo, sino que cAbl mantiene un dinámico y complejo equilibrio de estados que es afectado de forma diferencial dependiendo del estímulo y del contexto en el que se encuentra. Esto permite a cAbl responder de forma adecuada a una gran variedad de señales [10, 27].

En la célula, cAbl, se mantiene autoinhibida, principalmente gracias a una serie de factores, tanto intrínsecos como extrínsecos [23, 132]. La sobreexpresión de cAbl en células no produce ningún cambio en la homeostasis celular, pero por otro lado, la introducción de formas constitutivamente activas de cAbl ha demostrado tener la capacidad de transformar células sanas en cancerígenas [133], lo que pone de manifiesto la importancia de la correcta regulación de cAbl para la viabilidad celular. En humanos, la translocación génica del gen BCR con la del gen ABL produce una proteína quimérica, BCR-ABL, que se encuentra constitutivamente activa, siendo la principal causa de la leucemia crónica mieloide [123].

Los fármacos existentes para tratar la leucemia crónica mieloide (CML), como Gleevec, actúan bloqueando el centro activo del dominio catalítico, pero no es raro que los pacientes dejen de responder al tratamiento por el desarrollo de resistencia [134]. El análisis de las formas resistentes reveló que,

## CAPÍTULO 4

al contrario de lo que se esperaba, no todas las mutaciones que inducían resistencia se encontraban en el dominio catalítico, sino que aparecían agrupadas en el dominio SH3, en el dominio SH2 y en las regiones interdominio [134-136]. Esto, sumado a la baja especificidad que tienen los fármacos, generalmente basados en homología con la molécula de ATP, ha motivado numerosos estudios para comprender los mecanismos de regulación de esta quinasa de tirosina que permitan el desarrollo de nuevas terapias anticancerígenas [137-139].

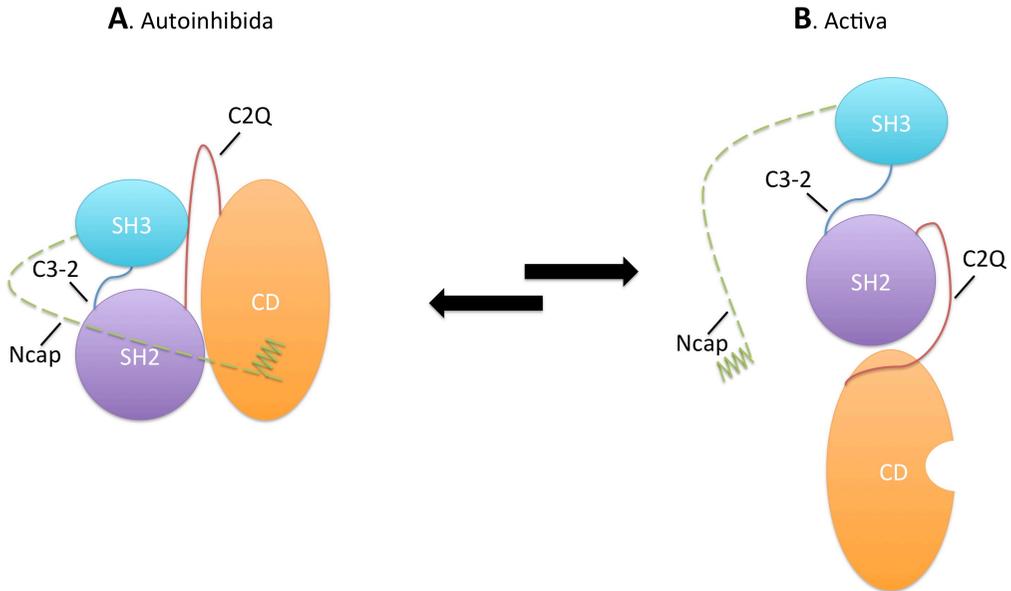


**Figura 4.2 Detalle de la estructura del dominio catalítico de la quinasa de tirosina cAbl.** Se ha resaltado en verde el lóbulo N, y en púrpura el lóbulo C. Se destaca en estructura de varillas la tirosina 412, responsable de la activación, y también la región colindante, conocida como lazo de activación (en color morado). En la cavidad catalítica se destaca una molécula de ATP en estructura de varillas

#### 4.1.1 Regulación de cAbl por parte de los dominios modulares SH3 y SH2

El análisis de las estructuras cristalográficas de cAbl ha permitido confirmar que la conformación autoinhibida guarda una alta homología con las quinasas de tirosina de la subfamilia Src y, al igual que en estas, el tándem de los dominios modulares (SH32) mantiene una serie de contactos intramoleculares con el dominio catalítico, limitando su libertad conformacional, necesaria para regular su actividad enzimática. Como se observa en la Figura 4.2, el dominio catalítico de cAbl es una estructura formada por dos lóbulos, denominados lóbulo C y lóbulo N. Entre ambos lóbulos se encuentra el centro activo, cuya activación depende de la fosforilación de la TYR439 y de un cambio conformacional que permite la posición idónea de los elementos para llevar a cabo la actividad enzimática y que, a su vez, permite acomodar la molécula de ATP y la secuencia diana [23, 27]. Esta conformación está restringida en la forma autoinhibida de cAbl debido a las interacciones intramoleculares específicas entre los dominios modulares, el dominio catalítico y posiblemente la región N-terminal [23, 140], permaneciendo todo el conjunto en una conformación muy característica y compacta. Una vez activada, los dominios modulares dejan de interactuar con el dominio catalítico, lo que permite a éste último que pueda llevar a cabo su actividad enzimática. De esta una conformación activa no ha sido posible conseguir información estructural detallada hasta la fecha. No obstante, gracias a técnicas biofísicas como la dispersión de rayos X de ángulo bajo (*Small Angle X-ray Scattering* ó SAXS), poseemos algunas pistas de los movimientos y cambios conformacionales relacionados con la activación de cAbl, Figura 4.3 [32].

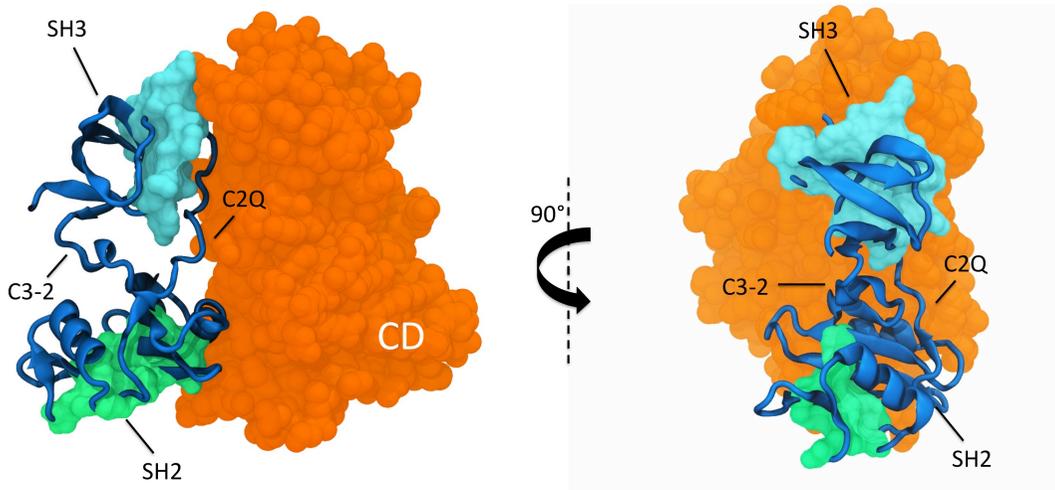
## CAPÍTULO 4



**Figura 4.3 Esquema de la organización estructural de la tirosina quinasa cAbl.** En el panel A; Organización de los dominios de cAbl cuando se encuentra autoinhibida según la información cristalográfica disponible. En el panel B; Organización de los dominios más probable cuando la quinasa de tirosina cAbl se encuentra activa de acuerdo con la estructura de baja resolución obtenida por SAXS (dipersión de rayos X de ángulo bajo) [32].

Respecto de la conformación autoinhibida, como podemos observar en la Figura 4.4, cómo el dominio SH2 interacciona directamente con el lóbulo C del dominio catalítico, dejando su sitio de reconocimiento de secuencia parcialmente oculto, mientras que el dominio SH3 se une a la secuencia interdominio que enlaza el dominio SH2 con el dominio quinasa, denominada C2Q y comprendida entre los residuos PRO237 a la LYS253. Esta secuencia queda enterrada entre el sitio de reconocimiento de secuencias ricas en prolina del dominio SH3 y el lóbulo N del dominio catalítico [32]. Esta unión, aunque débil, parece ser uno de los componentes críticos de la regulación de la actividad de cAbl, como sugieren la activación de cAbl en presencia de proteínas con motivos poliprolina [30, 31, 141] y los estudios de actividad con mutantes donde se ha eliminado el dominio SH3 [142] o se ha disminuido la

afinidad de la secuencia C2Q mediante la mutación de las prolinas 242 y 249 de la secuencia conectora por ácido glutámico [143]. Estas prolinas son responsables de que esta secuencia adopte una conformación cercana a la hélice PPII típica de las secuencias reconocidas por SH3 [22].



**Figura 4.4** Detalle de la estructura autoinhibida de cAbl a partir del PDB con código 2FO0. En azul se representa la estructura del tándem de los dominios modulares (SH3-SH2). La superficie de las interfaces de unión de ambos dominios modulares ha sido resaltada en color cian y verde, para el dominio SH3 y para el SH2 respectivamente. El dominio catalítico ha sido representado en naranja.

Otros indicios de la importancia de la interacción del dominio SH3 en el mecanismo de activación de cAbl provienen de la posibilidad de bloquear la interacción del dominio SH3 con la secuencia de conexión SH2-quinasa (C2Q) mediante fosforilación [24]. Hasta la fecha se han descrito cuatro puntos de fosforilación en esta región de cAbl. La fosforilación en cualquiera de estos cuatro restos, parece tener efectos positivos en la activación de cAbl, seguramente al bloquear el reconocimiento de la secuencia C2Q por parte del dominio SH3, lo que impediría que la quinasa de tirosina adopte la conformación autoinhibida. Sorprendentemente, ninguna de estas posiciones

## CAPÍTULO 4

es accesible en la conformación inactiva de cAbl, lo que sugiere que los eventos de fosforilación deberían ocurrir después de la activación de la quinasa por otro estímulo. Por otro lado, la existencia de varios sitios sensibles a fosforilación, podría responder a diferentes mecanismos. Por ejemplo, encontramos dos ( TYR89 y TYR134), en el sitio de unión del dominio SH3, y dos más (TYR245 y TYR251 ) en la secuencia de conexión del dominios SH2 con el dominio catalítico (C2Q). Aunque todas las fosforilaciones impedirían que el dominio SH3 volviera a la conformación autoinhibida, la fosforilación en el sitio de unión de SH3 debe tener un papel diferente, ya que a su vez imposibilitaría el reconocimiento de substratos por parte de este dominio modular [24]

La activación *in vivo* de cAbl mediante el dominio SH3 en la célula, como la de todas las proteínas implicadas en los procesos de señalización, debe ser transitoria, de forma que una vez cAbl ha fosforilado el sustrato debe volver rápidamente a la conformación inhibida. Existen evidencias que indican que en este proceso la secuencia que conecta los dominios modulares, C3-2 (comprendida entre los residuos VAL138 al GLU142), aunque no presenta estructura secundaria es un elemento clave que podría estar actuando como un muelle entre la conformación autoinhibida y la activa, modulando el cambio de orientación relativa de los dominios modulares [58].

Los primeros indicios de la importancia de esta secuencia interdominio se derivaron de un estudio de mutagénesis aleatoria en la secuencia de cAbl, donde se detectó que la substitución de la SER140 del segmento C3-2 por una isoleucina producía la activación de cAbl, tanto *in vivo* como *in vitro*. Un estudio posterior en los que se mutaron a glicina la SER140 y sus residuos colindantes, LEU141 y GLU142, a glicina, produjo una forma que tenía una actividad aún mayor que la forma S140I. En ambos casos, aunque cAbl se

encontraba activa, la capacidad del dominio SH3 de unir motivos poliprolina no sólo no se vió afectada, sino que parecía aumentar [58]. Incluso, se han descrito formas de BCR-ABL resistentes a Gleevec que presentan mutaciones en la SER140 [134]. Por consiguiente, se ha propuesto que la secuencia C3-2 se encuentra involucrada en los mecanismos que permiten la liberación del dominio SH3 de la secuencia C2Q para permitir el reconocimiento de ligandos exógenos, algo que se perturbaría al introducir las glicinas [58].

El papel del dominio SH2 en el mecanismo de regulación esta aún por esclarecer [21, 23]. Aunque las interacciones que establece el dominio SH2 con el dominio catalítico son importantes para la correcta regulación de cAbl, parece que su misión principal está más relacionada con el reconocimiento de sustratos para el dominio catalítico. Como hemos mencionado, la estructura autoinhibida de cAbl conserva una alta homología con la de los miembros de la subfamilia de quinasas de tirosina de Src. No obstante, aunque encontramos numerosas similitudes entre ambas, existen pequeñas, aunque importantes, diferencias. Por ejemplo, en las quinasas de tirosina de Src, el papel del dominio SH2 se manifiesta porque el extremo C-terminal contiene una secuencia con una tirosina, cuya fosforilación permite que ésta sea reconocida por el propio dominio SH2, de forma que el extremo C-terminal se repliega sobre la estructura estabilizando la conformación autoinhibida. La acción de fosfatasas sobre esta posición es un conocido interruptor de la actividad de la quinasas de tirosina de la subfamilia de Src.

Sin embargo, en cAbl este motivo no existe [144, 145]. La ausencia de la interacción del C-terminal con el dominio SH2 podría estar compensada por un aumento de la afinidad de la secuencia C2Q por el dominio SH3, tal como proponen Engen y colaboradores a partir de los resultados de experimentos de intercambio hidrógeno/deuterio y de espectrometría de masas con

## CAPÍTULO 4

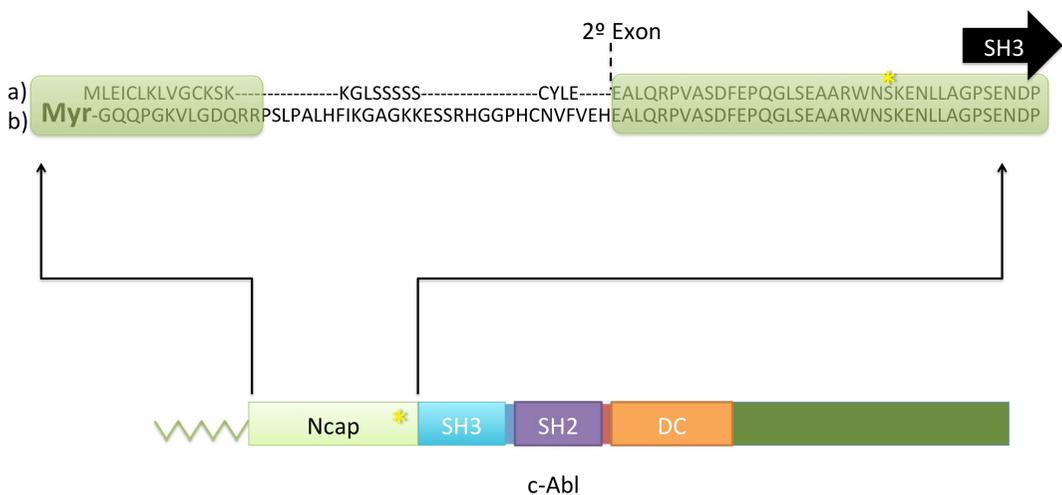
diferentes construcciones de la quinasa cAbl [26]. Otro sustituto posible de las interacciones del C-terminal con el dominio SH2 en la regulación de la actividad quinasa de cAbl, es la secuencia Ncap, que no existe en las quinasas de tirosina de la familia de Src [143]. Esto explicaría por qué la quimera oncogénica BCR-ABL, que carece de esta región Ncap, se encuentra constitutivamente activa, tanto *in vitro* como *in vivo*.

### 4.1.2 El papel de la región N-terminal (Ncap) en la regulación de cAbl

La región N-terminal de cAbl (Ncap) se transcribe en dos isoformas, cAbl-1a y cAbl-1b, por empalme alternativo (*alternative splicing*) del primer exón. La isoforma cAbl-1b se caracteriza por tener una mayor longitud que la cAbl-1a (15 residuos más) y por la adición de un grupo miristoílo en la posición 2 por modificación post-transcripcional. Esta región carece de estructura en ambas isoformas, habiéndose comprobado por delección que la región central de Ncap no es necesaria para la inactivación de cAbl *in vitro*. Únicamente la región adyacente al dominio SH3 (restos 61-82) y los primeros residuos (restos 1-15) son necesarios para mantener la quinasa regulada [23, 24]. En la Figura 4.5 muestra la secuencia del primer exón de ambas isoformas, alineadas y resaltando con un cuadrado verde aquellas regiones que son indispensables para la regulación de cAbl.

En un principio se pensó que el grupo miristoílo y la región Ncap se encargaban, al igual que en la subfamilia Src, de anclar cAbl a la membrana plasmática, pero, gracias a un completo estudio estructural de la isoforma cAbl-1b, se demostró que el grupo miristoílo se encuentra interaccionando con el dominio catalítico, insertado en una cavidad del lóbulo C. Aunque no fue posible determinar la posición de la mayor parte de la secuencia Ncap debido a su naturaleza desestructurada, se propuso que esta unión induciría un

cambio conformacional que facilitaría el acoplamiento del dominio SH2 al dominio catalítico, actuando, por tanto, Ncap como cierre al igual que lo hace el extremo C-terminal en cSrc [32]. La isoforma cAbl-1a se encuentra regulada aún careciendo de este grupo miristoílo, lo que indica que éste no es vital para la regulación, o que *in vivo* puede ser sustituido por la acción de fosfoinositoles, lípidos u otras interacciones [24].



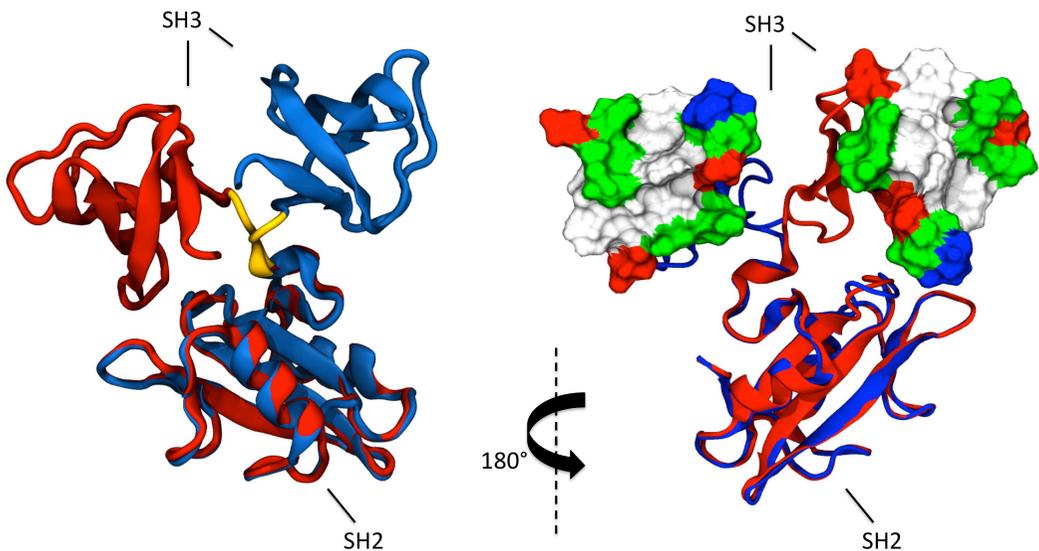
**Figura 4.5 Comparación del primer exón de las dos isoformas de cAbl.** En la parte superior se muestra el alineamiento del extremo N-terminal de las dos isoformas (a y b) de cAbl. En verde se resaltan las regiones necesarias para la regulación *in vivo* e *in vitro* de cAbl. En la parte inferior se muestra la organización de dominios para la quinasa de tirosina cAbl. Con un asterisco se resalta la posición donde se encuentra la SER69 fosforilada.

En el mismo estudio cristalográfico, fue posible determinar la estructura de la región Ncap adyacente al dominio SH3 (residuos 61 a 82), ya que ésta se encontraba interaccionando con el tándem SH32, principalmente con la secuencia conectora C3-2 y con el dominio SH2. Esta región de Ncap, al igual que la anterior, es necesaria para que cAbl se mantenga autoinhibida, con lo que mutaciones puntuales en las posiciones LYS70, GLU71, ASN72, LEU73 y ALA75, desregulan cAbl [140]. Dichas posiciones se encuentran

## CAPÍTULO 4

interaccionando con el dominio SH2, la región conectora C3-2 y el dominio SH3. Esta región de Ncap, es codificada en el segundo exón, de forma que se comparte por ambas isoformas, como se puede apreciar en detalle en la Figura 4.5. También en este estudio cristalográfico se observó que la SER69 situada en Ncap, se encontraba fosforilada, lo que sugiere una posible función como interruptor en el mecanismo de regulación. Esto ha motivado posteriores estudios en los que no queda clara la importancia de esta fosforilación en la regulación de la actividad de cAbl. [146]

A pesar de la abundante información que poseemos sobre cAbl, su regulación sigue siendo un complicado rompecabezas en el que aún existen interrogantes sin respuesta. Sabemos que cAbl requiere de un estímulo para llevar a cabo su actividad enzimática, que dicha activación conlleva un cambio conformacional y que existen, además de los dominios modulares, numerosos elementos implicados en la regulación de su actividad, lo que nos permite afirmar que se trata de una estructura muy cooperativa. Sin embargo, aún seguimos sin conocer muchos aspectos del mecanismo de activación e inhibición de su actividad. Incluso, desconocemos si la existencia de diferentes elementos reguladores responde a la necesidad de cAbl de responder de forma específica a diferentes estímulos y, de ser así, si existen diferentes estados activos, o si, por el contrario, sólo existe un estado activado independientemente del elemento regulador alterado. Dicho conocimiento permitiría el desarrollo de mejores y más efectivas terapias, al poder desarrollar fármacos específicos bloqueando la activación de elementos concretos de la quinasa o para cada uno de los estados de activación que pudieran existir.



**Figura 4.6** Superposición del tándem de los dominios SH3 y SH2 de la estructura cristalográfica con código PDB 2ABL (rojo) y de la estructura con código PDB 2F00 en azul. En ambos paneles el tándem ha sido superpuesto usando como referencia el dominio SH2. En el panel de la izquierda se resalta en amarillo la región del conector de ambos dominios que presenta un cambio conformacional. En el panel de la izquierda se ha representado la superficie de la interfaz de unión del dominio SH3 en ambas PDB. La superficie ha sido coloreada en concordancia con las propiedades de los residuos que la componen: verde para los residuos polares, azul para los básicos, blanco para los apolares y rojo para los ácidos.

En todo caso, la regulación de cAbl depende de un proceso reversible de ensamblaje y desensamblaje de los dominios modulares, cuya dinámica conformacional es modulada por interacciones intramoleculares, intermoleculares, fosfatasas y otras quinasas [21, 23]. En este sentido, llama la atención la conformación que adopta el tándem de dominios modulares de cAbl en la estructura cristalográfica con código PDB 2ABL. Esta construcción del tándem de los dominios modulares SH32 de cAbl aislados se encuentra en una conformación diferente a la de los tándems de la estructura cristalográfica con código PDB 2F00, resuelta 5 años después, donde el tándem SH32 está acompañado de otras regiones de cAbl, Ncap y el dominio catalítico (CD), todo ello en su conformación inactiva [32]. Al superponer los dominios modulares

## CAPÍTULO 4

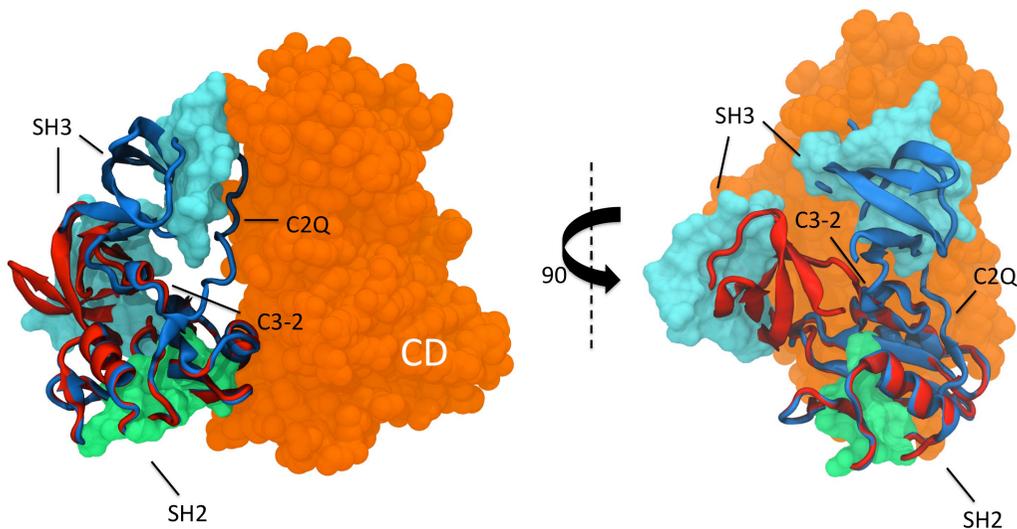
de ambas estructuras obtenemos un RMSD superior a 10 Å. Esta gran diferencia no corresponde a cambios conformacionales de los dominios modulares, sino, como se aprecia en la Figura 4.6, a un ligero cambio conformacional en la secuencia de conexión C3-2, que provoca un cambio en la orientación relativa de los dominios SH2 y SH3. *A priori*, una hipótesis que nos planteamos es que este cambio de conformación se podría deberse a la ausencia de las interacciones con elementos reguladores como el Ncap o la secuencia conectora C2Q, desplazando el equilibrio conformacional del tándem hacia una conformación que correspondería con la que adopta el tándem cuando cAbl está activada.

Con la intención de confirmar esta hipótesis, se inició un estudio de la dinámica del tándem SH32 de cAbl, para estudiar cómo ésta se veía afectada por los diferentes elementos de la quinasa, como son la región Ncap, el conector C2Q y la fosforilación en la SER69 del Ncap.

### 4.2 RESULTADOS Y DISCUSIÓN

Como se puede observar en la Figura 4.6, la conformación que adopta el tándem SH32 en la estructura cristalográfica con código PDB 2ABL deja los sitios de reconocimiento del dominio SH3 en una orientación relativa totalmente diferente a la observada en la estructura cristalográfica 2F00. Para que el tándem pueda adquirir esta conformación en el contexto de la quinasa completa, es necesario que el dominio SH3 se desplace, dejando de interactuar con la secuencia de conexión C2Q, como podemos ver en la Figura 4.7, donde hemos superpuesto ambas estructuras usando como referencia el dominio SH2. Esta conformación dejaría el sitio de reconocimiento de SH3 libre para reconocer ligandos externos, reforzando la

idea de que la conformación en la que se encuentra el tándem SH32 en la estructura cristalográfica 2ABL correspondería a la que éste adopta cuando la quinasa completa se encuentra activa o en una de las posibles conformaciones activas. Por eso, y para evitar confusiones, definiremos a esta conformación observada en 2ABL como abierta, y a la proveniente de cAbl inactiva de la estructura cristalográfica 2FO0, como cerrada.



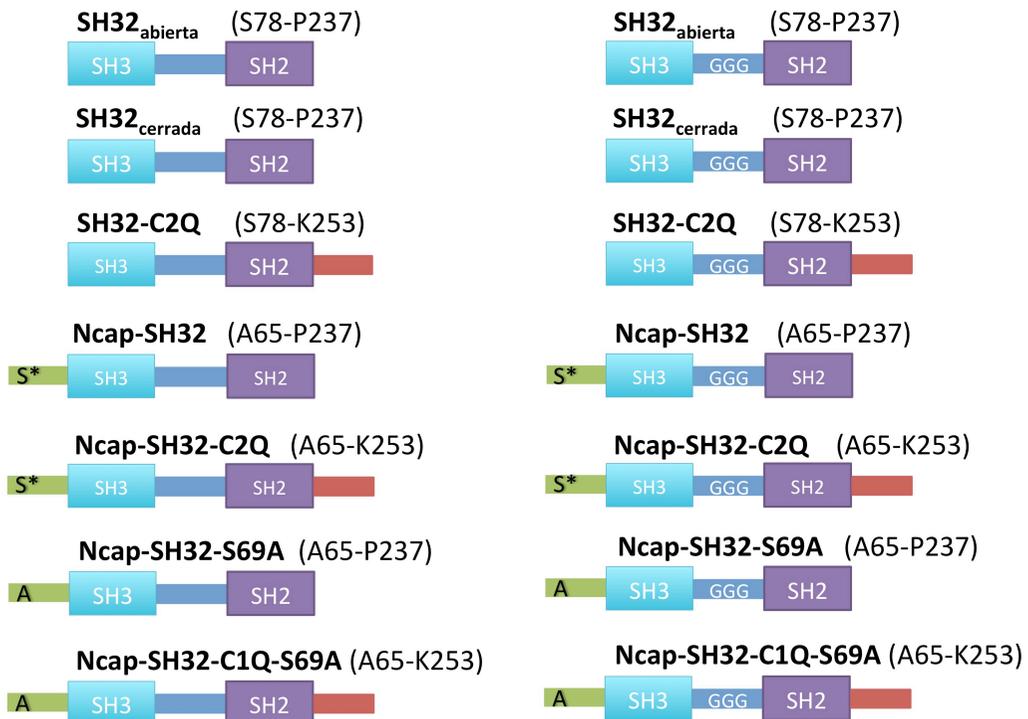
**Figura 4.7** Superposición de la estructura cristalográfica del tándem SH32 aislado (Rojo) con la estructura del tándem SH32 en el contexto de cAbl autoinhibida (Azul). Para la estructura cristalográfica del tándem aislado se utilizó el PDB con código 2ABL. Para la estructura autoinhibida de cAbl a partir del PDB con código 2FO0. En azul se representa la estructura del tándem de los dominios modulares (SH32). La superficie de las interfaces de unión de ambos dominios modulares ha sido resaltada en color cian y verde, para el dominio SH3 y para el SH2 respectivamente. El dominio catalítico ha sido representado en naranja.

#### 4.2.1 Dinámica conformacional del tándem SH32

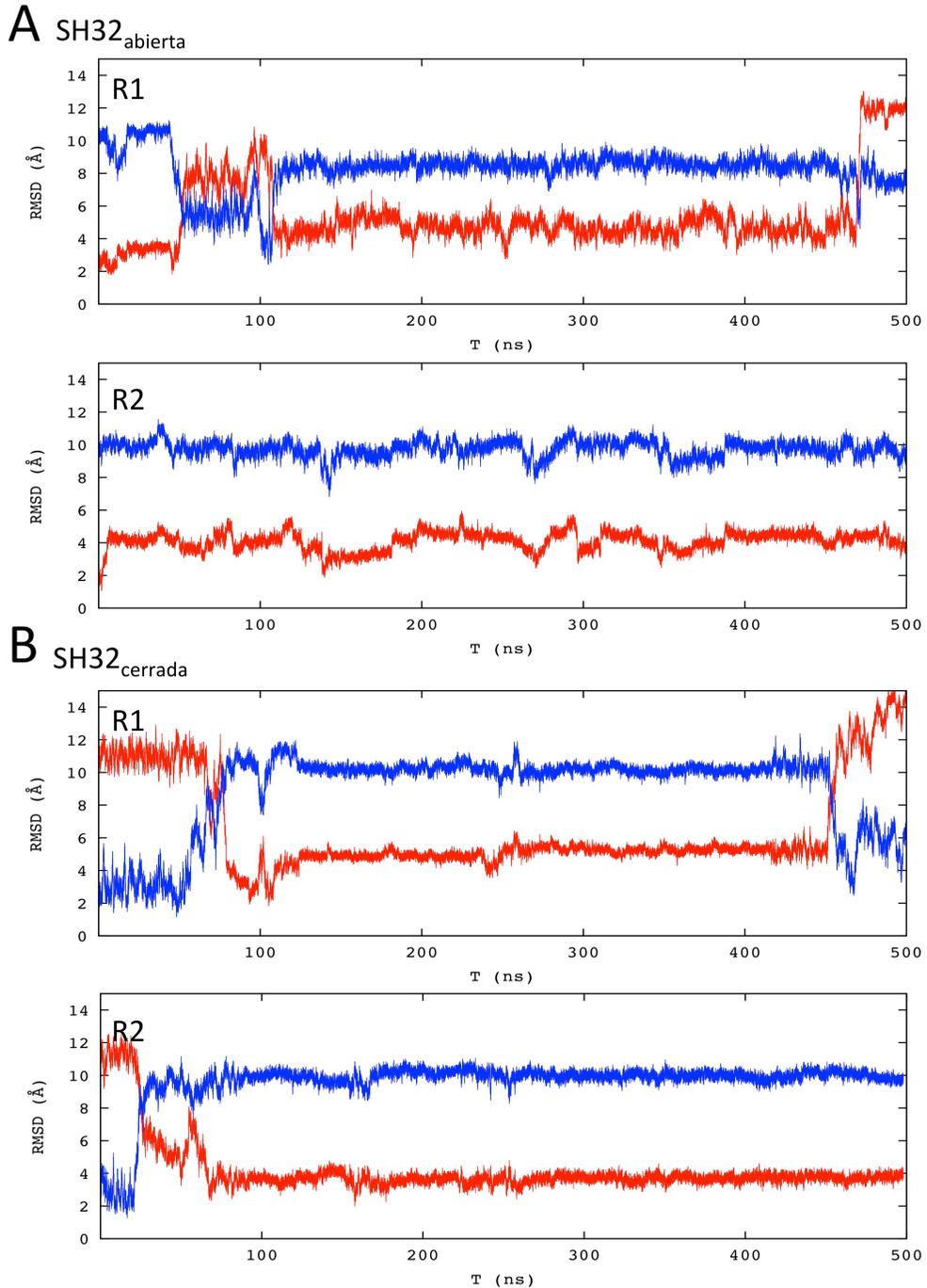
Con el objetivo de comparar el comportamiento conformacional de SH3 y SH2, se generaron cuatro trayectorias de dinámica molecular de 500 ns, siguiendo el protocolo descrito en la Sección 5.3 del capítulo de Materiales y Métodos. En dos de ellas usamos como conformación inicial la abierta y en las

## CAPÍTULO 4

otras dos usamos la conformación cerrada a la que previamente se le habían eliminado las regiones Ncap, C2Q y el dominio catalítico, quedando por tanto los residuos comprendidos entre SER78 y PRO237. En la Figura 4.8 quedan especificados en detalle los residuos que comprenden cada construcción utilizada para este estudio.



**Figura 4.8 Esquema de las diferentes construcciones utilizadas en los experimentos de dinámica molecular del tándem SH32.** Cada construcción utilizada en el estudio es acompañada de un esquema de la organización de los dominios y se marca también el primer y último residuo de la construcción, tomando como referencia la numeración de cAbl-1b. La sección verde acompañada de "S\*" corresponde al Ncap silvestre, con la SER69 fosforilada. Por otro lado, la sección verde con una A, corresponde al Ncap con la SER69 mutada por una alanina. La sección azul marino sin ningún indicativo corresponde a la secuencia de conexión C2-3 en su forma silvestre, mientras que la que está acompañada de "GGG" corresponde a la forma con las posiciones 139, 140 y 141 mutadas a glicina. Finalmente, la sección roja representa la secuencia de conexión del dominio SH2 con el dominio catalítico (C2Q).



**Figura 4.9** Representación de los valores de RMSD a lo largo de cada trayectoria del tándem SH32. La línea roja corresponde al RMSD de la trayectoria usando como referencia la estructura cristalográfica 2ABL o abierta, mientras que la azul es el RMSD usando como referencia la estructura 2FO0 o cerrada. En el panel A; se representan los valores de RMSD obtenidos para la construcción SH32<sub>abierta</sub> y su replica. En el panel B; se representación de los valores de RMSD para la construcción SH32<sub>cerrada</sub> y su replica.

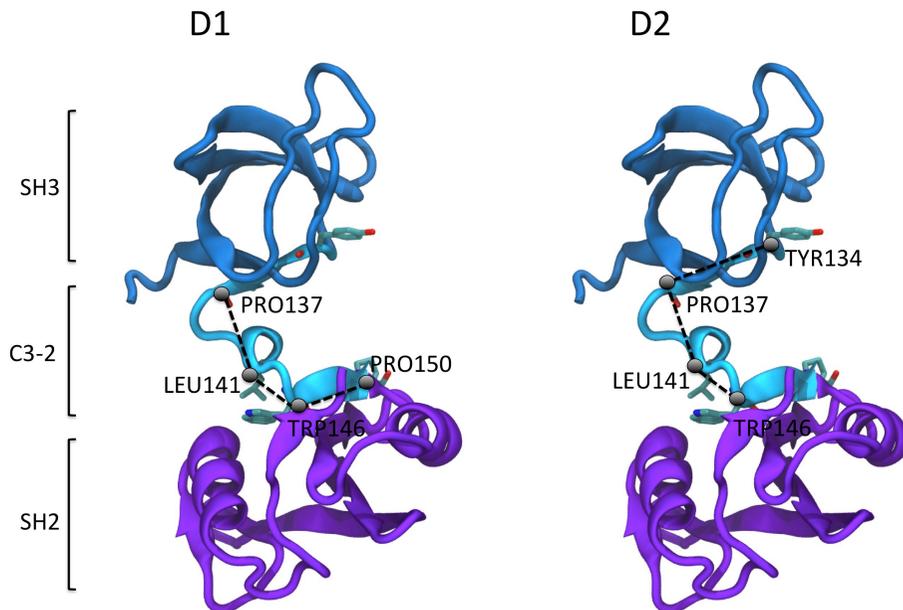
## CAPÍTULO 4

En la Figura 4.9 se representan los valores de RMSD a lo largo de cada trayectoria, donde la línea roja corresponde al RMSD de la trayectoria usando como referencia la estructura cristalográfica 2ABL o abierta, mientras que la azul es el RMSD usando como referencia la estructura 2F00 o cerrada. Como podemos observar, en todas las trayectorias en las que partimos de la configuración cerrada ( $SH32_{\text{cerrada}}$ ), el tándem acaba explorado una conformación muy cercana en RMSD a la conformación abierta, mientras que las trayectorias en las que se usó como estructura de partida la conformación abierta ( $SH32_{\text{abierta}}$ ) permanecen la mayor parte del tiempo de simulación alrededor de la estructura de partida, salvo un breve instante (50 ns) en los que la réplica  $SH32_{\text{abierta}} r1$ , adoptó la conformación cerrada.

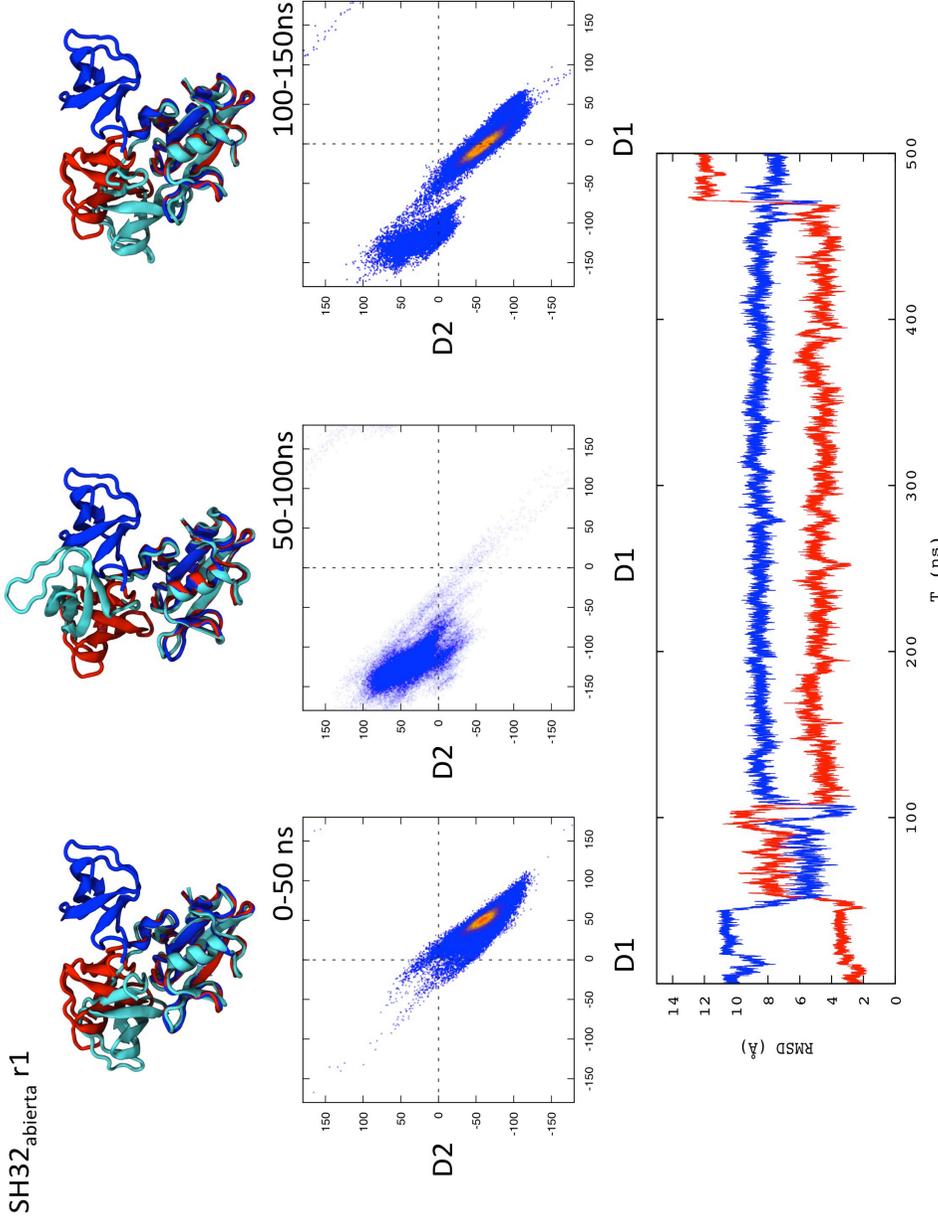
Por consiguiente, en la escala de tiempo de nuestras simulaciones de dinámica molecular realizadas (500 ns), hemos observado la capacidad del tándem de adoptar tanto la conformación abierta (2ABL) como la cerrada (2F00). Aunque no podemos estar completamente seguros solamente con estos datos, el comportamiento observado del tándem en ausencia de Ncap y C2Q, deja evidencias de un posible equilibrio conformacional entre ambas conformaciones cerrada y abierta, lo que podría apoyar la hipótesis de que la conformación abierta coincide con la conformación del tándem cuando cAbl se encuentra activa o que, al menos, éstas serían las conformaciones más probables dentro del equilibrio conformacional del tándem en ausencia de las interacciones intramoleculares de otras regiones, como Ncap o C2Q, que podrían modular este equilibrio conformacional. Curiosamente y como veremos con mas detalle en la Sección 4.2.3 de este Capítulo, este equilibrio se encuentra dirigido principalmente por la pequeña región que conecta los dominios SH3 y SH2.

#### 4.2.1.1 Descriptor del cambio conformacional del tándem SH32

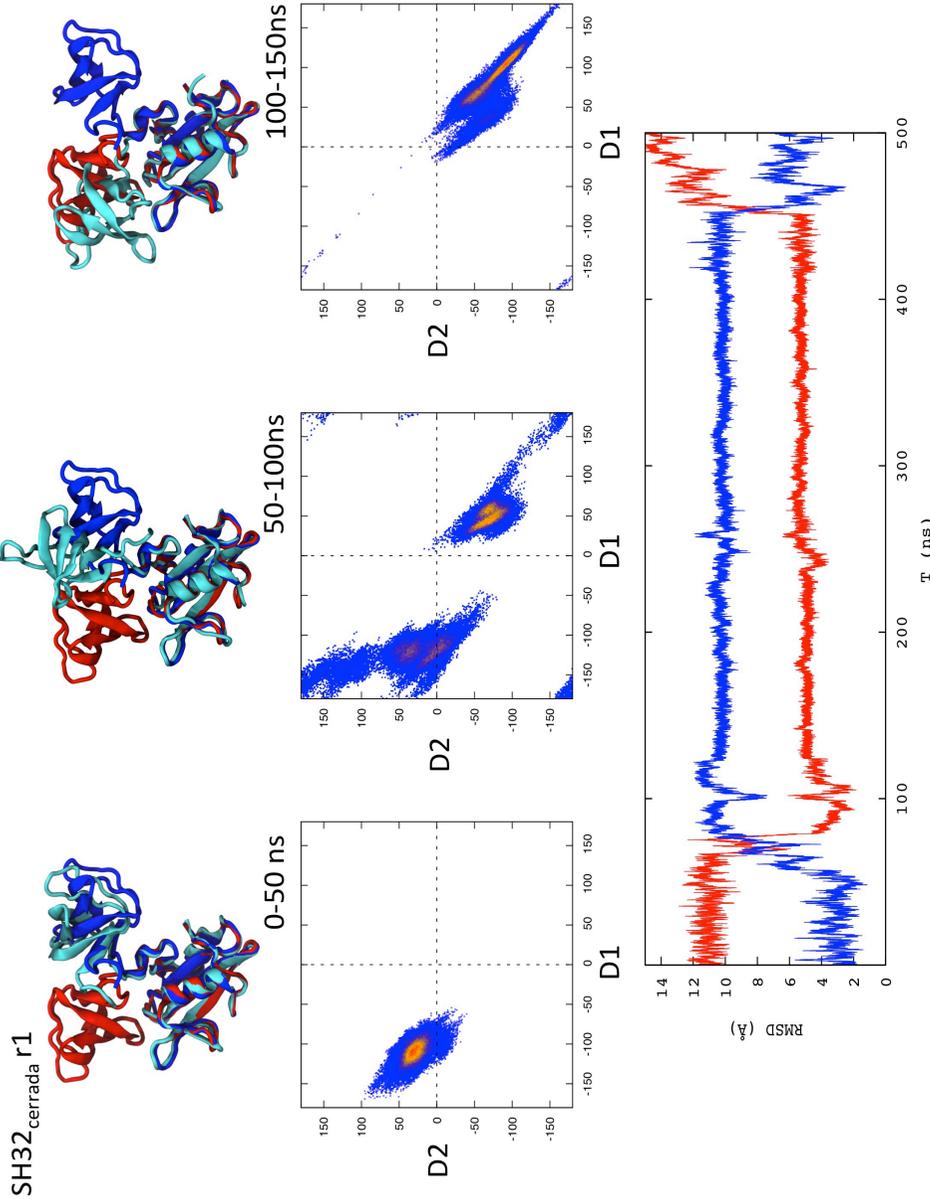
Aunque el RMSD es una gran herramienta para analizar las trayectorias de dinámica molecular, para el estudio del equilibrio conformacional del tándem es un parámetro con demasiada incertidumbre a la hora de discriminar entre las distintas conformaciones. Esto se debe a que dos o más conformaciones muy diferentes entre sí pueden compartir el mismo valor de RMSD. Por lo tanto, para el análisis exhaustivo de las trayectorias de dinámica molecular es necesario definir un mejor descriptor del cambio de conformacional.



**Figura 4.10** Detalle de los descriptores para el cambio conformacional del tándem SH32. Se representa el tándem SH32 a partir de la estructura cristalografica con código PDB; 2FO0. En ambos paneles se resalta el dominio SH2 en purpura, en azul el dominio SH3 y en cian C2Q, la secuencia conectora de ambos dominios. Se resaltan en varillas los residuos tomados para formar los ángulos diedros que actuarán como descriptores. En el panel de la izquierda el ángulo diedro denominado D1 y formados por los centro de masas de los residuos PRO137, LEU141, TRP146, PRO150. En el panel de la derecha el segundo ángulo diedro, denominado D2, y formado por Denominado por por los centros de masas de los residuos TYR134, PRO137, LEU141, TRP146.

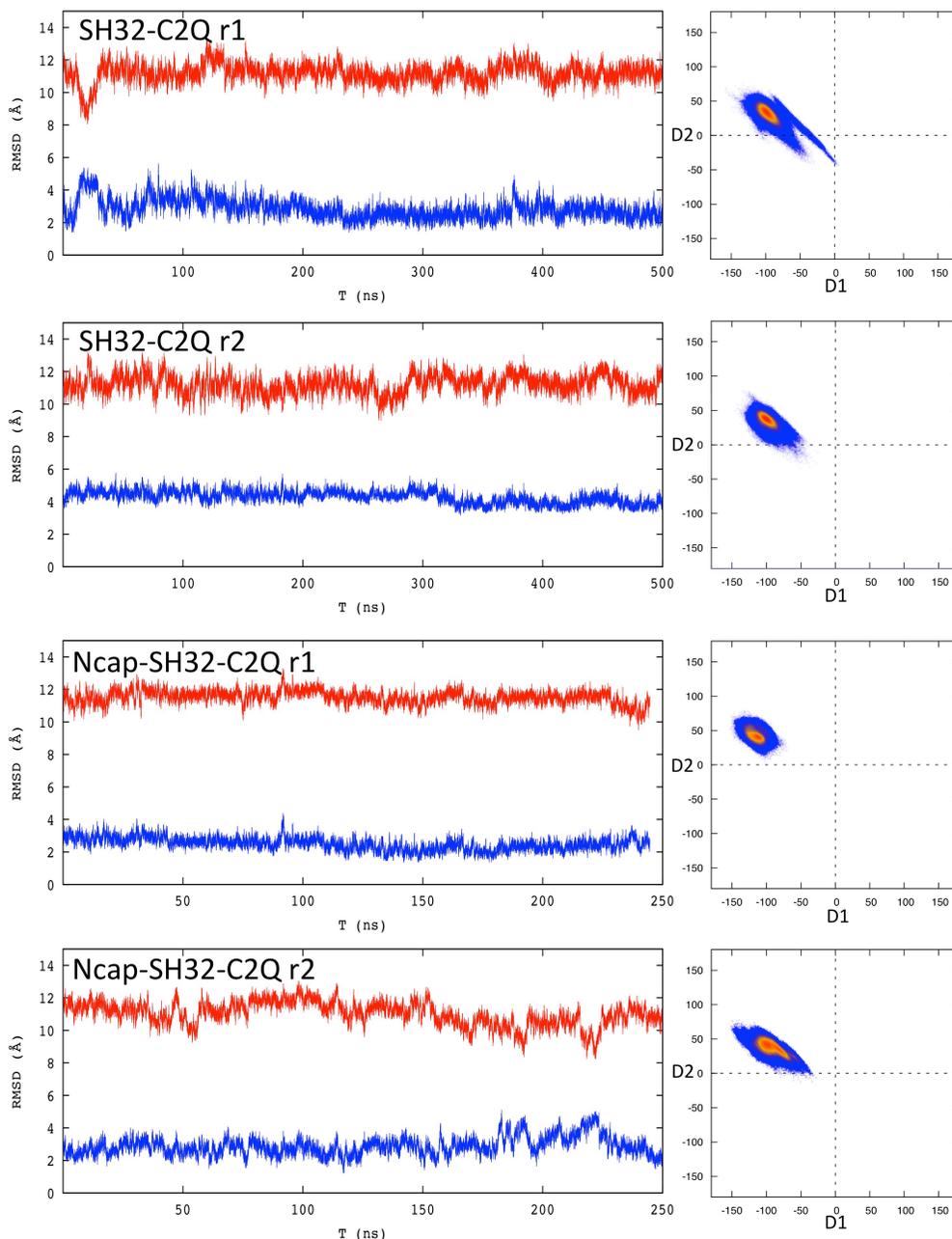


**Figura 4.11 Comportamiento de los descriptores D1 y D2 a lo largo de la trayectoria SH32<sub>abierto</sub> r1.** En el panel superior se muestra la superposición del tandem SH32 de la estructura cristalográfica con código PDB 2ABL en rojo con el de la estructura con código PDB 2FO0 en azul. En color cian se resalta una estructura representativa de la trayectoria en el intervalo de tiempo marcado. En el panel central se representa la frecuencia de valores que adoptan los descriptores D1 y D2 en el mismo intervalo de tiempo (50ns). Las coordenadas de los descriptores están coloradas según la frecuencia en las que son visitadas durante la trayectoria; En azul se resaltan las coordenadas visitadas de 1 a 10 veces, en naranja de 10 a 100 y en rojo más de 100. En el panel inferior se representan los valores de RMSD a lo largo de la trayectoria del tandem SH32, donde la línea roja corresponde al RMSD de la trayectoria usando como referencia la estructura cristalográfica 2ABL o abierta, mientras que la azul es el RMSD usando como referencia la estructura 2FO0 o cerrada



**Figura 4.12 Co** superposición del tándem SH32 de la estructura cristalográfica con código PDB 2ABL en rojo con el de la estructura con código PDB 2FO0 en azul. En color cian se resalta una estructura representativa de la trayectoria en el intervalo de tiempo marcado. En el panel central se representa la frecuencia de valores que adoptan los descriptores D1 y D2 en el mismo intervalo de tiempo (50ns). Las coordenadas de los descriptores están coloradas según la frecuencia en las que son visitadas durante la trayectoria; En azul se resaltan las coordenadas visitadas de 1 a 10 veces, en naranja de 10 a 100 y en rojo más de 100. En el panel inferior se representan los valores de RMSD a lo largo de la trayectoria del tándem SH32, donde la línea roja corresponde al RMSD de la trayectoria usando como referencia la estructura cristalográfica 2ABL o abierta, mientras que la azul es el RMSD usando como referencia la estructura 2FO0 o cerrada

## CAPÍTULO 4



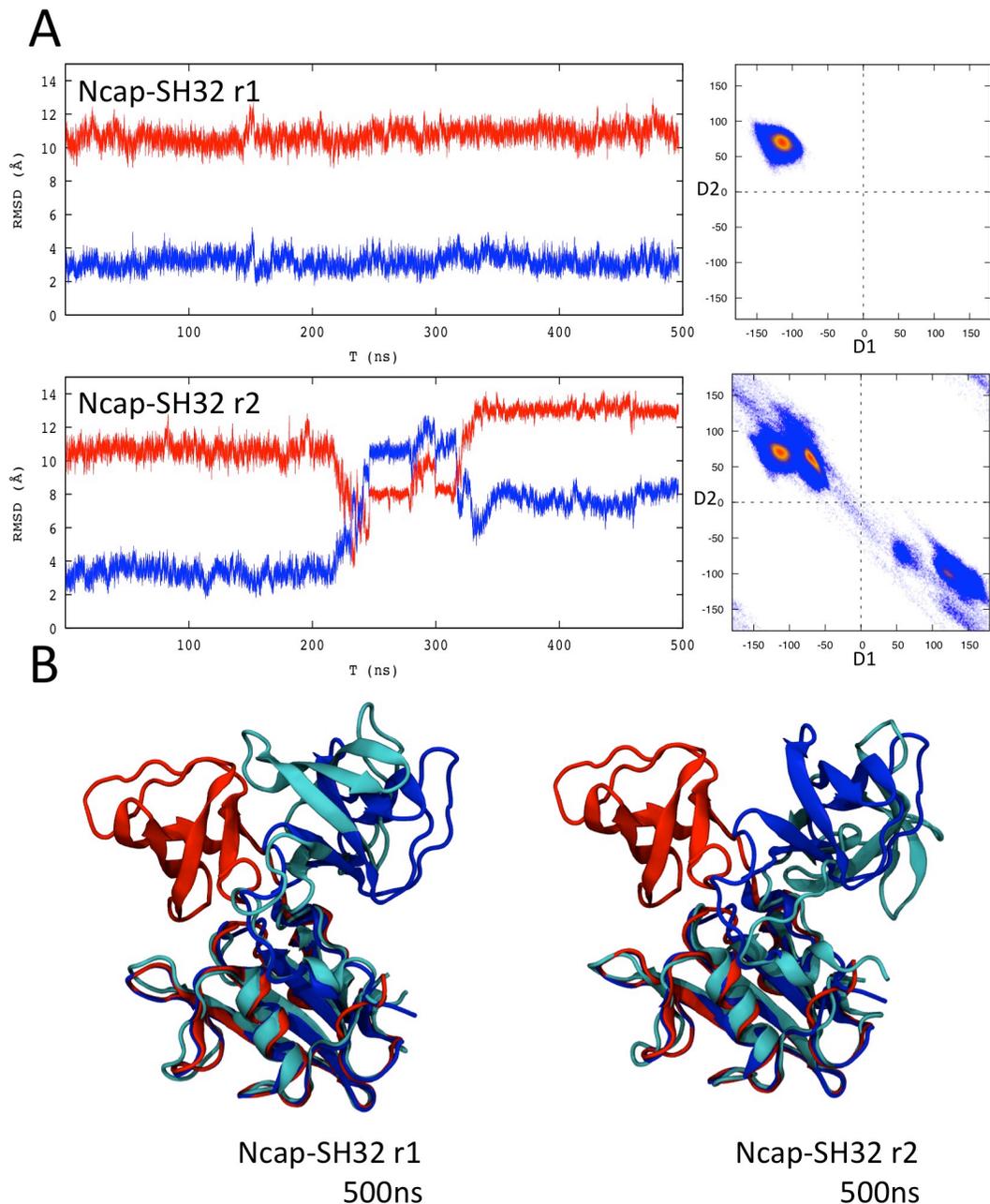
**Figura 4.13** El RMSD (paneles de la izquierda) acompañado de la frecuencia (paneles de la derecha) de los descriptores D1 y D2 a lo largo de las trayectorias SH32-C2Q y Ncap-SH32-C2Q y sus replicas. El color rojo corresponde a las trayectorias usando como referencia la estructura cristalográfica 2ABL o abierta, mientras que el azul corresponde a la estructura 2FO0 o cerrada. Las coordenadas de los descriptores están coloradas según la frecuencia en las que son visitadas durante la trayectoria; En azul se resaltan las coordenadas visitadas de 1 a 10 veces, en naranja de 10 a 100 y en rojo más de 100.

#### 4.2.2 Modificación de la dinámica conformacional del tándem SH32 en presencia de Ncap y C2Q

Siguiendo el mismo protocolo que en las anteriores simulaciones de dinámica molecular, se calcularon nuevas trayectorias de 500 ns. A partir de la estructura cristalográfica 2FO0 (cerrada) se generaron dos trayectorias incluyendo la región Ncap comprendida entre los residuos 61 al 77 (Ncap-SH32), dos con la región conectora del dominio SH2 y el catalítico comprendida entre los residuos 238 al 253 (SH32-C2Q) y finalmente, dos más con ambas regiones (Ncap-SH32-C2Q). El objetivo de estas nuevas simulaciones es, por tanto, determinar si estas regiones adicionales modifican el comportamiento conformacional del tándem SH32 aislado.

En primer lugar en la Figura 4.13 podemos observar la evolución del RMSD y de los descriptores D1 y D2 en las trayectorias en las que partimos del tándem acompañado de la secuencia conectora C2Q, SH32-C2Q. De nuevo, la línea roja representa el RMSD usando como referencia la estructura abierta, y la azul el RMSD usando la estructura cerrada. En todas las trayectorias en las que se partió de una estructura con la secuencia conectora C2Q el tándem se mantiene en una conformación cercana a la de partida, es decir, la cerrada, esté acompañado o no de la región Ncap.

En cambio, al observar el RMSD y los descriptores obtenidos de las trayectorias con el tándem que incluye únicamente la región Ncap, Figura 4.14, observamos que, mientras en la réplica r1 el tándem se mantiene en la conformación inicial (la cerrada) a lo largo de toda la trayectoria, en la segunda réplica la estructura sufre un cambio de conformación alrededor de los 200 ns. En este caso, adoptando una estructura alejada conformacionalmente tanto de la abierta como de la cerrada en RMSD.



**Figura 4.14** Comportamiento del tándem SH32 en las trayectorias para el constructo Ncap-SH32. En el panel A se representa el RMSD (paneles de la izquierda) acompañado de la frecuencia (paneles de la derecha) de los descriptores D1 y D2 a lo largo de las trayectorias Ncap-SH32. El color rojo corresponde a la trayectoria usando como referencia la estructura cristalográfica 2ABL o abierta, mientras que el color azul corresponde a la estructura 2FO0 o cerrada. En el panel B se muestra la superposición de tres estructuras, la estructura cristalográfica con código PDB 2ABL (rojo), la estructura con código PDB 2FO0 en azul y las coordenadas finales de la trayectoria . Ncap-SH32 y su replica (cian).

No obstante, si comprobamos la estructura final de la trayectoria, observamos cómo el tándem se encuentra en una conformación que aunque no es exactamente igual a la cerrada, la orientación relativa de los dominios es similar. Durante ambas trayectorias, incluso en la segunda réplica donde se produce el cambio conformacional, la región Ncap no pierde en ningún momento las interacciones que establece con el dominio SH2 y con la región C3-2, de forma que podemos intuir que la presencia de estas interacciones no limita el espacio conformacional del tándem de la misma forma que lo hace C2Q.

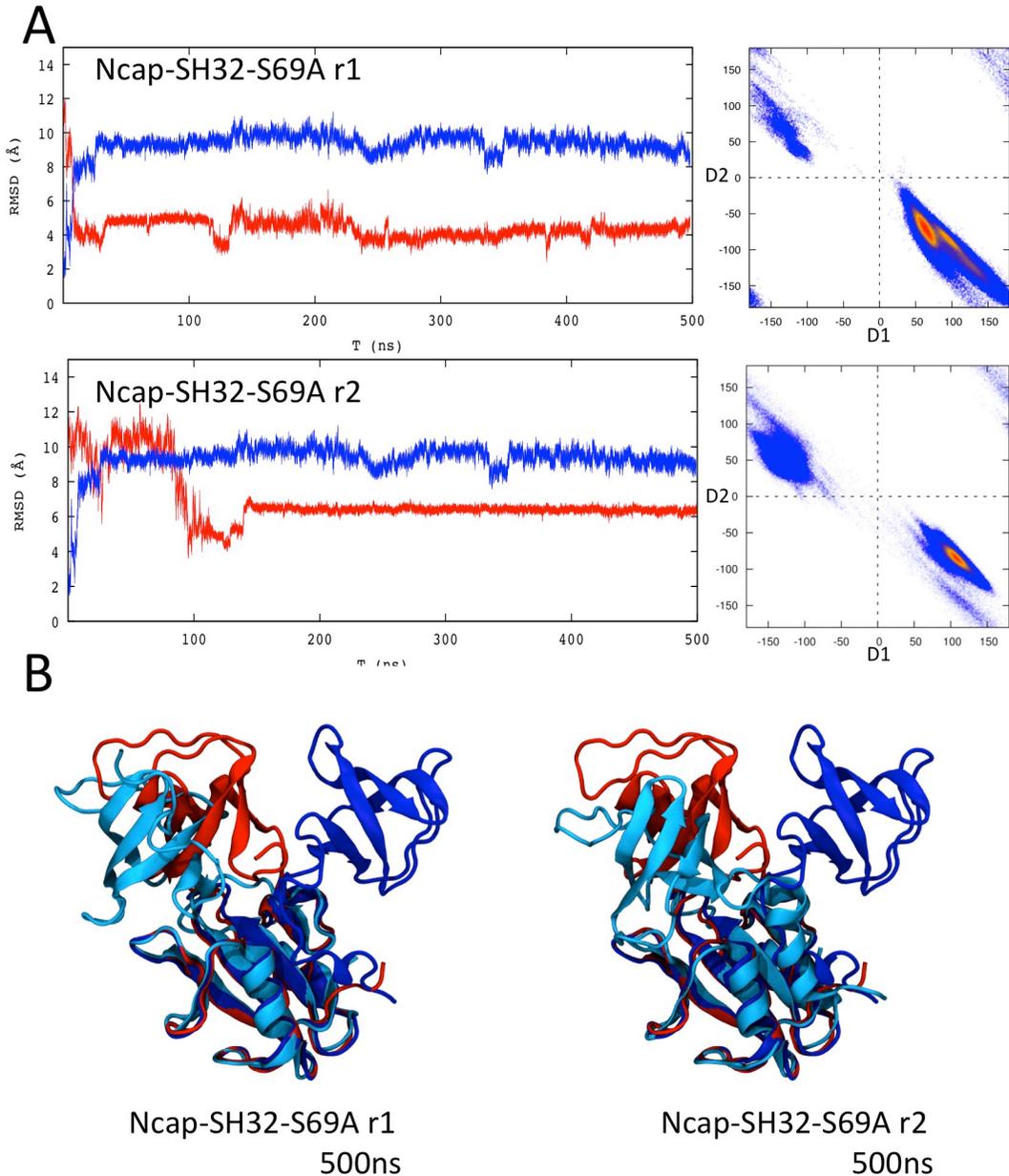
En experimentos previos con quinasas de tirosina de la subfamilia Src, se determinó que la afinidad del dominio SH3 por la secuencia de conexión C2Q era muy baja, estando favorecida principalmente por el empaquetamiento de esta secuencia entre los dominios quinasa y SH3 [144, 145]. Por otro lado, sabemos que alteraciones del segmento Ncap producen la activación constitutiva de cAbl [140, 143]. Es decir, la modificación o eliminación de Ncap o C2Q produce la activación constitutiva de cAbl de forma que, *a priori*, si el cambio conformacional observado en las trayectorias de SH32 corresponde a la conformación activa o a uno de los intermedios en el mecanismo de activación de cAbl, deberíamos observar un cambio conformacional del tándem en todas las trayectorias. Entonces, ¿por qué solamente observamos un cambio conformacional en una de las trayectorias de la construcción Ncap-SH32 (la r2) y no en el resto de trayectorias? ¿No carecen todos los constructos de elementos reguladores?

Una posible explicación, sería que cada elemento de la quinasa participa de una forma totalmente diferente en el mecanismo de activación y de regulación. Como hemos mencionado anteriormente, la secuencia de C2Q de cAbl podría en efecto tener una afinidad muy superior por el dominio SH3

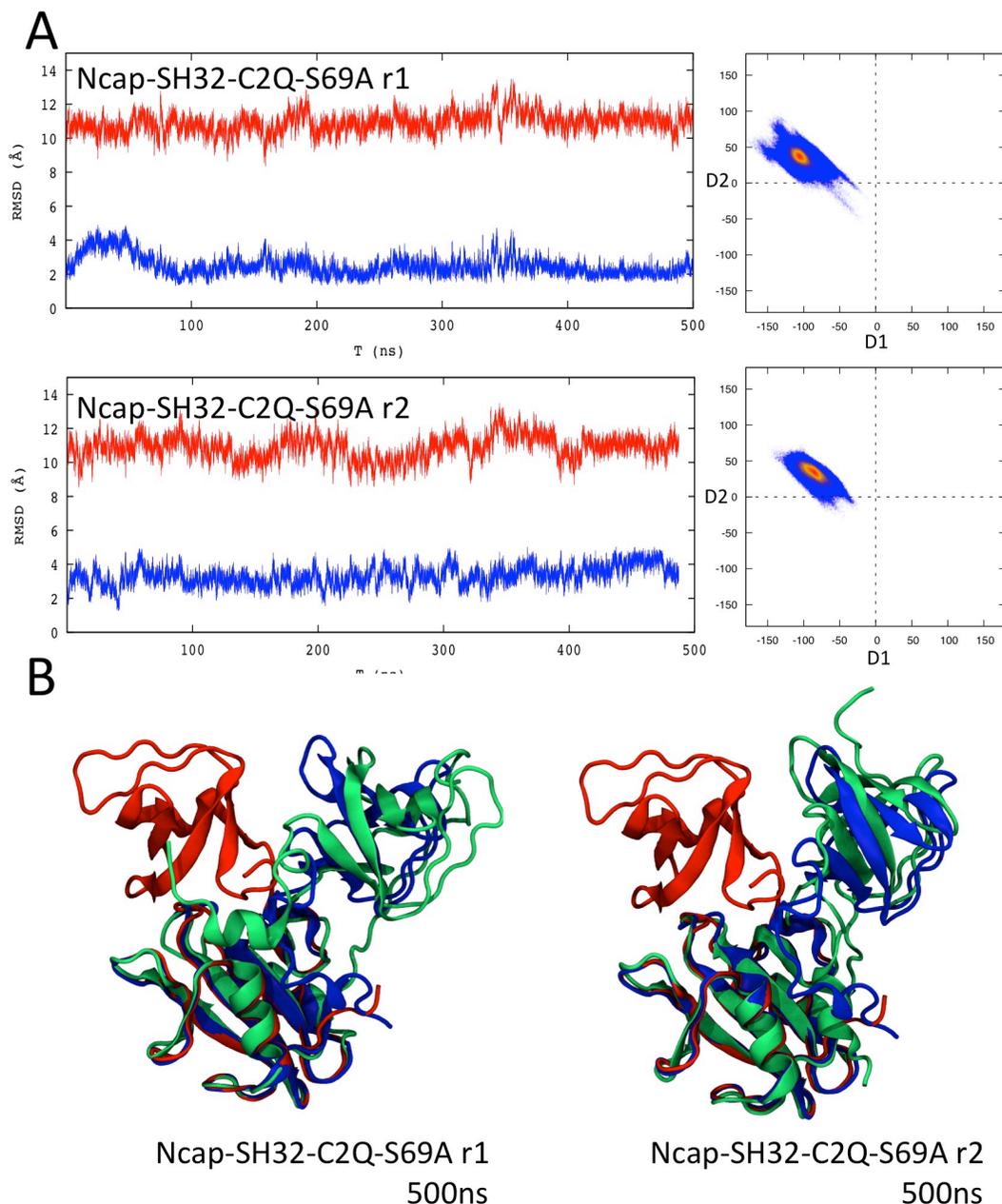
## CAPÍTULO 4

que en la subfamilia de Src, como proponen Engen y colaboradores [26]. Así, la unión de C2Q a SH3 en la conformación inicial de la trayectoria es suficiente para que el tándem permanezca durante toda la simulación en la conformación cerrada. Hay que tener siempre presente que el tándem SH32 no puede adoptar la conformación abierta sin dejar de interactuar con la región C2Q, de forma que esta secuencia podría constituir un elemento clave para mantener cAbl en la conformación autoinhibida. En cambio, con el mismo tiempo de simulación, las interacciones de Ncap con el tándem no impiden un cambio conformacional en el tándem. Esto parece sugerir que la región Ncap podría actuar estabilizando la conformación inhibida, ya que ciertamente el comportamiento del tándem con esta región no es equivalente al del tándem aislado, aunque también podría tener un papel como activador de cAbl. Sabemos que la simple mutación de las posiciones LYS70, GLU71 o ASN72 en cAbl, son suficientes para desregular su actividad, de forma que esta región de Ncap (residuos 61-82) debe, por sí sola, jugar un papel en el mecanismo de regulación. Además, sabemos por experimentos de intercambio hidrógeno/deuterio, que la presencia de Ncap estabiliza el tándem de los dominios modulares SH32 [146].

De este modo, la pérdida de interacciones de Ncap en la región adyacente al tándem SH32 podría alterar el comportamiento del tándem. En esta línea, llama la atención la fosforilación de la SER69 en la región Ncap. Los residuos fosforilados son considerados comúnmente como interruptores moleculares mediados por la acción de quinasas y fosfatasas. Esto parece indicar que la desfosforilación de la serina podría modular el comportamiento de esta parte de Ncap.



**Figura 4.15 Comportamiento del tándem SH32 en las trayectorias para el constructo Ncap-SH32-S69A.** En el panel A se muestra el RMSD (paneles de la izquierda) acompañado de la frecuencia (paneles de la derecha) de los descriptores D1 y D2 a lo largo de las trayectorias Ncap-SH32-S69A. El color rojo corresponde a la trayectoria usando como referencia la estructura cristalográfica 2ABL o abierta, mientras que el color azul corresponde a la estructura 2FO0 o cerrada. En el panel B se muestra la superposición de tres estructuras, la estructura cristalográfica con código PDB 2ABL (rojo), la estructura con código PDB 2FO0 en azul y las coordenadas finales de la trayectoria . Ncap-SH32-S69A y su replica (azul).



**Figura 4.16** Comportamiento del tándem SH32 en las trayectorias para el constructo Ncap-SH32-C2Q-S69A. En el panel A se muestra el RMSD (paneles de la izquierda) acompañado de la frecuencia (paneles de la derecha) de los descriptores D1 y D2 a lo largo de las trayectorias. El color rojo corresponde a la trayectoria usando como referencia la estructura cristalográfica 2ABL o abierta, mientras que el color azul corresponde a la estructura 2FO0 o cerrada. En el panel B se muestra la superposición de tres estructuras, la estructura cristalográfica con código PDB 2ABL (rojo), la estructura con código PDB 2FO0 en azul y las coordenadas finales de la trayectoria . Ncap-SH32-C2Q-S69A y su replica (azul).

Así, con la intención de observar cómo afectaba esta fosforilación al segmento Ncap y al tándem se crearon dos nuevas estructuras para iniciar trayectorias, siguiendo el mismo protocolo utilizado para las simulaciones anteriores, y así analizar los efectos de estos elementos. Utilizamos, para ello el tándem acompañado del Ncap, así como el tándem acompañado del Ncap y del segmento C2Q, pero en ambos casos la fosfoserina en posición 69 fue mutada a alanina.

En la Figura 4.15 mostramos los valores de RMSD y descriptores D1 y D2 obtenidos para las trayectorias del mutante S69A sin la presencia de la región C2Q (Ncap-SH32-S69A). En ambas replicas el tándem SH32 cambia rápidamente hacia la conformación abierta. Sin embargo, como podemos observar en la Figura 4.16 por los valores de RMSD y de los descriptores D1 y D2, la construcción con la fosfoserina mutada y con la región C2Q (Ncap-SH32-C2Q-S69A), el tándem SH32 se mantiene durante el tiempo de simulación en la conformación cerrada. No obstante, es importante señalar que en todas las simulaciones con la fosfoserina sustituida por alanina, independientemente de la presencia del segmento C2Q, la región Ncap perdió sus interacciones con el tándem o cambio de conformación, algo que no sucedió en ninguna de las trayectorias donde la región Ncap no fue alterada.

Esto indica que la fosfoserina 69 es un elemento clave para la interacción de la secuencia Ncap con el tándem SH32. Así mismo, la pérdida de estas interacciones facilitaría al tándem adquirir la conformación abierta en ausencia de otros elementos, como por ejemplo C2Q. Probablemente, una vez Ncap pierde las interacciones con el tándem SH32, el aumento de la entropía conformacional (recordemos que Ncap no tiene estructura) podría desestabilizar el tándem, en especial el dominio SH3, favoreciendo que éste adopte la conformación abierta. Este comportamiento encajaría con la

## CAPÍTULO 4

hipótesis de que Ncap actúa como un interruptor de cAbl. Además, es muy verosímil que la función de esta región de Ncap (residuos 61-82) esté íntimamente relacionada con la región Ncap que no ha sido resuelta en la estructura cristalográfica, (residuos 1-15). Como sabemos, la presencia de ambas regiones, es necesaria para que cAbl permanezca constitutivamente inactiva. De hecho, la forma oncogénica BCR-ABL, que se encuentra desregulada, carece del extremo Ncap miristoilado (residuos 1-15), pero no de la región de Ncap (residuos (61-82).

Las diferentes trayectorias de dinámica molecular no permiten asegurar que las interacciones que establece Ncap con el tándem sean responsables de estabilizar el tándem en la conformación cerrada, o al menos no con la misma intensidad que la región C2Q. De esta forma, que todo sugiere a que la región Ncap se trataría de un interruptor de la activación más que de un elemento que participe en la autoinhibición, mediado muy probablemente por una fosfatasa. Por otro lado la capacidad del tándem para adoptar la conformación abierta sin perder las interacciones con Ncap, podría ser un indicio de la capacidad de esta región de modificar el comportamiento del equilibrio conformacional del tándem hacia otras conformaciones no observadas.

### **4.2.3 Papel del conector C3-2 en la dinámica conformacional del tándem SH32**

Al superponer las estructuras cristalográficas 2ABL y 2F00 (usando como referencia solamente el dominio SH2) observamos un cambio en la orientación relativa de los dominios modulares, que implica el desplazamiento y rotación del dominio SH3 respecto el dominio SH2. Como se puede apreciar en la Figura 4.7, en la zona de la secuencia de conexión entre dominios, C3-2,

sólo podemos percibir un ligero cambio en algunos de los residuos más cercanos al dominio SH3. Además, los cambios conformacionales observados en las trayectorias anteriores, no implican cambios estructurales ni en el dominio SH3, ni en el dominio SH2, como se puede observar en el valor de RMSD medio obtenido para cada dominio Tabla 4.1. El RMSD medio de los dominios se encuentra muy por debajo de 1 Å en todas las trayectorias, incluso en las que ocurrió el cambio conformacional. En cambio, el valor de RMSD de la secuencia de conexión de los dominios modulares, C3-2, pasa de adquirir valores alrededor de 0,5 Å para aquellas trayectorias en las que no hay cambio conformacional, a llegar a 1,5 Å de media en las trayectorias en las que el tándem cambia de orientación relativa.

**Tabla X.1** RMSD medio de las trayectorias, considerando solamente los átomos del esqueleto peptídico de las diferentes regiones. Las réplicas en las que a lo largo de la trayectoria los dominios han cambiado de orientación se resaltan con un asterisco.

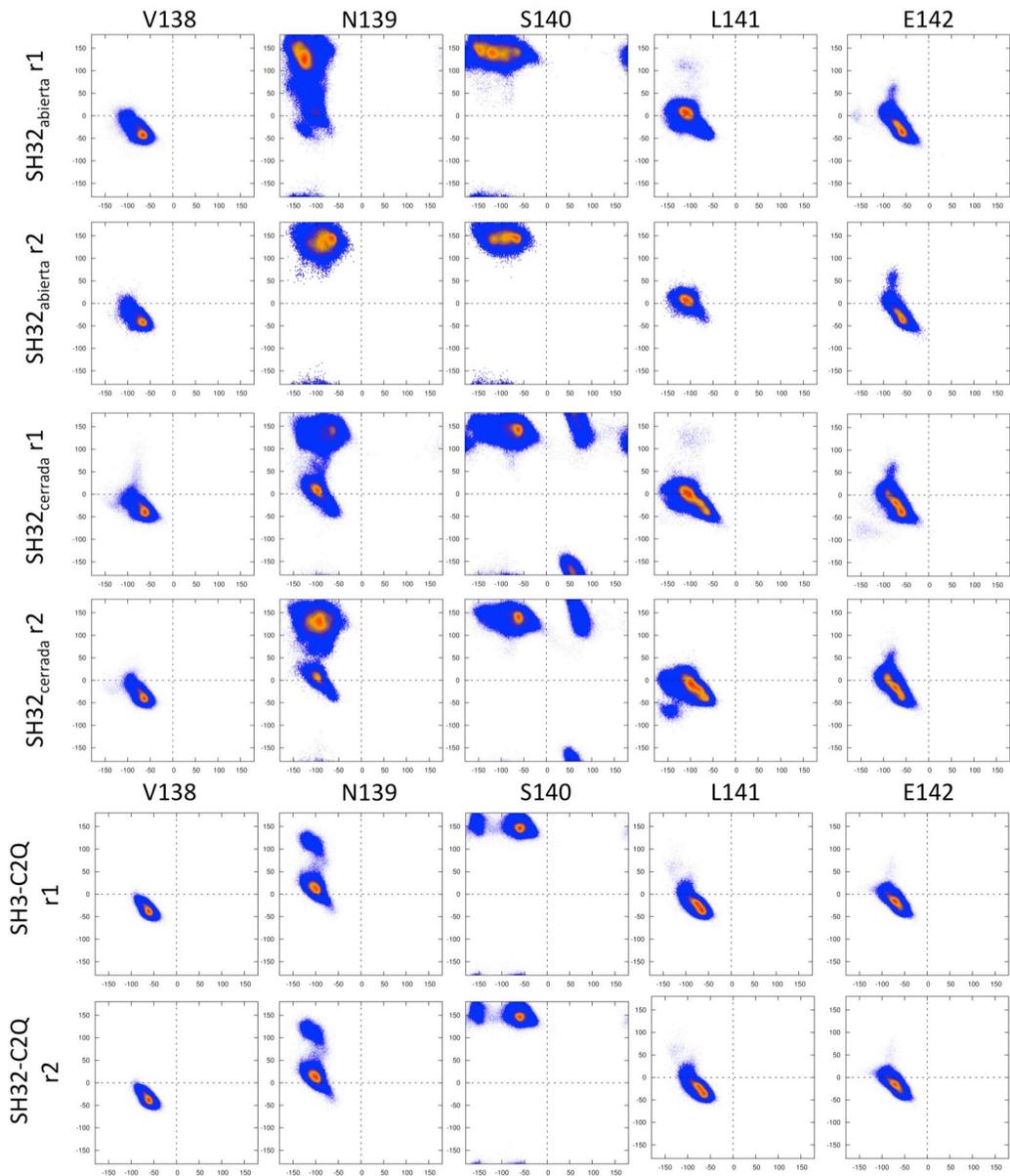
	SH32-Abierta		SH32-Cerrada		SH32-C2Q	
	R1*	R2	R1*	R2*	R1	R2
SH3	0,55	0,58	0,53	0,52	0,49	0,49
SH3-2	1,21	1,52	0,99	0,94	0,48	0,47
SH2	0,57	0,70	0,59	0,59	0,55	0,55

Con el objetivo de localizar los restos más importantes en el cambio conformacional de la secuencia C3-2, se recogieron los valores de los ángulos diedros del esqueleto peptídico de los restos que componen la secuencia C3-2 a lo largo de las diferentes trayectorias. Los valores de los ángulos diedros phi ( $\phi$ ) y psi ( $\psi$ ) otorgan una valiosa información sobre la conformación primaria de la cadena peptídica de una proteína. Mediante la comparación de los valores que adoptan a lo largo de las diferentes trayectorias podemos detectar aquellos residuos que han cambiado de conformación y, por tanto, los restos que gobiernan el cambio de orientación relativa de los dominios SH3 y SH2. En

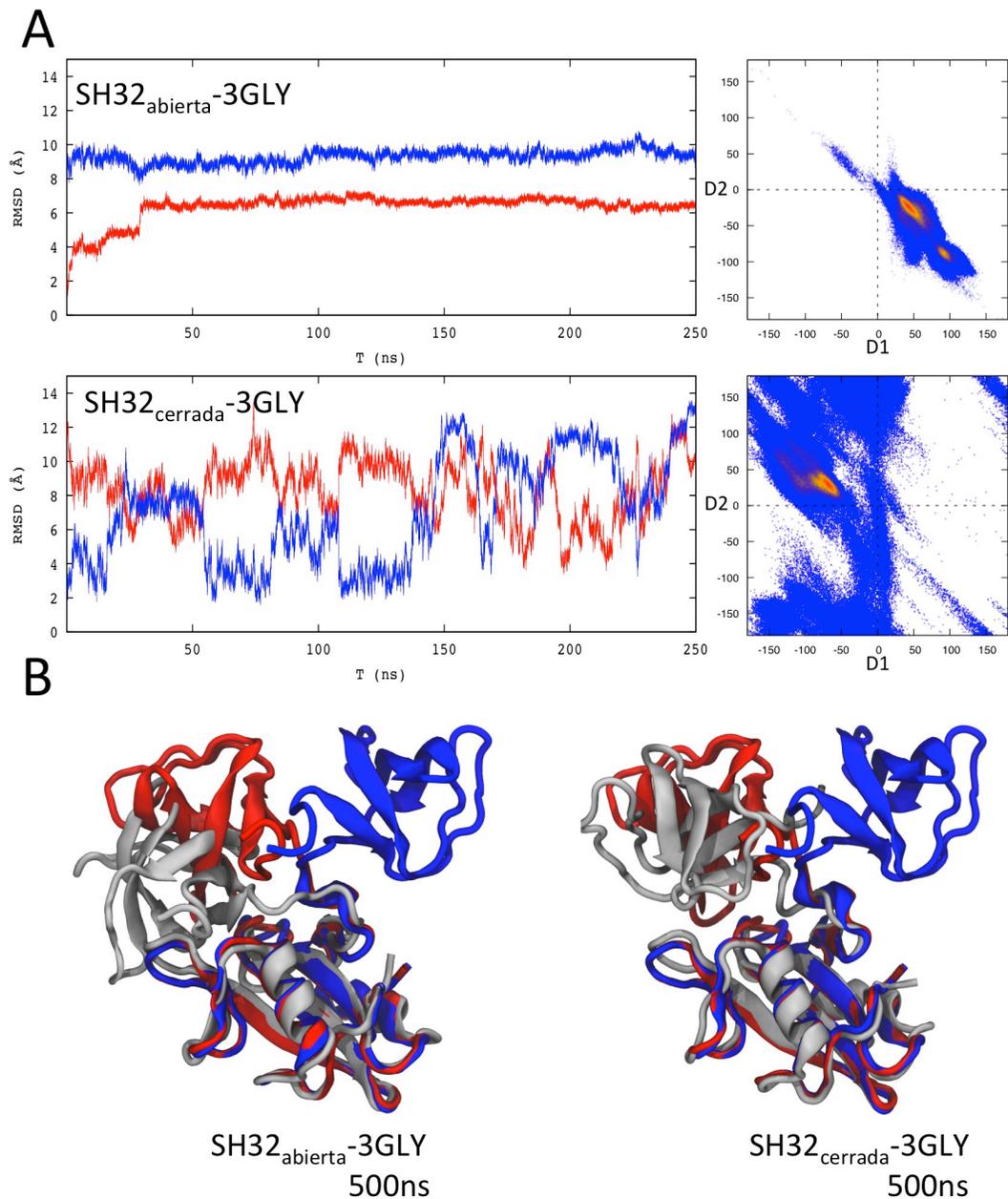
## CAPÍTULO 4

la Figura 4.17 podemos observar la comparación de los valores de los ángulos diedros entre las trayectorias en las que había ocurrido el cambio de conformación SH32<sub>abierta</sub> r1 y las dos replicas de SH32<sub>cerrada</sub> , comparadas respecto a los valores de las trayectorias SH32-C2Q r1, SH32-C2Q r2 y SH32<sub>abierta</sub> r2, donde el tándem había permanecido en la conformación inicial, es decir la cerrada. las posiciones 139, la 140 y aunque en menor medida, la 141 del segmento de C3-2 son los que muestran un cambio en la dispersión de los valores de los ángulos diedros coincidiendo con las trayectorias donde se a producido el cambio conformacional. Algo que, *a priori*, parece estar en buena sintonía con los datos experimentales que indican que la mutación a glicina de los restos SER140, LUE141 y GLU142 resulta en la activación de cAbl [58].

También llama la atención que en las trayectorias en las que el tándem SH32 es acompañado de la secuencia C2Q, la dispersión de angulos diedros es menor no solo en las posiciones mencionadas anteriormente, sino que también en todo el segmento C2-3. Lo que indicaría que la secuencia C2Q limita considerablemente el espacio conformacional explorado por el tándem SH32.



**Figura 4.17** Frecuencia de los ángulos diedros del esqueleto péptido para los residuos que componen el segmento C2Q. Comparación de las frecuencia que adquieren los ángulos diedros  $\phi$  (eje de las X) y  $\psi$  (eje de las y) a lo largo de las diferentes trayectorias. En azul se resaltan las regiones adoptadas por el residuo de 1 a 10 veces, en naranja de 10 a 100 y en rojo más de 100.



**Figura 4.18** Comportamiento del tándem SH32 en las trayectorias para los constructos SH32<sub>abierta</sub>-3GLY y SH32<sub>cerrada</sub>-3GLY. En el panel A se muestra el RMSD (paneles de la izquierda) acompañado de la frecuencia (paneles de la derecha) de los descriptores D1 y D2 a lo largo de las trayectorias. El color rojo corresponde a la trayectoria usando como referencia la estructura cristalográfica 2ABL o abierta, mientras que el color azul corresponde a la estructura 2FO0 o cerrada. En el panel B se muestra la superposición de tres estructuras, la estructura cristalográfica con código PDB 2ABL (rojo), la estructura con código PDB 2FO0 en azul y las coordenadas finales de la trayectoria . Ncap-SH32-C2Q-S69A y su replica (gris)

#### 4.2.3.1 Análisis de una triple mutación en residuos clave del conector

##### C3-2

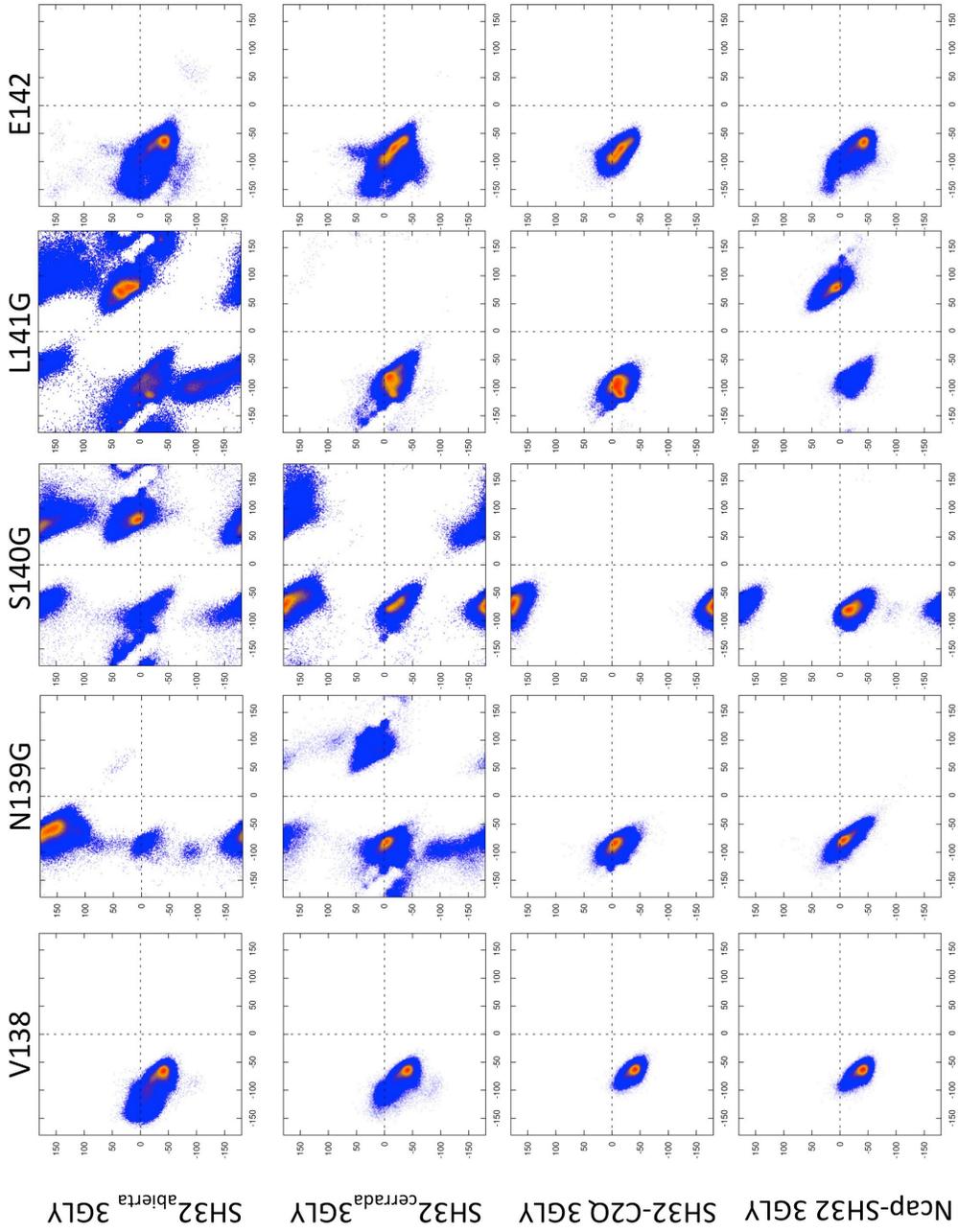
Estos resultados previos indican que los residuos ASN139, SER140 y LEU141 del conector actúan a modo articulación, modulando el movimiento de orientación relativa entre los dominios SH3 y SH2, fruto del equilibrio entre flexibilidad y rigidez conformacional de la secuencia conectora. Para comprobar que, efectivamente, estos residuos son clave para modular el equilibrio conformacional del tándem, se generaron cuatro trayectorias, donde los residuos ASN139, SER140 y LEU141 fueron substituidas por glicina, que tiene una mayor libertad conformacional. En concreto, se produjeron 250 ns de simulación de dinámica molecular con el mismo protocolo que las anteriores, una partiendo de la conformación cerrada (SH32<sub>cerrada</sub>-3GLY), otra partiendo de la conformación abierta (SH32<sub>abierta</sub>-3GLY), dos más partiendo de la conformación cerrada, una con el tándem acompañado de la región Ncap (Ncap-SH32-3GLY) y otra con la secuencia C2Q (SH32-C2Q-3GLY).

En la Figura 4.18, podemos observar el RMSD a lo largo de las trayectorias SH32<sub>abierta</sub>-3GLY y SH32<sub>cerrada</sub>-3GLY, respecto a la conformación abierta (línea roja) y cerrada (línea azul) y los descriptores D1 y D2. Los valores de RMSD obtenidos para estas trayectorias nos indican un fuerte efecto de esta mutaciones sobre el comportamiento del tándem. La trayectoria iniciada con las coordenadas de la conformación cerrada (SH32<sub>cerrada</sub>-3GLY), se encuentra explorado de forma muy errática el espacio conformacional, con una fluctuación continua del RMSD, mientras que en la trayectoria en la que se usó la conformación abierta (SH32<sub>abierta</sub>-3GLY), el tándem permanece durante toda la trayectoria en una conformación alejada en RMSD tanto de la estructura cristalográfica abierta, como de la cerrada. No obstante, como podemos apreciar por los valores de D1 y D2, aún siendo la conformación

## CAPÍTULO 4

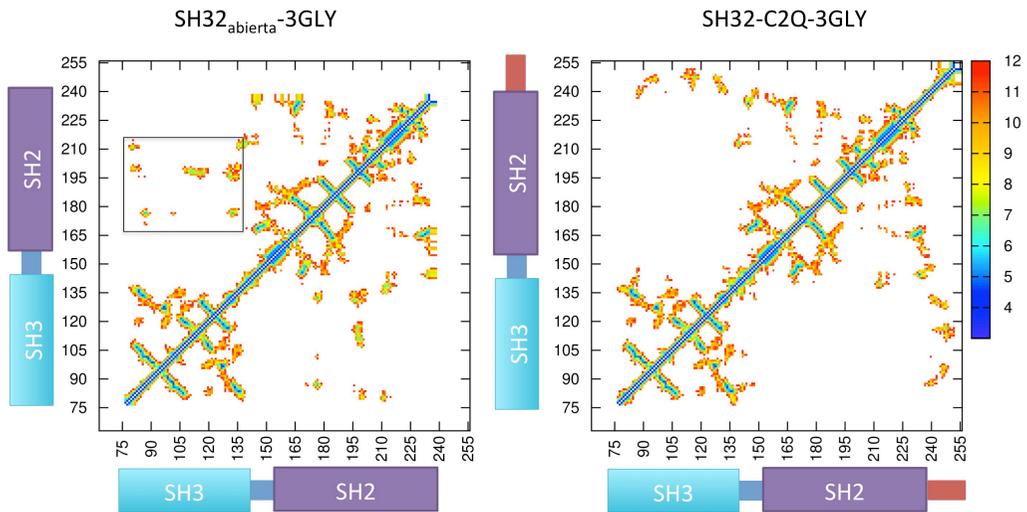
global del tándem diferente de ambas estructuras cristalográficas de partida en RMSD, el dominio SH3 permanece en una orientación relativa al SH2 cercana a la abierta. Esta variación de comportamiento de estas construcciones mutadas respecto al observado en las trayectorias con el C3-2 en forma silvestre, ponen de manifiesto un desajuste en la dinámica conformacional del tándem al introducir las glicinas en el conector C3-2, tal y como se postuló a partir de los experimentos, que conduce a la desestabilización de la conformación cerrada y posiblemente a una estabilización de la abierta.

En la Figura 4.19 se muestran los ángulos diedros del esqueleto peptídico a lo largo de las trayectorias con la triple mutación a glicina. Al comparar estos con los obtenidos en las trayectorias con el conector en su forma silvestre, observamos una mayor dispersión en los valores de los diedros, especialmente de las posiciones 139, 140 y 141. Es decir, al introducir glicinas se produce un aumento de la flexibilidad de esta región, lo que parece favorecer la adopción de la conformación abierta. Pero, ¿por qué la abierta y no otra? Al observar las diferentes trayectorias y, en especial la correspondiente a la construcción SH3<sub>abierta</sub>-3GLY, observamos cómo al adoptar la conformación abierta, los residuos de ambos dominios modulares establecen una serie de contactos intermoleculares, en concreto entre el lazo distal del dominio SH3 y los lazos  $\beta_B$ - $\beta_C$  y  $\beta_{D'}$ - $\beta_C$  de SH2. Como podemos observar en la Figura 4.20. estos contactos se encuentran ausentes en la conformación cerrada, lo que sumado a la pérdida de rigidez del conector C3-2 y la carencia de otros elementos que estabilicen la conformación cerrada, se estaría favoreciendo que el tándem adopte la conformación abierta.



**Figura 4.19** Frecuencia de los ángulos diedros del esqueleto péptido para los residuos que componen el segmento C2Q-3GLY). Comparación de las frecuencia que adquieren los ángulos diedros phi (eje de las X) y psi (eje de las Y) a lo largo de las diferentes trayectorias . En azul se resaltan las regiones adoptadas por el residuo de 1 a 10 veces, en naranja de 10 a 100 y en rojo más de 100.

## CAPÍTULO 4



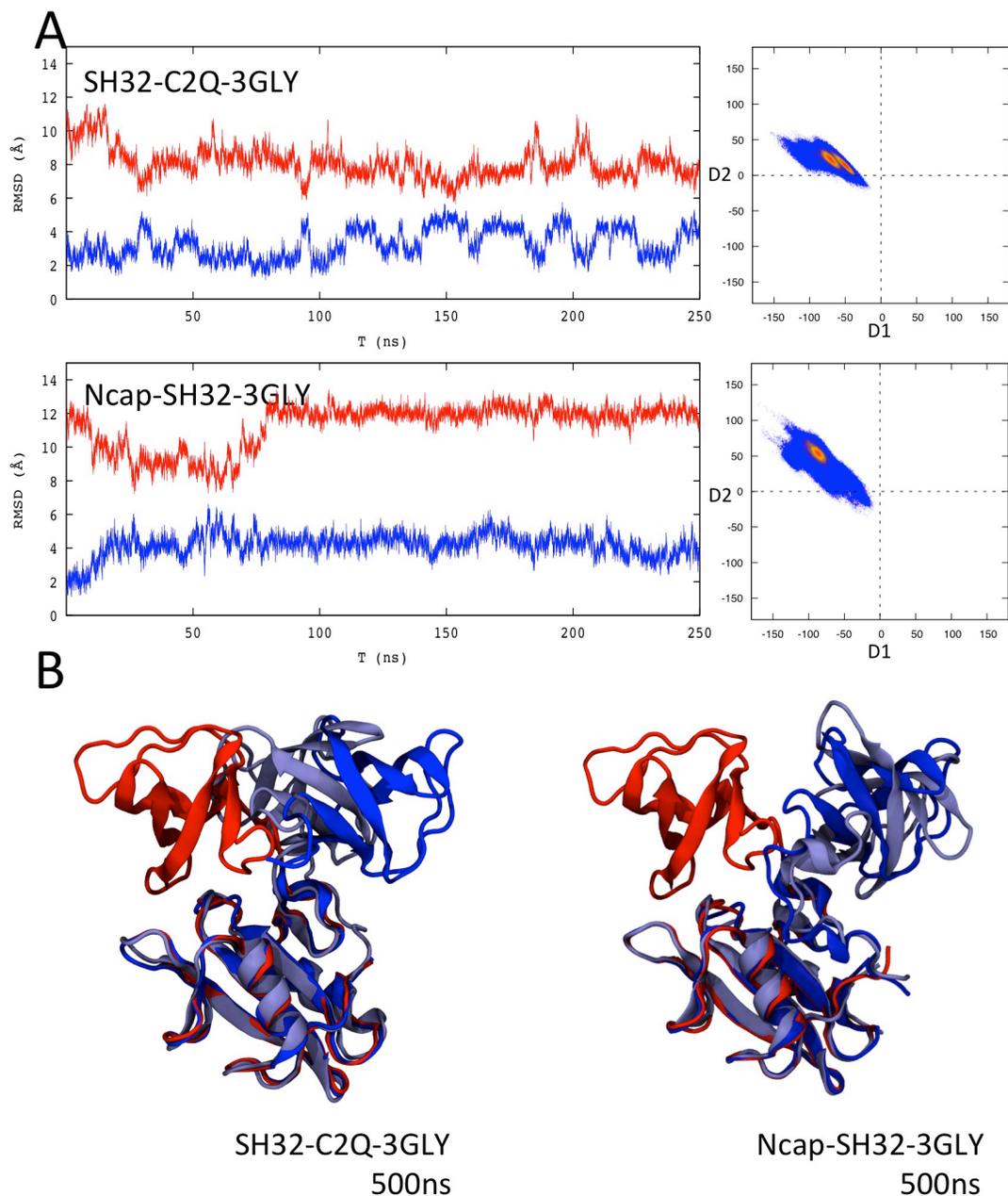
**Figura 4.20** Mapa de contactos de los  $\text{Ca}$  a lo largo de la trayectoria  $\text{SH32}_{\text{abierta}}\text{-3GLY}$  y  $\text{SH32-C2Q-3GLY}$ . Representación de la distancia promedio a lo largo de la trayectoria de los  $\text{Ca}$  de los residuos de toda la construcción mediante la escala de color que se encuentra a la derecha de la figura. No se tomaron en consideración las distancias superiores a 12 Å. La posición numérica es acompañada por un esquema para resaltar a qué dominio corresponde cada residuo. En el panel de la izquierda se ha recuadrado los contactos entre el dominio SH2 y SH3 que estabilizan la conformación que adopta a lo largo de la trayectoria.

Por ello, en la trayectoria que parte de una conformación abierta ( $\text{SH32}_{\text{abierta}}\text{-3GLY}$ ), la estructura permanece a lo largo de la trayectoria en esa conformación, mientras que en la que se partía de la conformación cerrada, ahora muy desfavorecida, la trayectoria explora diferentes conformaciones, en busca de un mínimo local, apareciendo así conformaciones que antes no estaban favorecidas o no eran accesibles. Todas estas evidencias indican que el conector de los dominios modulares, C3-2, y en concreto los residuos en las posiciones 139, 140 y 141, limita las conformaciones accesibles. Esto sugiere que, una vez activada cAbl, la orientación relativa de los dominios modulares es muy importante y, por ello, se encuentra controlada, lo que probaría que, posiblemente, el tándem esté diseñado para adoptar preferentemente un número limitado de conformaciones específicas funcionalmente relevantes, aún en ausencia de interacciones intramoleculares.

Por otro lado, en las trayectorias Ncap-SH32-3GLY, Ncap-SH32-C2Q-3GLY y SH32-C2Q-3GLY, el tándem se mantuvo en la conformación cerrada a lo largo de toda la trayectoria, como podemos apreciar por los valores de RMSD y los descriptores D1 y D2 en la Figura 4.21. En este contexto, las interacciones que establecen ambas regiones con el tándem, aunque de forma individual no son capaces de mantener la quinasa en la forma inactiva, sí que parecen modificar el espacio conformacional del tándem lo suficiente como para que, incluso mutando el conector, no seamos capaces de ver cambios conformacionales en el tiempo de simulación. Incluso, en la Figura 4.19, podemos observar cómo en las trayectorias SH32-C2Q, Ncap-SH32, Ncap-SH32-C2Q, SH32-C2Q-3GLY y Ncap-SH32-3GLY los valores de los ángulos phi y psi de los residuos del conector C3-2, adoptan y se mantienen en los valores característicos de la conformación cerrada.

Todos estos datos podrían explicar por qué las variantes resistentes de las formas oncogénicas de BCR-ABL con mutaciones en la región del conector C3-2 no responden a los inhibidores como Gleevec [134-136]. En la forma oncogénica BCR-ABL, aunque el mecanismo de autoinhibición esté dañado, el dominio SH3 debe seguir inhibiendo la actividad gracias a la unión a la secuencia C2Q, ya que este mecanismo se encuentra intacto. Una mutación en la región C3-2 puede acabar por desplazar este equilibrio hacia la conformación abierta, contrarrestando el efecto inhibitor de la interacción del dominio SH3 con C2Q, lo que aumentaría la actividad de la forma oncogénica y disminuiría la efectividad del inhibidor.

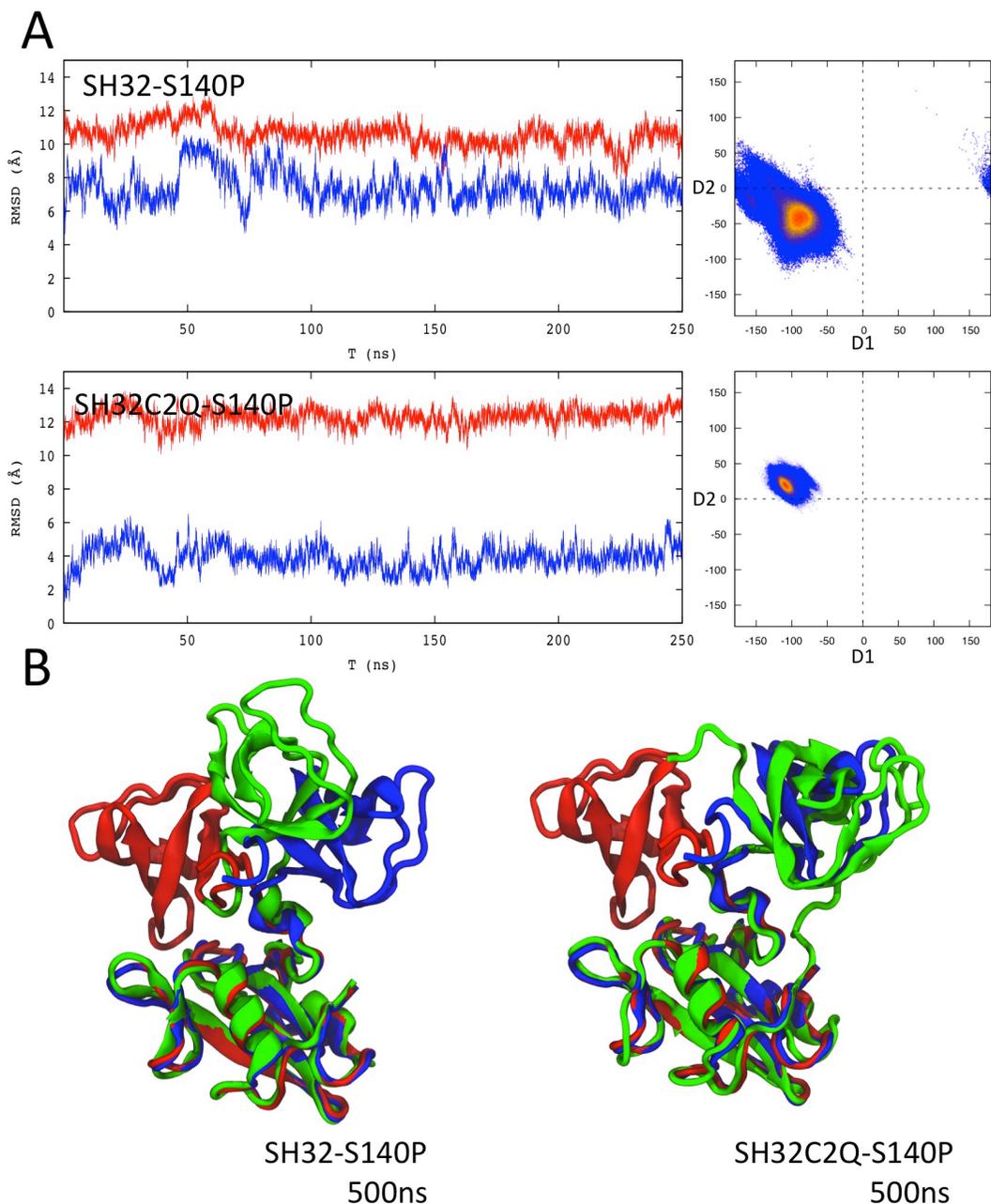
## CAPÍTULO 4



**Figura 4.21** Comportamiento del tándem SH32 en las trayectorias para los constructo Ncap-SH32-3GLY y SH32-C2Q-3GLY. En el panel A se muestra el RMSD (paneles de la izquierda) acompañado de la frecuencia (paneles de la derecha) de los descriptores D1 y D2 a lo largo de las trayectorias Ncap-SH32-3GLY y SH32-C2Q-3GLY. El color rojo corresponde a la trayectoria usando como referencia la estructura cristalográfica 2ABL o abierta, mientras que el color azul corresponde a la estructura 2FO0 o cerrada. En el panel B se muestra la superposición de tres estructuras, la estructura cristalográfica con código PDB 2ABL (rojo), la estructura con código PDB 2FO0 en azul y las coordenadas finales de la trayectoria . Ncap-SH32-C2Q-S69A y su replica (morado)

#### 4.2.3.2 Análisis de la rigidificación del segmento C3-2, mediante la mutación S140P

En este punto, se decidió probar el efecto contrario, es decir, en lugar de aumentar la flexibilidad del conector, se sustituyó la SER140 por una prolina y se observó cómo esta mutación afectaba al comportamiento dinámico. Se realizaron dos simulaciones de dinámica molecular, siguiendo el mismo protocolo que en las anteriores, para generar sendas trayectorias de 250 ns de duración. Una simulación se realizó con la estructura que corresponde al tándem en conformación cerrada, en ausencia de Ncap y C2Q, y la otra con la misma mutación, incluyendo la secuencia C2Q. Como se observa en la Figura 4.22, el RMSD de la trayectoria se mantiene alejado de las dos estructuras cristalográficas de referencia en la trayectoria SH32-S140P, mientras que en la SH32-C2Q-S140P, observamos cómo a lo largo de la trayectoria el tándem se mantiene con un valor de RMSD más cercano a la estructura cristalográfica cerrada. Si analizamos los valores de los descriptores D1 y D2, podemos observar cómo la trayectoria con la construcción SH32-S140P adopta una conformación que no coincide con los valores de la conformación abierta ni cerrada, mientras que en el caso de SH32-C2Q-S140P se mantienen cercanos a los de la conformación cerrada. Estos resultados indican que, efectivamente, la introducción de la prolina en C3-2 modifica el comportamiento dinámico del tándem y confirman la capacidad de la región C2Q de estabilizar la conformación cerrada. Al analizar los ángulos diedros de los residuos de la región C3-2, observamos cómo la introducción de la prolina ha modificado, no sólo el comportamiento la posición 140, sino también el comportamiento de los restos colindantes ASN139 y la VAL138, algo que se corrige parcialmente en la trayectoria SH32-C2Q-S140P.



**Figura 4.22 Comportamiento del tándem SH32 en las trayectorias para los constructos SH32-S140P y SH32-C2Q-S140P.** En el panel A se muestra el RMSD (paneles de la izquierda) acompañado de la frecuencia (paneles de la derecha) de los descriptores D1 y D2 a lo largo de las trayectorias Ncap-SH32-3GLY y SH32-C2Q-3GLY. El color rojo corresponde a la trayectoria usando como referencia la estructura cristalográfica 2ABL o abierta, mientras que el color azul corresponde a la estructura 2FO0 o cerrada. En el panel B se muestra la superposición de tres estructuras, la estructura cristalográfica con código PDB 2ABL (rojo), la estructura con código PDB 2FO0 en azul y las coordenadas finales de la trayectoria . Ncap-SH32-C2Q-S69A y su replica (verde)

En la Tabla 4.2 se muestra el RMSD medio para cada uno de los dominios modulares y el de la secuencia C3-2, calculado de forma individual para las trayectorias con la prolina en la posición 140. Al comparar los valores obtenidos con los anteriores Tabla 4.1 observamos cómo, a diferencia del resto, la introducción de la prolina produce un aumento generalizado del RMSD de los dominios modulares, lo que indicaría que la rigidificación del C3-2 ha provocado un cambio estructural en los dominios modulares, en especial en el dominio SH3; algo que se agrava en la trayectoria en la que se encuentra presente la región C2Q. Esto podría ser debido a la incompatibilidad en la unión del dominio SH3 a la secuencia C2Q en presencia de la prolina en la posición 140. Esto evidencia que los residuos que ocupan tanto la posición 140 como los colindantes son claves en el equilibrio conformacional del tándem y, por lo tanto, vitales para la regulación de la actividad de la quinasa de tirosina cAbl. Estos residuos no sólo ejercen de articulación en el cambio conformacional de cAbl, sino que, además, parecen modular las conformaciones posibles.

**Tabla 4.2** RMSD medio a lo largo de las trayectorias, calculado individualmente para cada dominio y el segmento C3-2, considerando solamente los átomos del esqueleto peptídico.

	SH32-S140P	SH32-C2Q S140P
<b>SH3</b>	1,95	2,85
<b>C3-2</b>	3,97	0,59
<b>SH2</b>	1,10	1,01

#### 4.2.4 Simulaciones de metadinámica

Gracias a la simulaciones de dinámica molecular hemos podido detectar los residuos clave en el equilibrio conformacional del tándem SH32 y observar cómo las regiones Ncap y C2Q modificaban este mismo equilibrio.

## CAPÍTULO 4

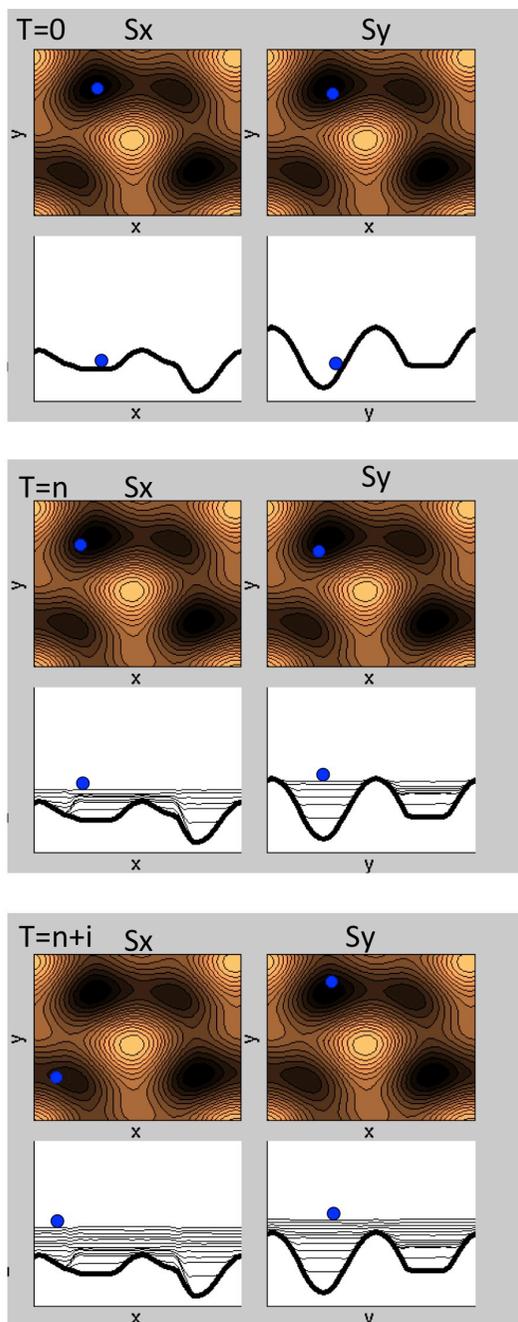
Sin embargo, seguimos sin saber si es posible que existan otras conformaciones del tándem una vez la quinasa es activada, ni tampoco si la conformación que hemos denominado como abierta observada en las trayectorias y que coincide con la estructura cristalográfica 2ABL, es la conformación final que adopta el tándem una vez la quinasa es activada, o si es sólo un estado intermedio que puede adoptar otras conformaciones por interacciones con otras proteínas o con otras regiones de la propia cAbl. Decir en este sentido que aunque la conformación sea accesible durante las trayectorias de dinámica molecular, aunque nos aporta mucha información no nos puede asegurar que se trate de la conformación mas poblada. Para responder a estos interrogantes, se decidió aplicar metodologías computacionales que nos permitieran obtener más información sobre el equilibrio conformacional del tándem y como éste cambia en presencia de los diferentes elementos estructurales.

Como se describe con detalle en la Sección de Materiales y Métodos 5.1.6, la metadinámica es una metodología que permite realizar una exploración intensiva en el espacio de una o varias coordenadas de reacción, añadiendo un sesgo en forma de potencial energético para facilitar que el sistema visite el mayor número de estados diferentes. Esto permite una estimación de la superficie de energía libre en el espacio de las coordenadas de reacción.

Durante la metadinámica el sistema es libre de explorar el espacio conformacional, al igual que en una simulación de dinámica molecular común, pero se registran todas las posiciones que visita el sistema dentro de las coordenadas de reacción. A intervalos de tiempo fijos se registra y se añade un potencial de energía siguiendo la función gaussiana de distribución. Centrada sobre el eje de las coordenadas en las que se encuentra el sistema y con una

variancia predefinida. Cada vez que el sistema vuelve a visitar las mismas coordenadas se vuelve añadir un potencial de energía, que se suma al que existía anteriormente. Esto permite que el sistema salga de mínimos locales y explore todo el espacio disponible. Una vez todos los mínimos locales han sido rellenados con este potencial energético, el sistema puede explorar el espacio de las coordenadas de reacción como si no existieran diferencias de energía libre entre ellos Figura 4.23. Una vez finalizada la metadinámica, se reconstruye la superficie de energía libre en las coordenadas de reacción mediante la contabilización de la distribución de estados. Esta superficie de energía libre nos permitirá conocer cómo de poblado está un estado según las coordenadas de reacción, siendo las áreas de baja energía libre, las más probable, así como las de alta energía, que serán las menos pobladas. Evidentemente, dicha distribución debe ser normalizada por el potencial energético añadido, para lo que se aplica la metodología conocida como WHAM (*Weighted Histogram Analysis Method*), que se describe en detalle en las siguientes referencias [226, 227].

De esta forma, la metadinámica aplicada a nuestro sistema, nos puede permitir conocer las diferencias de energía libre que existen entre las conformaciones abierta y cerrada, así como si existen otras conformaciones con posibilidad de ser pobladas. Además, podemos hacer un estudio complementario para entender cómo afectan a la distribución de conformaciones, tanto la triple mutación a glicina de la secuencia de conexión C3-2 como la presencia de las regiones Ncap y C2Q.



**Figura 4.23** Ejemplo hipotético de una metadinámica realizada con dos coordenadas de reacción  $S_x$  y  $S_y$  a lo largo del tiempo. Cada panel esta dividido en 4 secciones, las dos superiores representan la superficie de energía libre en una proyección bidimensional, es decir, para ambas coordenadas de reacción. En azul se representan las coordenadas que explora la simulación para el tiempo  $t$ . En la parte inferior de cada panel se representa la superficie de energía libre unidimensional con una línea negra y gruesa. Con líneas finas se representa el potencial de energía añadido a medida que la simulación de metadinámica avanza en el tiempo, siendo de nuevo el punto azul las coordenadas que se encuentra explorando para ese instante  $t$  [147]

Como está detallado en la Sección 5.6 de Materiales y Métodos, la selección de las coordenadas de reacción es un aspecto crítico en estos cálculos, ya que debemos definir con la mayor exactitud posible el proceso bajo estudio. Estas coordenadas han de ser lo suficientemente amplias como para no enmascarar estados relevantes para el proceso bajo estudio, al tiempo que no deben ser demasiado grandes, ya que llevaría mucho tiempo completar la exploración de todo el espacio. Una estrategia para asegurar que estamos usando unas coordenadas de reacción que describan correctamente todos los grados de libertad del proceso es aumentar el número de coordenadas; es decir, aumentar la dimensionalidad del espacio a explorar. Desafortunadamente, utilizar más de tres coordenadas, es sencillamente demasiado costoso. Pero el uso de más de tres coordenadas de reacción es posible si aplicamos la metodología conocida como intercambio de sesgos (Bias Exchange [36]) a las metadinámicas. Esta metodología consiste en lanzar varias réplicas de forma paralela, cada una de ellas explorando una o dos coordenadas de reacción diferentes y, a intervalos de tiempo fijado, se emparejan al azar para intercambiar las conformaciones que se encuentran explorando entre ellas, siempre y cuando se acepte el criterio probabilístico conocido como METROPOLIS, detallado en la Sección 5.1.6 de Materiales y Métodos, donde la probabilidad de cambio es inversamente proporcional a la diferencia de energía potencial entre las coordenadas de reacción de las conformaciones. Con la intención de realizar una exploración lo más intensiva posible del equilibrio conformacional del tándem, se seleccionaron cinco coordenadas de reacción. Las tres primeras, llamadas S1, S2 y S3, se basan en el RMSD de la estructura durante la dinámica respecto al de las cristalográficas (abierta y cerrada), mientras que las que hemos llamado D1 y D2 son ángulos diedros formados por los centros de masas de varios residuos que, como

## CAPÍTULO 4

hemos mencionado anteriormente, definen la orientación relativa de los dominios modulares (Sección 4.2.1.1)

Para las coordenadas de reacción basadas en el RMSD es necesario definir unos límites para impedir la exploración indefinidamente del espacio de RMSD. Estos límites pueden, por un lado, enmascarar estados intermedios inesperados, y, por otro, aumentar el espacio de exploración innecesariamente. Por ello, y con la intención de llegar a un compromiso, hemos utilizado lo que se conoce como coordenadas de reacción de camino (*path collective variables*). Como se describe en detalle en la Sección 5.6.1 de Materiales y Métodos, esta metodología ha demostrado su eficiencia para describir procesos de dos estados finales A y B definidos *a priori*. En nuestro caso estos dos estados finales corresponden a las estructuras la abierta y la cerrada. Las coordenadas de reacción de camino nos indican en qué punto se encuentra una configuración dada en el espacio que conecta ambas formas, obteniendo un valor que indicaría a cuál de estas dos se encuentra más próxima. Tanto S1 como S3 usaron como referencia la conformación abierta (2ABL) y la cerrada (2F00), pero mientras en S1 evaluamos el RMSD de todo el tándem, en S3 solamente se calculó el RMSD de la región de conexión C3-2, comprendida entre los residuos (238 a 253). Por otro lado, la coordenada de reacción S2 es complementaria a las anteriores, siendo su misión la de impedir la deriva excesiva de las estructuras en el espacio del RMSD que separa ambas estructuras finales, y nos indica la distancia a la que se encuentra una configuración dada del camino que conecta ambos estados finales.

Estas coordenadas de reacción aunque son ideales para este tipo de casos en los que queremos comparar dos estados y averiguar si existen otros intermedios no esperados. Sin embargo, es de vital importancia el uso del los

descriptores D1 y D2 para completar el estudio, ya que estos descriptores definen muy bien la orientación relativa de los dominios .

Para el estudio del tándem de dominios modulares de cAbl se produjeron 4 metadinámicas diferentes. Una partiendo de la estructura del tándem aislado (SH32<sub>m</sub>), una segunda del tándem con la triple mutación a glicina en el conector C3-2 (SH32-3GLY<sub>m</sub>), una tercera con la región Ncap (Ncap-SH32<sub>m</sub>) y finalmente el tándem acompañado de secuencia de conexión C2Q (SH32-C2Q<sub>m</sub>). Como estructuras de partida se tomaron las coordenadas resultantes del último paso de equilibración de las dinámicas de la sección anterior. Siguiendo el protocolo de la Sección 5.6 de Materiales y Métodos, se generaron 32 replicas para cada metadinámica, cada una sesgada en dos coordenadas de reacción diferentes permitiendo el intercambio entre ellas (Bias Exchange) y produciendo, para cada replica, 400 ns de simulación, de forma que para cada estructura se llevó a cabo un total de 12800 ns de simulación.

En el Apéndice 9.3, se muestran las superficies bidimensionales de energía libre reconstruidas a partir de cada simulación de metadinámica en función de las coordenadas de reacción de camino S1 y S2, S3 y S1 y por los descriptores D1 y D2. No obstante, la proyección de la superficie de energía libre en estas coordenadas de reacción, aunque aporta información interesante es inusual y requiere una descripción más detallada. Con el objetivo de mejorar la compresión de los datos obtenidos por estas simulaciones de metadinámicas se decidió reconstruir la superficie de energía libre en un espacio bidimensional pero utilizando como variables de reacción el RMSD respecto a la conformación abierta y cerrada del tándem. Es decir, para cada instante de las trayectorias de las metadinámicas, se obtuvo el RMSD respecto a la conformación abierta (2ABL) y cerrada (2FO0), y se

## CAPÍTULO 4

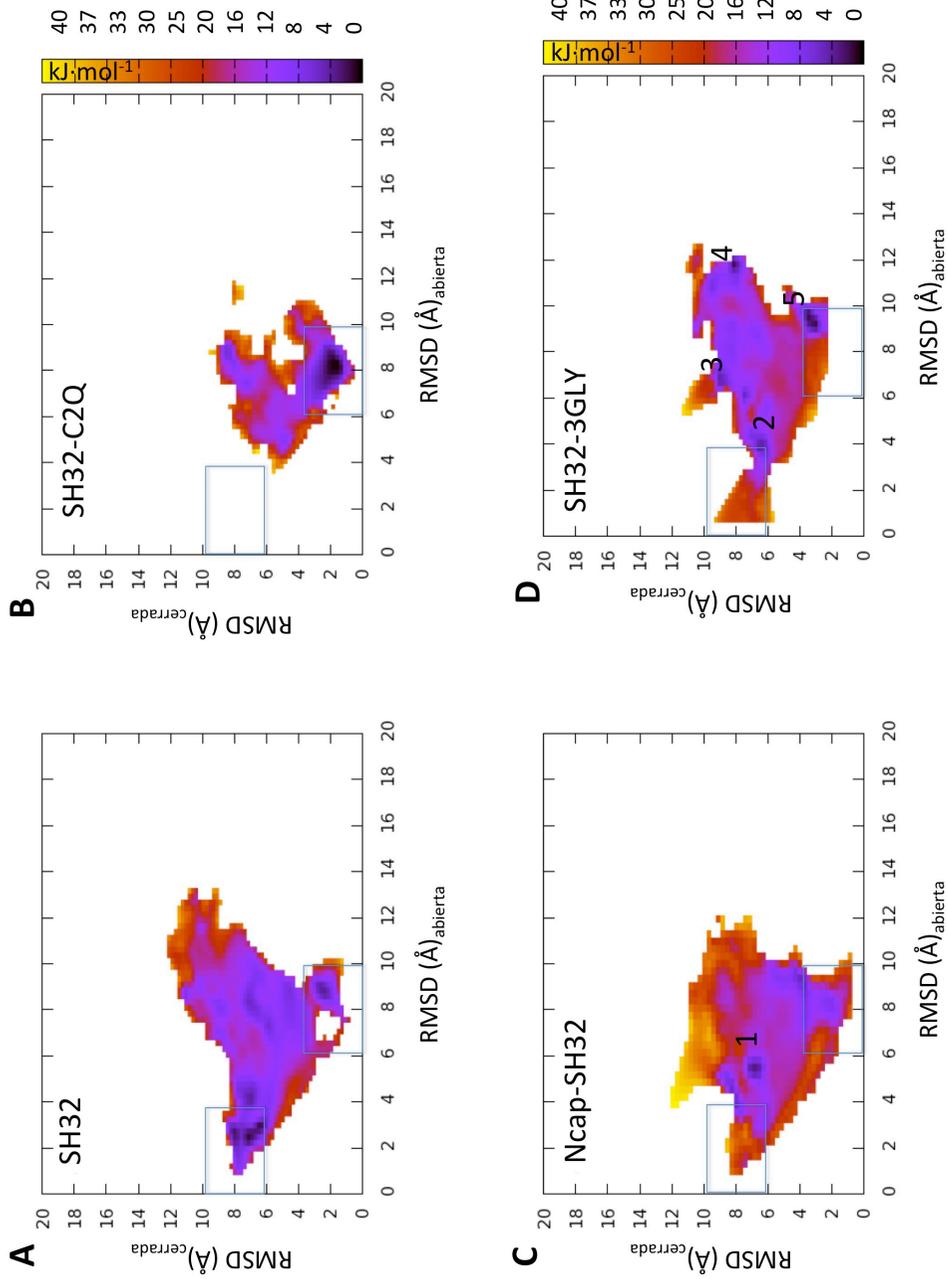
reconstruyó la superficie de energía libre como si hubiéramos usado el RMSD como coordenadas de reacción. Con ello obtenemos unas superficies de energía libre de más fácil comprensión, que se muestran en la Figura 4.24, donde dedicaremos especial atención a los valores de energía libre que corresponden a la conformación cerrada y abierta de las estructuras cristalográficas, y resaltadas en la figura con un rectángulo.

En la Figura 4.24 panel A se representa la superficie de energía libre que corresponde a la metadinámica del tándem aislado SH32<sub>M</sub>, donde observamos la existencia de dos regiones principales donde se aglutinan las coordenadas con los valores más bajos de energía libre ( $< 4 \text{ kJ}\cdot\text{mol}^{-1}$ ), uno situado en el área que corresponde a la conformación cerrada y el segundo en la conformación que corresponde a la abierta. En un segundo plano encontramos una amplia región alejada en RMSD de ambas estructuras, con valores de energía libre entre 4 y 8  $\text{kJ}\cdot\text{mol}^{-1}$ . Esta conformación, al igual que la abierta, presenta una disposición relativa de los dominios que resulta difícil de adoptar si se mantienen las interacciones intermoleculares típicas de cAbl cuando se encuentra autoinhibida. Esto nos indica que, en ausencia de otros elementos de la quinasa o de otras proteínas, el tándem presenta una distribución conformacional principalmente bimodal. La región que corresponde a la conformación cerrada, se encuentra limitada por una pequeña barrera con valores de energía libre de 20  $\text{kJ}\cdot\text{mol}^{-1}$ .

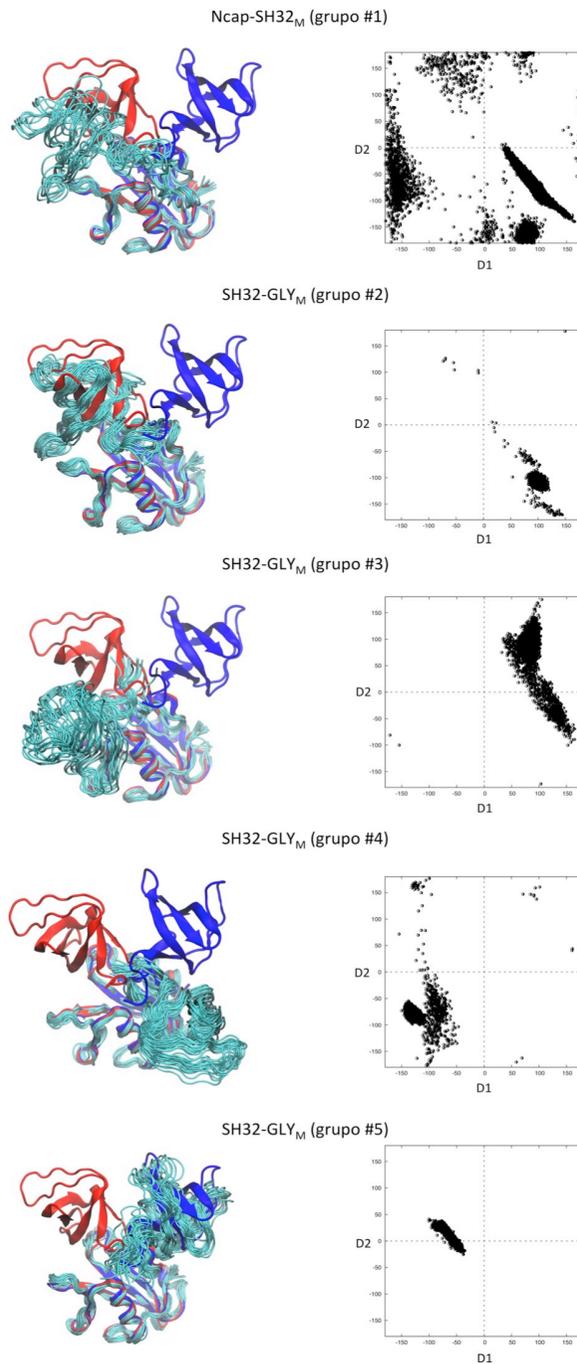
En la Figura 4.24 panel B mostramos la superficie de la energía libre obtenida de las simulaciones de metadinámica de la construcción SH32-C2Q<sub>M</sub> en el espacio del RMSD, donde observamos una única y amplia región donde se concentran los valores mínimos de energía libre. Esta área corresponde a la configuración cerrada del tándem. El área colindante a esta región incrementa rápidamente su energía libre a medida que nos alejamos en RMSD de las

coordenadas cercanas a la configuración cerrada, estando la región más cercana a la configuración abierta por encima de los 50 kJ.mol<sup>-1</sup>. Todo esto confirma que la presencia del conector C2-Q, desplazara la distribución conformacional del tándem hacia la conformación cerrada, de forma que debe ser indispensable la pérdida de interacciones entre SH3 y el C2-Q para el cambio de conformación y activación de cAbl. Esto refuerza el papel del conector C2Q como un elemento vital en el mecanismo de autoinhibición de cAbl.

Es interesante señalar que cuando se compara la superficie de energía libre obtenida para el tándem aislado SH32<sub>M</sub> con la obtenida para el constructo con la región conectora C2Q (SH32-C2Q) el área de mas baja energía y de menor RMSD respecto a la estructura cerrada no se encuentra en la misma posición en ambos casos. Así, en la simulación de metadinámica del constructo SH32<sub>M</sub>, se encuentra ligeramente desplazado y aislado, es decir rodeado por valores de energía libre elevados, alrededor de los 20 kJ.mol<sup>-1</sup>. Esto podría indicar, por un lado, que la conformación observada en la estructura cristalográfica 2FO0 solamente se encuentra poblada siempre y cuando el conector C2Q se encuentre disponible para interactuar con el dominio SH3 y, por otro, que una vez el tándem adopta una conformación cercana a la cerrada, ésta puede mantenerse en durante un tiempo al quedar dentro de una región con baja energía rodeada por una barrera energética elevada, C2Q.



**Figura 2.24** Proyección de la superficie de energía libre en el espacio del RMSD respecto a las configuraciones cerrada (correspondiente al PDB con código 2FO0) y abierta (correspondiente al PDB con código 2ABL). Con un cuadrado azul de 4 Å de lado se resalta la región que corresponde a la conformación cristalográfica abierta y cerrada.



**Figura 4.25** Muestra representativa de las estructuras que se encuentran en coordenadas de interés de la superficie de energía libre. En los paneles de la derecha se presenta 1% de las estructuras que forman el grupo (cian) superpuesta con las estructuras cristalográficas correspondientes a la conformación abierta, (rojo) y la conformación cerrada (azul). En los paneles de la izquierda se representa los valores de D1 y D2 para todos los miembros del grupo.

## CAPÍTULO 4

Al estudiar la superficie de energía libre para el constructo con la región Ncap en el panel C de la Figura 4.24, lo primero que observamos es que la región que corresponde a la conformación abierta presenta valores de energía libre muy altos, entre 25 y 16  $\text{kJ}\cdot\text{mol}^{-1}$ , es decir, que esta conformación en presencia del segmento Ncap es muy poco probable en la distribución conformacional del tándem. Por el contrario, la conformación cerrada presenta valores de 8  $\text{kJ}\cdot\text{mol}^{-1}$ , estando más favorecida. Además observamos que, a diferencia de lo observado en la metadinámica de SH32<sub>M</sub>, esta región no se encuentra limitada por un área de alta energía libre, lo que nos indicaría que la unión de Ncap al tándem desplaza el equilibrio conformacional del tándem hacia la conformación cerrada, y, por tanto, facilita la unión de la región C2Q al dominio SH3. Esto podría indicar que la interacción con el Ncap pueda constituir un paso previo a la unión del C2Q cuando cAbl pasa del estado activado al autoinhibido.

En estas gráficas también observamos dos regiones de menor energía libre de la superficie, ( $<4 \text{ kJ}\cdot\text{mol}^{-1}$ ) en un área alejada en RMSD de la conformación abierta y cerrada. Las estructuras que se encuentran en esta área, como se observa en detalle en la Figura 4.25, aunque presentan un RMSD de más de 6 Å respecto a la conformación abierta, tienen una orientación relativa de los dominios modulares parecida. De esta forma, en presencia de Ncap el tándem puede poblar con facilidad conformaciones con los dominios en una orientación relativa diferente a la de la conformación cerrada, pero ambos mínimos se encuentran separados por una amplia barrera con valores de energía libre altos. Esto junto a los resultados de las simulaciones de dinámica molecular sugiere que las interacciones de Ncap con el tándem, tienen una doble función. Por un lado, como hemos observado en las simulaciones llevadas a cabo con la construcción Ncap-SH32-S69A, la pérdida

de las interacciones que establece Ncap con el tándem desencadena que el tándem cambie rápidamente de conformación. Esto indicaría que Ncap actúa como interruptor del cambio conformacional del tándem hacia la conformación abierta. Pero otro lado, como observamos de la superficie de energía libre, mientras Ncap permanece unido al tándem no se favorece únicamente la conformación cerrada, sino que aparece una tercera conformación no observada en ningún otro caso. De forma que además de interruptor Ncap actuaría como modulador del equilibrio conformacional, favoreciendo una conformación del tándem que en el contexto de cAbl podría corresponder a un estado activo con sus propiedades únicas o bien a un estado intermedio.

El efecto modulador de Ncap sobre el equilibrio conformacional del tándem que especialmente patente al comparar la superficie de energía libre proyectada para los descriptores D1 y D2 (ver Apéndice 9.2) para las diferentes simulaciones de metadinámica. En contraste al resto de constructos, en la metadinámica Ncap-SH32M encontramos que la mayoría de las coordenadas en el espacio conformacional de D1 y D2, presenta valores por encima de los 50 kJ·mol<sup>-1</sup>, quedando solamente dos pequeñas áreas con valores por debajo los 4 kJ·mol<sup>-1</sup> y que por tanto se encontrarán pobladas en la distribución conformacional.

Esto sería un buen indicativo, de que en efecto existen diferentes estados de activación de la quinasa de tirosina cAbl dependiendo del estímulo de activación. Es decir, si cAbl es activada mediante la disrupción de las interacciones intramoleculares del dominio SH3, el tándem puede no encontrarse en la conformación cerrada, pero mientras Ncap permanezca unido al tándem, esta conformación no será la que hemos definido como abierta, sino la tercera conformación mencionada anteriormente. Aunque es

## CAPÍTULO 4

cierto que ambas conformaciones se encuentran separadas por una amplia barrera. Por otro lado, si cAbl, una vez activada, también pierde las interacciones de Ncap, por ejemplo por la acción de una fosfatasa de serina, el tándem adoptaría mayoritariamente la conformación que hemos definido como abierta.

Aunque con la información disponible tampoco podemos descartar la posibilidad que para que el tándem en conformación abierta pueda volver a la conformación autoinhibida, la unión de Ncap al tándem fuera un paso previo, es decir que sería necesaria la acción de una quinasa de serina para fosforilar la SER69 y así restablecer las interacciones con el tándem. Es decir se trataría de un paso intermediario de la activación de cAbl, mas que un estado activado diferente.

Por otro lado en el mapa de superficie para la energía libre de la estructura con la secuencia de conexión C3-2 cuyos residuos clave han sido substituidos por glicinas (SH32-3GLY<sub>M</sub>), Figura 4.24 panel D observamos la existencia de varias regiones con bajos valores de energía libre (<4 kJ·mol<sup>-1</sup>) repartidas por varias zonas del espacio del RMSD, la mayoría en regiones de RMSD alejadas tanto de la estructura cerrada como de la abierta y adoptando orientaciones relativas de los dominios (Figura 4.25) incompatibles con la estructura de cAbl autoinhibida. Asimismo estas regiones se encuentran delimitadas por amplias zonas con valores de mas de 20 kJmol<sup>-1</sup> de energía libre. El enorme cambio de la distribución conformacional respecto a la forma nativa del tándem confirman la importancia de los residuos en las posiciones 139, 140 y 141. en la determinación del equilibrio conformacional del tándem. También es interesante destacar cómo las dos conformaciones mayoritarias de la distribución conformacional en el tándem en su forma nativa, SH32M, están separadas por una barrera de energía libre que no presenta valores

superiores a los 16 kJ·mol<sup>-1</sup>, mientras que en SH32-3GLY, las conformaciones más probables se encuentran separadas por regiones con valores que llegan a superar los 25 kJ·mol<sup>-1</sup>. Es decir, una molécula del tándem nativo que se encuentre en la conformación abierta, puede cambiar con relativa frecuencia a la conformación cerrada y viceversa, lo que a su vez se correlaciona muy bien con el carácter transitorio de la activación de cAbl. Todo lo contrario que sucede al mutar el conector C3-2, SH32-3GLY, donde una vez una molécula del tándem adopta una de las conformaciones de baja energía libre, debe superar una alta barrera energética para cambiar de conformación, de forma que no se observaran cambios conformacionales con la misma frecuencia. Este comportamiento podría explicar porque las formas de BCR-ABL con mutaciones en la posición 140 presenta una especial resistencia al inhibidor Gleevec, ya que no solamente la conformación cerrada se encontraría poco poblada, sino que además, el equilibrio conformacional estaría desplazado hacia conformaciones donde el tándem presenta una organización de dominios incompatibles con el mecanismo de autoinhibición, eliminando cualquier efecto remanente que pudiera tener el dominio SH3 de la forma oncogénica sobre la regulación.

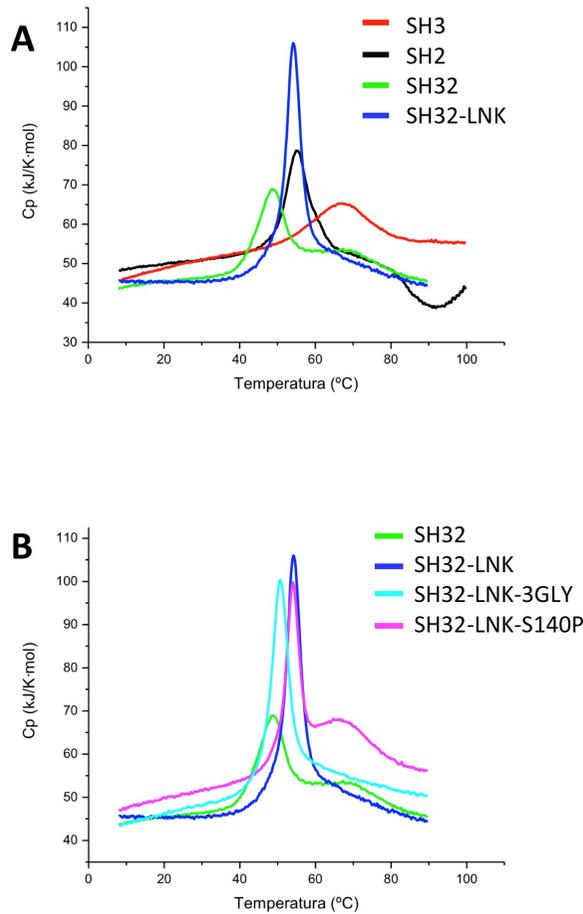
#### **4.2.5 Estudio por calorimetría diferencial de barrido de las construcciones del tándem SH32**

Con la intención de validar experimentalmente las conclusiones obtenidas de las simulaciones de dinámica molecular, en nuestro grupo de investigación se ha abordado, de forma paralela, el estudio termodinámico del equilibrio conformacional de las diferentes construcciones del tándem SH32 de cAbl (Apéndice 9.2). Para ello se clonaron las secuencias que corresponden al dominio SH3, al dominio SH2, al tándem SH32, al tándem con la secuencia

## CAPÍTULO 4

C2Q (SH32-C2Q) en *E. Coli* Dichas construcciones se purificaron y expresaron siguiendo el protocolo del Apéndice 9.1. Se procedió a la caracterización mediante calorimetría diferencial de barrido de su equilibrio conformacional. Desgraciadamente, no ha sido posible encontrar condiciones en las que el despliegamiento térmico de los dos dominios, SH2-cAbl y SH3-cAbl aislados sea simultáneamente reversible, lo que ha imposibilitado el análisis cuantitativo de las trazas calorimétricas y la obtención de parámetros termodinámicos fiables. No obstante, como se describirá a continuación, la comparación de las trazas calorimétricas nos permitirá inferir algunos resultados interesantes de forma cualitativa.

En la Figura 4.26 panel A, se comparan los termogramas del tándem en ausencia y presencia de la secuencia C2Q, así como los correspondientes a los dominios SH3 y SH2 de forma individual. En ausencia de la secuencia C2Q, podemos observar que existen dos transiciones, que corresponderían aparentemente con la desnaturalización de los dominios individuales. En presencia de C2Q, se observa un único pico muy agudo, lo que indica que la interacción con la secuencia de conexión induce una cierta cooperatividad en el desplegamiento del tándem. Este cambio en el comportamiento, junto a los experimentos de intercambio de hidrógeno/deuterio y de espectrometría de masas, así como el comportamiento del tándem en presencia del C2Q indican que el C2Q de cAbl tiene una afinidad suficiente por el dominio SH3 como para mantenerse unido a él, en ausencia del dominio quinasa y que esa interacción de alguna manera se transmite al dominio SH2 mediante interacciones cooperativas.



**Figura 4.26 Dependencia de la capacidad calorífica molar parcial con la temperatura para las diferentes construcciones ensayadas, obtenidas a partir de los experimentos en fosfato sódico 50 mM pH7.0 .** Los símbolos representan los datos experimentales, mientras que las líneas continuas corresponden al mejor ajuste al modelo de equilibrio de desplegamiento de dos estados. En el Panel A, se representan los termogramas para las construcciones: dominio SH3 (rojo), dominio SH2 (negro), tándem SH32 (verde) y tándem SH32 acompañado de la secuencia de conexión C2Q (azul). En el panel inferior se representan los termogramas correspondientes a las siguientes construcciones: tándem de los dominios SH32 (verde), tándem SH3 SH2 acompañado de la secuencia de conexión C2Q (azul), tándem SH32 acompañado de la secuencia de conexión C2Q con el conector de los dominios C2Q con las posiciones 139, 140 y 141 mutadas a glicina (cian) y tándem SH32 acompañado de la secuencia de conexión C2Q con el conector de los dominios C2Q con la posición 140 mutada a prolina (rosa).

En la Figura 4.26 panel B, se muestran los termogramas para las diferentes construcciones del tándem aislado (SH32), con la secuencia C2Q (SH32-C2Q), así como las respectivas construcciones con las mutaciones en la secuencia de conexión entre dominios C3-2 (SH32-C2Q-S140P y SH32-C2Q-

## CAPÍTULO 4

3GLY). Podemos observar que en la presencia de la triple mutación a glicina, el comportamiento del tándem no cambia significativamente con respecto a la forma salvaje, existiendo, un sólo pico muy agudo, aunque la  $T_m$  se ha desplazado algunos grados a temperaturas más bajas, lo que indica una desestabilización asociada, probablemente, al aumento de la entropía conformacional del estado desplegado.

En cambio, la presencia de la prolina en la construcción SH32-C2Q-S140P genera un termograma en el que aparecen dos picos, al igual que en el tándem SH32 aislado, de modo que, aún en presencia de la secuencia de conexión C2Q, los dominios permanecen desacoplados. Esto indica la pérdida de cooperatividad con respecto a la construcción silvestre SH32-C2Q. Esto está de acuerdo con los resultados de las simulaciones, indicando que el resto prolina en la secuencia de conexión entre ambos dominios impide la adopción de la conformación necesaria para interactuar con la secuencia C2Q. En las trayectorias de dinámica molecular, el dominio SH3 sufre un cambio de conformación, algo que puede estar relacionado con la tensión entre el cambio de conformación que provoca la prolina y la interacción con la secuencia C2Q.

Por consiguiente, aún a nivel cualitativo, los experimentos de DSC validan las principales conclusiones del estudio computacional. Esto es, confirman la relevancia de la región C3-2 en la determinación del equilibrio conformacional del tándem SH32. Así mismo, confirman que la secuencia C2Q es capaz de modificar el comportamiento del tándem en ausencia de ningún otro elemento de la quinasa, lo que indica que esta interacción podría tener una afinidad mayor que la detectada para la subfamilia Src, adquiriendo, por lo tanto una mayor relevancia en la inactivación de la quinasa.

Finalmente, hemos visto cómo las interacciones de la secuencia C2Q, en ausencia de la quinasa, reconfiguran el espacio conformacional del tándem,

haciendo que la configuración más estable sea la cerrada. Esto, sumado al desplegamiento cooperativo observado por DSC del tándem SH32 en presencia de C2Q, reforzaría la idea de que la interacción entre C2Q y el dominio SH3 es mayor de lo que se pensaba y que respondería a la necesidad de aumentar la estabilidad de la conformación cerrada en ausencia de las interacciones que establece el extremo C-terminal con SH2 en la subfamilia de quinasas de tirosina de Src.

### 4.3 CONCLUSIONES

Del análisis del equilibrio conformacional del tándem de los dominios SH3 y SH2 de cAbl realizado se desprenden las siguientes conclusiones:

1) El tándem SH32 de cAbl en ausencia de las interacciones de Ncap o C2Q se encuentra en un equilibrio conformacional en el que las dos orientaciones relativas detectadas mediante el análisis de las estructuras cristalográficas son mayoritarias.

2) El equilibrio conformacional del tándem se encuentra modulado gracias a las propiedades de la secuencia que conecta ambos dominios modulares, C3-2. El conector C3-2 es responsable de limitar la entropía conformacional del tándem de dominios modulares, otorgándole suficiente libertad para que el dominio SH3 se desplace para reconocer ligandos y liberar el dominio catalítico, a la vez que limita las conformaciones posibles para que el dominio SH3 se mantenga unido a la secuencia enlace C3-2. Se destaca asimismo el papel de los residuos que se encuentran en las posiciones 139, 140 y 141.

## CAPÍTULO 4

3) Tanto las simulaciones de dinámica molecular, como las de metadinámica indican que la región Ncap tendrá un papel destacado en el mecanismo de regulación de la quinasa cAbl, pudiendo actuar tanto como interruptor de la actividad como un modulador del estado activado de cAbl.

---

**5 MATERIALS AND METHODS**

CAPÍTULO 5

## 5.1 An introduction to molecular dynamics of biomolecules

A dynamic model has replaced the early view of proteins as relatively rigid structures. The internal motions and conformational changes play an essential role in their function. These can be small scale motions involving atomic fluctuations around an average structure, or can involve large-scale reorganization of the molecular machinery. The experimental study of these internal motions is not trivial because most experimental techniques provide average values for the entire ensemble of conformations making it hard to dissect the individual contribution of each element of the biomolecule to the macroscopic observable. In contrast, molecular dynamics (MD) simulations can provide detailed information of a system at atomic level, calculating the positions of all atoms in the macromolecule as a function of time from an initial set of coordinates. This technique provides mechanistic information at the microscopic level of relevance to address specific questions about the molecular origin of the macroscopic properties of the system. Consequently, over the last decades, MD simulations have emerged as an extremely valuable and widely used technique, providing a complementary perspective to experimental studies.

### 5.1.1 The fundamentals of the methodology

The essential principles of MD were established in the late 1950s [148] and early 1960s [149]. Fifteen years later, with more computational power available and more elaborated methodological concepts, the first protein simulation was performed [150, 151]. In principle, the simulation of the behaviour of atomic particles would require an exact quantum mechanical treatment of every atom within a molecular system [152]. However, this *ab initio* method often takes enormous amounts of computer time and memory,

## CAPÍTULO 5

being only analytically possible for small molecules. The problem of computational expense can be alleviated through simplification schemes, using approximations to the wave functions. Nonetheless, for large molecules this approach introduces errors that escalate exponentially as the molecule increases in size and complexity. An alternative way to describe the interactions between atoms is based in the application of Newton's laws of motion to the atoms of a particular system. Eq (5.1).

$$(Eq. 5.1) \quad \vec{F} = m\vec{a}$$

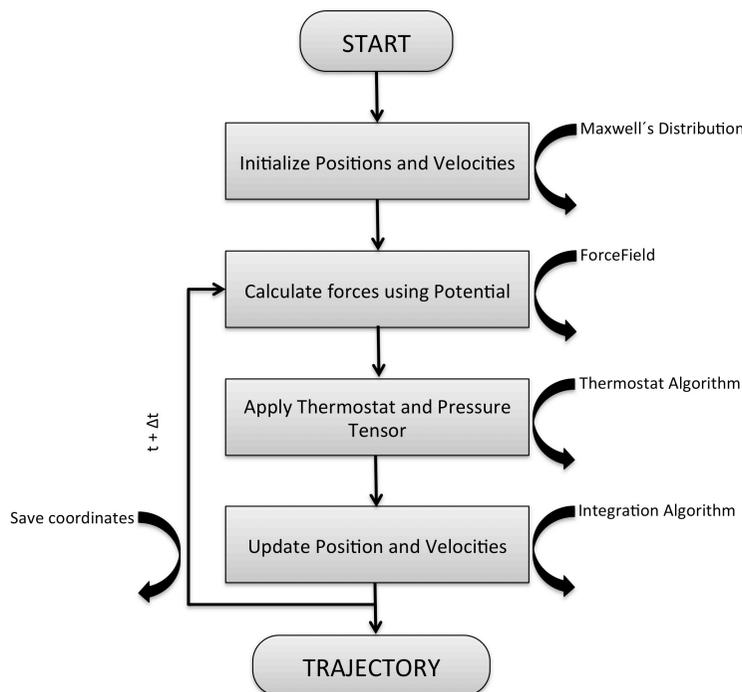
In this approach, each atom in the system is considered as a point with a mass ( $m_i$ ), corresponding to the nuclear mass of the atom. Then, the molecule is treated as a classical system of particles where the motion is driven by the momentum of each particle. Every particle  $i$  in the system is described by Newton's differential equation of motion Eq (5.2).

$$(Eq. 5.2) \quad \frac{d^2r_i(t)}{dt^2} = \frac{F_i}{m_i}$$

$$(Eq. 5.3) \quad F_i = -\partial V(r_1, \dots, r_N) / \partial r_i$$

For this purpose we need to be able to calculate the force, denoted by  $F_i$ , at certain time ( $t$ ) acting on the atom  $i$  at coordinates  $r_i$ . This approach requires calculation of the local gradient of the potential energy  $V(r)$ , which, therefore, must be a differentiable function of atomic coordinates  $r_i$ . (Eq. 3). The potential energy map is defined by a set of semi-empirical equations, known as the force field, and defines the behaviour of the atomic relations, such as electrostatic and van der Waals interactions, as a function of the position of every particle in the system. This will be described in detail in the Section 5.1.4

Any simulation will start from an initial set of atomic positions and velocities. The initial configuration for a molecular system can be obtained from different sources: X-ray structures, nuclear magnetic resonance (NMR), computational models, etc. On the other hand, the kinetic energy of any atom is related to the state of the system, and more specifically to its temperature. At any time, the velocities for each atom are chosen randomly from a Maxwell-Boltzmann or Gaussian distribution at a given temperature, which gives the probability that an atom  $i$  has a velocity  $v_{ix}$  in the  $x$  direction at a temperature  $T$ .



**Figure 5.1** The Global Molecular dynamics algorithm.

Finally, the integration of equation 2 in a small time-step ( $\Delta t$ ), returns the position of the particles in next time interval. The iteration of this process, results in a trajectory, *i.e.*, a set of discrete coordinates for each atom for each

## CAPÍTULO 5

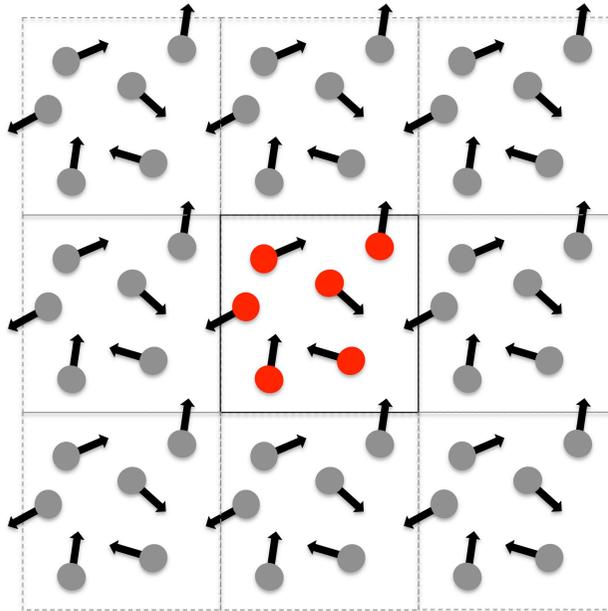
time step Figure 5.1. The potential energy is a function of the atomic positions ( $3N$ ) of all the atoms in the system. Due to the complicated nature of this function, there is no analytical solution; it must be solved numerically, following the algorithm called leap-frog/Verlet scheme [150, 151, 153]. That scheme is one of the most accurate, stable, and yet simple and efficient algorithm available for molecular dynamics [152].

Some MD packages, like GROMACS or NAMD, in order to reduce the cost of computing full electrostatics, uses a multiple time-stepping integration scheme. In this scheme, the total force acting on each atom is broken into two pieces, a quickly varying local component and a slower long-range component. The local force component is defined in terms of a splitting function. The long-range component consists only of electrostatic interactions outside of the local interaction distance. Since the long range forces are slowly varying, they are not evaluated every time-step [153].

Highest frequency motions occurring in the system limit the length of the time-step  $\Delta t$  in a MD. By freezing the high-frequency internal vibrations, such as bond-length or possibly bond-angle vibrations, the highest frequency motions threshold is increased, allowing longer time-steps. Several methods exist for integrating the equations of motion of a molecular system in presence of constraints, like the so-called SHAKE method [154], which is especially appropriate for macromolecules. When simulating a macromolecular system without constraints, a value of time-step of 0.5 fs is appropriate, whereas with a time-step of 1.0 fs is recommended when bonds to hydrogen atoms are constrained. When all bonds are constrained a time step of 2.0 fs is appropriate. This would allow a factor of two to four times in computing effort to be saved.

### 5.1.2 An important matter: the treatment of boundaries

We run computer simulation programs to predict and study the properties of a system in bulk. However, the size of the molecular system in a particular simulation is limited by the memory capacity and speed of the computer. It is important to consider that, no matter how large the simulated system, the number of atoms in the simulation would be negligible compared with the number of atoms contained in a macroscopic piece of matter. In order to carry out a simulation that adequately represent a real system, the nature of the boundaries of the model has to be carefully considered. This is generally done by imposing a periodicity on the simulated system. In this way, the biomolecule and the solvent are enclosed within a limited space (a periodic boundary box) with penetrable walls to make the number of particles constant throughout the simulation. We can imagine this box to be replicated to infinity by rigid translation in all three dimensions, completely filling the space. The content of one box is identical to that of all other boxes and any molecule leaving the box on one direction must simultaneously enter it from the opposite direction, so the concentration of material remains constant. Figure 5.2. In molecular simulations with periodic boundary conditions the computational box may have five different shapes: triclinic, hexagonal , two types of dodecahedrons, or a truncated octahedron [155]



**Figure 5.2 Periodic boundary conditions.** The figure below illustrates the concept of periodic boundary conditions in 2 dimensions. The central box represents the system we are simulating, while the surrounding boxes are exact copies in every detail - every particle in the simulation box has an exact duplicate in each of the surrounding cells. Even the velocities (indicated by the arrows) are the same. This arrangement is imagined to fill the whole of space. A result of this is that whenever an atom leaves the simulation cell, it is replaced by another with exactly the same velocity, entering from the opposite cell face. So the number of atoms in the cell is conserved. Furthermore, no atom feels any surface forces, as these are now completely removed

### 5.1.3 Constant temperature and pressure ensemble

Inside the box, where the simulation is carried on, the temperature and pressure can be controlled. Usually the simulations are carried out under isothermal-isobaric ensemble, where moles ( $n$ ), pressure ( $P$ ) and temperature ( $T$ ) are conserved. This corresponds most closely to laboratory conditions with a flask open to ambient temperature and pressure. Also, the use of a thermostat can be motivated by other reasons, such as the study of

temperature dependent processes or enhancing the efficiency of a conformational search.

For a finite size system one may define an instantaneous temperature  $T(t)$ , commonly defined in terms of kinetic energy through equipartition (Eq 5.4),

$$(Eq. 5.4) \quad T(t) = \frac{1}{k_B N_{df}} \sum_{i,\alpha} m v_{i,\alpha}^2$$

where  $N_{df}$  is the number of degrees of freedom in the system,  $k_B$  is Boltzmann's constant and  $v_{i,\alpha}$  is the velocity of the atom  $i$  in one of the  $\alpha$  coordinate axis ( $x,y,z$ ). Where the initial distribution of velocities is drawn from the Maxwell-Boltzmann distribution

$$(Eq. 5.5) \quad p(v_{i,\alpha}) = \left(\frac{m_i}{2\pi k_B T}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2} \frac{m_i v_{i,\alpha}^2}{K_B T}\right]$$

In order to control the temperature of the MD, several methods have been proposed [156-160]. An obvious way to alter the temperature is to scale the velocities by a factor lambda.

$$(Eq. 5.6) \quad \lambda(t) = \sqrt{1 + \frac{\Delta T}{\tau_T} \left(\frac{T_0}{T(t)} - 1\right)}$$

The factor  $\lambda$  is used to scale the velocities at each time-step, in order to relax the temperature toward the desired temperature value  $T_0$ . The relaxation rate is controlled by the time coupling constant  $\tau_t$ . This algorithm proposed by Berendsen [157] although it is efficient for relaxing the system to the reference temperature, it does not allow for a proper sampling of the canonical ensemble.

The Nosé–Hoover is a deterministic thermostat, first proposed by Nosé [156] and later modified by Hoover [158] and consider the heat bath as an

## CAPÍTULO 5

integral part of the system by addition of an artificial variable playing the role of the thermodynamic friction coefficient evolving in time according to the difference between  $T(t)$  and  $T_0$  [161, 162].

$$(Eq. 5.7) \quad \frac{d^2 r_i(t)}{dt^2} = \frac{F_i}{m_i} - \xi \frac{dr_i}{dt}$$

$$(Eq. 5.8) \quad \frac{d\xi}{dt} = \frac{\sum m \frac{dr_i}{dt} - N_d f k_B T}{Q} = \frac{T(t) - T_0}{Q}$$

The strength of the coupling is determined by a constant term  $Q$  and  $E$  is the friction factor.

A possible alternative to the deterministic thermostat is the use of stochastic Langevin thermostat [151]. This thermostat models the influence of a heat bath by adding to the velocity of each particle a small random noise (fluctuation) and a frictional force directly proportional to the velocity of that particle. The random force term  $\eta_i(t)$  is randomly extracted from a Gaussian distribution. These two factors are balanced to give a constant temperature. The drawback of this thermostat is that momentum transfer is destroyed. So, it is unadvisable to use Langevin thermostat for runs in which you wish to compute diffusion coefficients

$$(Eq. 5.9) \quad \frac{d^2 r_i(t)}{dt^2} = \frac{F_i}{m_i} - \xi \frac{dr_i}{dt} + \eta_i(t)$$

In constant pressure dynamics, the volume of the unit cell is adjusted and rescales all atomic coordinates each step to make the computed pressure approach the target pressure. The pressure is given by the Equation 10 where  $\Theta$  is the interval virial for pair-additive potentials,  $F_{ij}$  is the force on particle  $i$  due to particle  $j$  and  $r_{ij}$  is the distance between particles ( $r_i - r_j$ ).

$$(Eq. 5.10) \quad P = \frac{2}{3V} (\sum_{i,\alpha} m v_{i,\alpha}^2 - \Theta)$$

$$(Eq. 5.11) \quad \Theta = -\frac{1}{2} \sum_{i < j} r_{ij} F_{ij}$$

The various methods for carrying out MD at constant pressure are based on the same principles as the methods of temperature control, by the modification of the equations of motions of each atom by the introduction of an additional degree of freedom such as Langevin piston scheme [163-165] or by embedding in the temperature control scheme the degrees of freedom of unit cell vectors like Nosè-Hoover barostat .

#### 5.1.4 The molecular mechanics force field in biological macromolecules

In the context of MD of biological macromolecules, a force field refers to the form and parameters of mathematical functions used to describe the potential energy of a system of atoms. The core of any force field is the potential energy function used to describe the relationship of the coordinates of an atom to the potential energy,  $V$ , of the system of interest. The following equation is a typical effective potential for a biomolecule of  $N$  atoms and Cartesian position vectors  $r_i$ .

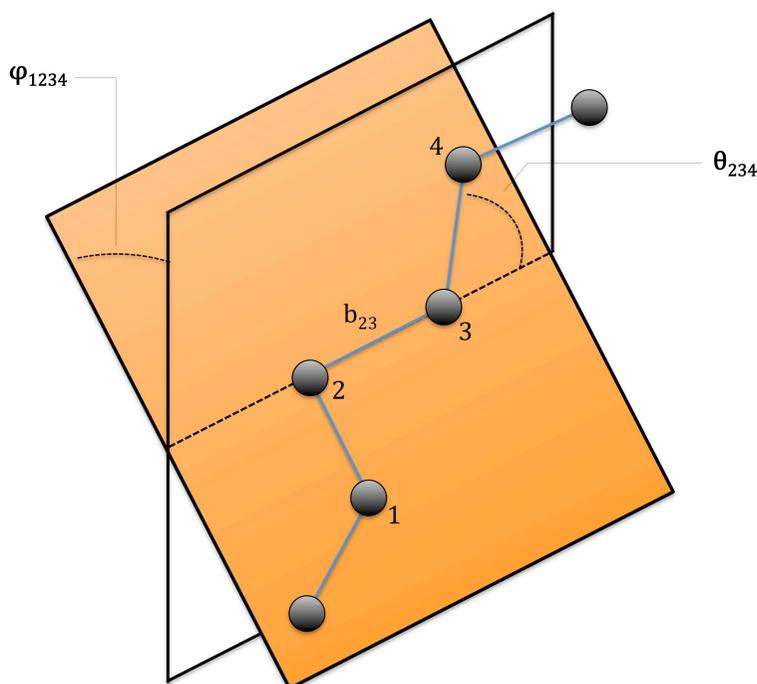
(Eq. 5.12)

$$\begin{aligned} V(r_1, \dots, r_N) = & \sum_{\text{Bonds}} \frac{1}{2} K_b (b - b_0)^2 + \sum_{\text{angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2 \\ & + \sum_{\text{dihedrals}} K_x (1 + \cos(n\chi - \delta)) + \sum_{\text{impropers}} \frac{1}{2} K_{imp} (\varphi - \varphi_0)^2 \\ & + \sum_{\text{nonbond}} \varepsilon_{ij} \left( \left[ \left( \frac{Rmin_{ij}}{r_{ij}} \right)^{12} - \left( \frac{Rmin_{ij}}{r_{ij}} \right)^6 \right] \right) + \frac{q_i q_j}{\varepsilon r_{ij}} \end{aligned}$$

The first term represents the covalent bond stretching interaction along bond length  $b$ . It is a harmonic potential in which the minimum energy bond length  $b_0$  and the force constant  $K_b$  vary with the particular type of bond.

## CAPÍTULO 5

The second term describes the bond angle bending interaction in similar form. Two forms are used for the dihedral angle interactions: a harmonic term for dihedral angles  $\varphi$  that are not allowed to make transitions, and a sinusoidal term for the other dihedral angles  $\chi$ , which may make 360 degree turns. (Figure 5.3)



**Figure 5.3 Force field bounded conditions term.** Geometry of a simple chain molecule, illustrating the definition of interatomic distance  $r_{23}$ , bond angle  $\theta_{234}$  and torsion angle  $\varphi_{1234}$ .

The last term is a sum over all pairs of atoms and represents the effective non-bonded interaction, composed of the Lennard-Jones potential, used to treat the van der Waals (vdW) interactions and the Coulomb interactions between atoms  $i$  and  $j$  with charges  $q_i$  and  $q_j$  at distance  $r_{ij}$ , used to treat the electrostatic interactions. The choice of relative dielectric constant is a matter of dispute. Values ranging from 1 to 8 have been used [166], while

others take  $\epsilon$  proportional to the distance  $r_{ij}$  [167, 168]. The Lennard-Jones potential (also referred to as the LJ potential, 6-12 potential, or 12-6 potential) is a mathematically simple model that approximates the interaction between a pair of neutral atoms, where  $\epsilon_{ij}$  is the LJ well-depth,  $R_{min}$  is the minimum interaction radius and  $r_{ij}$  is the distance between the particles. Typically,  $\epsilon_{ij}$  and  $R_{min,ij}$  are obtained for individual atom types and combined for each pairwise [169]. The computation of the pair-wise non-bonded interactions are the most time-consuming part of a MD simulation; because of the non-local and long range nature of the charge-charge interactions. Accordingly, several approximations have been used during the last 20 years in order to reduce the computational effort needed [170]. One of the simplest approximations is the use of a cutoff distance that defines the maximum distance allowed between pairs of atoms for their interaction to be computed, under the assumption that the force between atoms can be neglected when they are largely separated. Since the Lennard-Jones forces are short-range in nature, their contribution seems to be properly modelled with such cutoffs, although more accurate strategies have recently been reported [171, 172]. However, dealing with the Coulombic term, the use of a cutoff is much more critical for the reliability of the simulations [173, 174]. Accordingly, the treatment of long-range interactions has been an active field of research in the last years and is still a topic of considerable interest [175, 176]. The most alternative to truncation most widely used is the Ewald summation procedure or the computationally more effective Particle Mesh Ewald or PME [177]. The Ewald summation is a technique for efficiently summing the interactions between ions and all its periodic images. The infinite sum of charge-charge interactions for a system in periodic boundary conditions is conditionally convergent, meaning that the result of the summation depends on the order in which it is taken. Ewald summation specifies the order as follows: sum over each box first, and then

## CAPÍTULO 5

sum over spheres of boxes of increasingly larger radii [178]. In PME, the charges are distributed over a grid points with weighting functions chosen according to the distance of the respective grid points to the location of the charge and the convolutions necessary to evaluate the sums are calculated via fast Fourier transforms. This method significantly reduces the computation time where the long-range electrostatic interactions are calculated in the more efficient way using Fourier spaces.

The form of the Eq 5.6 is similar to that applied in early force fields and is still the form used in biomolecular force fields. Nonetheless, a potential energy function alone does not make a force field. It is a combination of the potential energy function with the parameters used in that function. The parameters for the bonded and non-bonded terms of the force field are derived from quantum calculations and from thermodynamic, crystallographic and spectroscopic data on a wide range of the systems. These parameters have been developed to simulate a specific type of molecules. The force field defines a set of parameters for each type of atom. For example, a force field would include distinct parameters for an oxygen atom in a carbonyl functional group or in a hydroxyl group. The typical parameter set includes values for atomic mass, van der Waals radius, partial charge for individual atoms, equilibrium values of bond lengths, bond angles, dihedral angles for pairs, triplets and quadruplets of bonded atoms, and values corresponding to the effective spring constant for each potential.

The most commonly used force fields to perform MD simulations of biomolecules are, in alphabetical order, AMBER [179], CHARMM [180], GROMOS96 [181] and Optimized Potentials for Liquid Simulations (OPLS) [182]. The differences between these biomolecular force fields are subtle, regarding details such as how improper (out of plane) torsions are described

or how differences in the local charge distributions are dealt with. Nonetheless, they all reproduce experimental data in a similar fashion [183].

### 5.1.5 The solvent models for biomolecules

The aqueous environment is an essential aspect for three-dimensional structure and activity of the biomolecules [184, 185]. Many water models are available in the literature for an accurate representation of the liquid water [186, 187]. These models have been parameterized to reproduce physical and thermodynamic properties such as the density, enthalpy of vaporization, radial distribution functions, energies of hydration or dipole moment. Such treatment may be performed using explicit or implicit models.

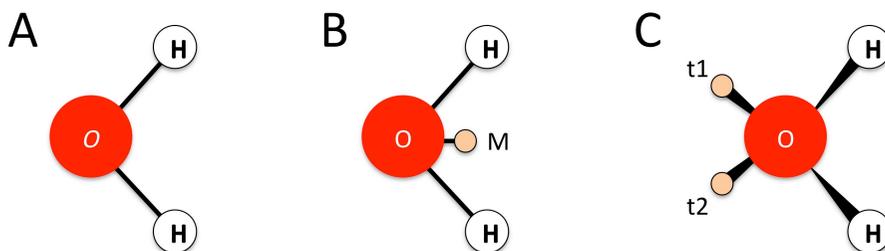
Implicit solvation models allow for a significant reduction in computational time by assuming the solvent as a continuous medium. Early implicit solvation models include simple treatments such as distance-dependent dielectric constants ( $r$ -dielectric) and atomic solvent accessibility based free energy contributions to solvation [188]. A significant improvement in the treatment of solvation is the use of the Poisson-Boltzmann (PB) model, where contributions from solvent polarization along with the asymmetric shapes of biological molecules are taken into account [189]. Although the PB equation has solid theoretical justification, it is computationally expensive to apply without additional approximations. Effective alternatives are the generalized-Born (GB) based solvation models. These models are based on modelling the protein as a set of spheres whose internal dielectric constant differs from the external solvent [190]. In addition, both the PB or GB methods can be combined with free energy solvent accessibility (SA) terms that account for the hydrophobic effect, referred to as PB/SA or GB/SA approaches [191, 192]. Implicit solvation models have made significant advances in recent years

## CAPÍTULO 5

[193], and are typically used in studies where an extensive sampling of conformational space is required, such as in protein folding. However, these models can fail when highly specific water-biomolecule interactions are important.

The explicit water models used in biomolecular simulations are TIP3P, TIP4P [194], TIP5P [195], the empirical simple point-charge (SPC) model (SPC) [196] and extended SPC/E models [197].

In all of these models the movements of individual water molecules are explicitly calculated. They can be classified attending to the number of points used to define the model, the consideration of rigid or flexible structures, the inclusion in the model of polarization effects. The simplest and probably the most commonly used model for MD is TIP3P, which has three interaction sites, corresponding to the three atoms of the water molecule with rigid geometry (Figure 5.4). A point charge is assigned to each atom and, also, Lennard-Jones parameters are considered for the oxygen atom. The more complex 4-sites or 5-sites models, such as TIP4P and TIP5P respectively, place the negative charge either on a dummy atom (point M) placed at a distance of 0.15 Å in the oxygen along the bisector of the HOH angle, or on two dummy atoms representing the lone pairs of the oxygen atom (Figure 5.4 Panel B). The TIP5P is a 5-point model with four charges arranged tetrahedrally around the oxygen atom (Figure 5.4 Panel C). The positive charges are located at the hydrogen atoms at a distance of 1 Å from the oxygen atom, close to the real distance in the water molecule. The negative charge is located at the other two vertices of the tetrahedron (sites t1 and t2) but at a distance of only 0.8 Å from the oxygen. These models improve the electrostatic distribution around the water molecule, but result in significantly higher computational costs because of the higher number of electrostatic interactions to compute.



**Figure 5.4 Schema of most common rigid Water Models.** In the Panel A; The simplest model used for TIP3P, SPC and SPC/E where all the atoms are in the same plane and where the partial positive charges on the hydrogen atoms are balanced by an appropriately negative charge located on the oxygen atom. The TIP3P and SPC models differ slightly in parameters of distance between atoms and the angle HOH. In the Panel B; the four-site models, TIP4P, in this model the negative charge shift from the oxygen atom to a point along the bisector of the HOH angle towards the hydrogen atoms. The parameters for these two models are. In the Panel C; five-site model, TIP5P is almost tetrahedral, charges are placed on the hydrogen atoms and on the two lone-pair sites close to the oxygen atom ( $t1$  and  $t2$ ).

The SPC models are similar to the TIP3P models, albeit considering an ideal tetrahedral geometry (i.e., H-O-H angle =  $109.47^\circ$ ) instead of the observed angle. The SPC/E model includes a correction for polarization self-energy that yields improved structure and diffusion properties [183].

Each water model has its own strengths, but all of them are far to be a perfect prediction of the behaviour of the bulk water [198]. In the work of this thesis we used TIP3P model for the entire simulations carried out in this work. This model reproduce close enough the behaviour of the water molecules for the purpose our analysis and also not increase the computational cost of our simulations in a magnitude not affordable.

## CAPÍTULO 5

**Table 5.1** Time scale of molecular processes in proteins.

Type of motion	Time	Amplitude
Atomic fluctuation, Side chain motion	fs To ps ( $10^{-15}$ – $10^{-12}$ s)	> 1 Å
Loop motion, Collective motion	ps To $\mu$ s ( $10^{-12}$ – $10^{-6}$ s)	1-5 Å
Domain motion, Subunit motion, Hinge- bending motion	$\mu$ s To ms ( $10^{-6}$ – $10^{-3}$ s)	5-10 Å
Subunit association, Folding	ms To s ( $10^{-3}$ – $10^0$ s)	< 10 Å

### 5.1.6 An introduction to metadynamics

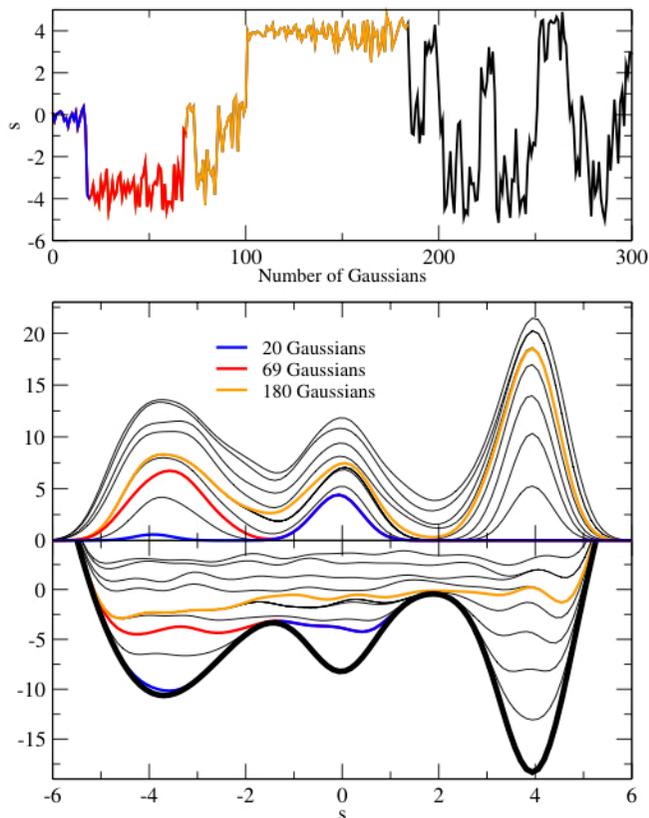
It has been 25 years since the first MD of a biomolecule was performed. The bovine pancreatic trypsin inhibitor (BPTI) was simulated in vacuum with a crude force field and the trajectory was calculated just for 9.2 ps long using a relatively accurate X-ray structure available [150]. Now, with a huge number of structures available in the protein data bank (PDB), the refined force fields and the speed of computers, MD simulations are progressively pushed to larger systems, to perform combined quantum/molecular mechanics, and longer simulation times. Such improvement has made possible to apply MD simulations to a wide range of problems related with biomolecules, from the characterizing the structural dynamics of proteins to elucidating mechanisms of protein function. For example, the study of dynamics of protein-ligand

systems makes possible the evaluation of important structural features of a ligand/protein interface, such as intermolecular hydrogen bonds, contact area and binding energy [199].

Also, MD simulations are a powerful approach to further our understanding of the microscopic factors responsible for the conformational stability of biomolecules, by means of Gibbs free energy calculations using an analysis based on thermodynamic cycles and statistical mechanics. The first approaches to get the free energy map (also known as a potential of mean force or PMF) of a protein were achieved by free energy perturbation (FEP) and thermodynamic integration (TI) methods [200, 201]. Sampling problems may be encountered in FEP or TI when there are large energy barriers, which do not allow an effective exploration of the conformational space within the available computer power. The design of a molecular dynamics simulation should account for the available computational power. Simulation size (number of particles), time-step and total time duration must be selected so that the calculation can finish within a reasonable time. With the computers available now a day allow, classical molecular dynamics simulations can be performed with time scales in the range of hundreds of nanoseconds. Nonetheless, for large and complex systems, such as proteins or other biomolecules, most phenomena of interest take place on time scales that are orders of magnitude larger. To make statistically valid conclusions from the simulations, the time span simulated should match the kinetics of the natural process Table 5.1. Also, there are many circumstances where the free energy surface (FES) has several local minima separated by large barriers. In such circumstances a simulation started in one minimum will be able to move spontaneously to the next minimum only under very favourable circumstances. To overcome this problem and enhance the conformational

## CAPÍTULO 5

space sampling, different techniques have been developed. For example, umbrella sampling [202], adaptive force bias [203], metadynamics [204], coarse-graining [205], finite-temperature string method [206], milestoning [207], hyperdynamics [208], etc [208-210].



**Figure 5.5** Filling a 1D model potential with hills starting in the lower well. The dynamics is biased with a metadynamics potential  $V_G$  as defined by equation 5.13. The parameters are  $s = 0.4$ ,  $w = 0.3$  and  $\tau = 300$ . Upper panel: time evolution of the metadynamics bias potential  $V_s$ . Blue line:  $V_s$  as when the first minimum is filled and the system ‘escapes’ to the second minimum; red line:  $V_s$  as when also the second minimum is filled; orange line:  $V_G$  when the entire profile is filled and the dynamics becomes diffusive. Lower panel: time evolution of the sum of the metadynamics potential  $V_G$  and of the external potential, represented as a thick black line [211-213]

Metadynamics is one of the methods aimed to reconstructing the probability distribution enhancing the sampling as a function of one or a few predefined collective variables (CVs), which are assumed to be able to describe the process of interest. Ideally the CVs should clearly distinguish

between the initial state, the final state and the intermediates, and they should describe all the slow events that are relevant to the process of interest. If the CVs are not chosen carefully, the metadynamics can be stuck on overfilling one minimum, or it will take a very long time to fill the free energy surface.

The metadynamics in the space of the chosen CVs ( $S$ ) is enhanced by a history-dependent potential constructed as a sum of gaussians centred along the trajectory, added during the simulation every certain number of time-steps, with a controlled width  $\delta s$  and height  $\omega$ . The external ('metadynamics') potential acting on the system at time  $t$ , if two or more CVs are used at the same time, is given by

$$\text{(Eq. 5.13)} \quad V_G(S(x), t) = \omega \sum_{\substack{t' = \tau_G, 2\tau_G, \dots \\ t' < t}} \exp\left(-\sum_{\alpha=1}^d \frac{(S_\alpha(x) - S_\alpha(t'))^2}{2\delta s_\alpha^2}\right)$$

At the beginning of the simulation  $V_G$  is set to zero, and is updated with a certain frequency  $\tau_G$ . The basic assumption of metadynamics is that  $V_G(S(x), t)$  defined in the latter equation after sufficient long time provides an estimate of the underlying Gibbs free energy Figure 5.5. After that time, the motion of the system becomes diffusive and unbound in the region of CV space. The metadynamics potential  $V_G$  is represented at different times in Figure 5.5, upper panel. The lower panel of Figure 5.5 reports the sum of  $V_G$  and the external potential. Clearly, as the simulation proceeds,  $V_G$  iteratively compensates the underlying potential.

This biasing avoids entrapment in a local minimum and forces the system to sampling the conformational space along the collective variables and also maps out the underlying free energy surface as the negative of the sum of the bias potential hills  $V_G(S, t)$ . This scheme is common in other methodologies [203]. However metadynamics has the advantage that the bias

## CAPÍTULO 5

potential recursively compensates the Gibbs free energy allowing first the automatic exploration of low Gibbs free energy regions [214].

A strategy for ensuring a correct description of the relevant physics of the biomolecule under study is to increase the number of CVs. Unfortunately this number cannot grow beyond a threshold of three because to explore a high dimensional Gibbs free energy surface is too costly. Some strategies have been recently proposed [207, 215-217]. In the BE approach the replicas of MD simulations (walkers) are run in parallel, biasing each walker in a different collective variable. Each different replica independently accumulates a history-dependent potential  $V_g(S_k(x_i), t)$  that, after a sufficiently long time, would provide an estimate of the free energy projected on  $S_k$ . If all the variables are relevant for the process, the convergence of each free energy profile can be extremely slow and hindered by hysteresis. We now allow the replicas to exchange their configurations, two replicas  $i$  and  $j$  are selected at random among the available walkers. The exchange move consists of swapping the atomic coordinates  $x_i$  and  $x_j$  of the two replicas evolved under the action of two different history-dependent potentials, for example  $V_G(S_k(x_i), t)$  and  $V_G(S_l(x_j), t)$ , when an appropriate probability function (Metropolis criterion) is satisfied [147]. The swap is accepted with a probability

(Eq. 5.14)

$$P = \min \left\{ 1, \exp \left[ \frac{1}{T} (V_G(S_k(x_i), t) + V_G(S_l(x_j), t) - V_G(S_k(x_j), t) - V_G(S_l(x_i), t)) \right] \right\}$$

The Metropolis scheme is implemented by attempting a trial displacement and calculating the change in history-dependent potentials. If  $\Delta V_G < 0$ , then  $P > 1$  and the move is accepted. Otherwise, if  $\Delta V_G > 0$ ,  $P < 1$  and

the move is accepted with a probability of  $P = \exp[-1/T(\Delta E)]$ . A random number  $R$ , between 1 and 0, is generated and the move is accepted if  $\exp[-1/T(\Delta V_G)] \geq R$ .

The aim of the bias exchange is to introduce a jump process on top of the ordinary molecular dynamics evolution. The jumps greatly increase the capability of each replica to diffuse in the CV space, and hence the accuracy of the free energy reconstruction, which is primarily determined by the correlation time of the dynamics in CV space.<sup>18</sup> Moreover, by using this algorithm each configuration evolves under the action of a history-dependent potential that changes every time an exchange move is accepted. Even if each bias is defined in a few collective variables at a time each configuration, after some accepted exchanges, will eventually explore the space spanned by all the collective variables. This improves greatly the capability of the system to explore the configuration space.

In low dimensional metadynamics, the best estimation of the Gibbs free energy is

$$(Eq. 5.15) \quad F_{meta}(s, t_{tot}) = - \frac{1}{t_{tot} - t_F} \int_{t_F}^{t_{tot}} V_G(s, t) dt$$

if  $t_F$  is the time which the CV starts diffusing in all the relevant region and  $t_{tot}$  is the total simulation time. In bias exchange the results of each simulation are low dimensional projections of the Gibbs free energy for each walker. The convergence of the bias potential to the corresponding Gibbs free energy projection is monitored like in standard metadynamics. Nevertheless, these low dimensional projections are often not very intuitive, as in complicated processes like protein conformational transitions each minimum in a low dimensional profile may correspond to different structures. A manner to estimate the relative probability of the different structures in a high

## CAPÍTULO 5

dimensional space is to use the low dimensional Gibbs free energies obtained from bias exchange to estimate, by weighted-histogram procedure (WHAM) [218], the free energy of a finite number of structures that are representative of all the configurations explored by the system. These structures are determined by performing a cluster analysis, grouping all the frames of all the walkers in sets (bins) in which all the frames are close to each other in CV space. The canonical weight of each bin is estimated by a weighted histogram procedure based on metadynamics bias potentials. The method is full explained in the following reference [219].

### 5.2 Molecular dynamics simulations carried out with AMBER

MD simulations from Chapter 2<sup>th</sup> and 3<sup>th</sup> were carried out with the AMBER 9.0 package using as initial configurations the structures from the PDB file as are detailed in each chapter. All crystallographic waters were removed, except those buried at the binding interface for the simulations performed at Chapter 3. Standard ff03 force field parameters were used [220, 221]. Each complex was solvated in a truncated octahedron periodic box filled with TIP3P water molecules and neutralized the net charge by counter ions. All hydrogen atoms, water molecules and ions were added using the fleap tool. MD simulations were preceded by two energy-minimization steps with the aim of remove any bad contacts created during the solvation and also to allow our starting structure to reach a stable local minimum. In the first stage the protein and crystallographic water molecules were fixed by a harmonic potential in order to optimize the positions of the added water and counter ions atoms. Then in a second stage the entire system will be minimized. To achieve this we performed 1000 cycles (500 steepest descent and 500 conjugate gradients) with 500 kcal·mol<sup>-1</sup> harmonic force restraints on ligand

and protein atoms followed by 2500 cycles (1500 steepest descent and 1000 conjugate gradients) without constraints. Once the system was minimized, the system was slowly heated. Doing the heating in stages reduces the chances that the system will blow up by allowing it to equilibrate at each temperature. An alternative option would be to use weight restraints to control the heating (as we did in the NAMD protocol in the next section). The system was heated from 0 to 300 K for 20 ps and equilibrated for 30 ps. To maintain the temperature a Langevin temperature equilibration scheme was used. Finally the production simulation at 300 K was run. MD runs were carried out in periodic boundary conditions in an isothermal isobaric ensemble (NPT) at 1 atm, with 2-fs time integration steps. The SHAKE algorithm was used to constrain all bonds that contain hydrogen atoms. A 12-Å cutoff was applied to treat non-bonded interactions, and the particle Mesh Ewald method was introduced for the treatment of long-range electrostatic interactions. MD trajectories were recorded every 2 ps. The simulations were run in the cluster at the University of Granada, UGRGRID, (a Sun Fire X2200/X4600 cluster) using 128 CPUs for each trajectory.

### 5.3 Molecular dynamics simulations carried out with NAMD

NAMD is a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems that works with AMBER and CHARMM force fields and file formats. In order to produce long simulations for the different structures (Chapter 4), NAMD version 2.7b1 [222] was chosen using the CHARMM22 force field.

The starting structures were solvated and neutralized with the CHARMM scientific package [167]. Solvating a protein requires a few simple

## CAPÍTULO 5

considerations and operations. The protein has to be oriented about the origin and then its spatial dimensions have to be computed. From this, one can estimate the size of the water box required. One should allow for at least an additional 10 Å in each direction. Based on such considerations, for the structures of the Chapter 4 a 75x75x75 Å truncated octahedron box of water was employed. Once the solvation was completed, the structure was neutralized replacing a number of water molecules with counter-ions in order to achieve the desired concentration of ions (0.15 M solution of potassium chloride). Next, the system was minimized with a harmonic restraint at the protein atoms and crystallographic water molecules for 2000 steps, in order to remove bad contacts of water molecules. Then, the full system was equilibrated by carrying out conventional MD simulations with gradually weaker constraints every 1 ns on the structure, during 4 ns. All trajectories were generated at constant temperature (298 K) and pressure (1 atm), under periodic-boundary conditions. Electrostatic interactions were computed using the Particle Mesh-Ewald algorithm with a real-space cutoff distance at 12 Å, also used for the van der Waals interactions. To control the temperature and the pressure the Langevin scheme was used and MD trajectories were recorded every 2 ps. The simulations were run into the University of Granada, UGRGRID (a Sun Fire X2200/X4600 cluster) using 64 CPUs for each trajectory.

### 5.4 Modelling of the 3D structures

To perform molecular dynamics simulations is necessary to have the files of coordinates of the structures of the mutants of interest in pdb format (Protein Data Bank). Because there are no experimental structures for some of the interest mutants, it is necessary to model them from the existing structures.

In order to obtain the structures of the mutants of the Chapter 2 , we used as a template the structure of the SH3 domain of c-SRC bound to VSL12 (code pdb: 1QWF). To produce the coordinate files of the different mutants the side chain of residue 128 fro the NMR structure 1QWF was modified using the design tools contained in the Discovery Studio 2.1 Suite [223], we modified just the side chain of the target residue for the desire mutant. To obtain the structure of the domain without ligand we selective remove the chain corresponding to the VSL12 ligand.

The models for the Chapter 4 have been generated from the pdb file with code 2F00 and 2ABL. In the Table 4.8 we show with residues were removed to build each model. In the other hand to build the mutants we used the charmm tools to substitute the residue of interest, after a minimization step were applied to the structures.

## 5.5 Analysis of molecular dynamics trajectories

The data obtained from MD trajectories can be used to calculate several structural and dynamical properties. The most common analysis is the Root Mean Square Displacement (RMSD) use to estimate the displacement of the atoms or group of atoms along the trajectory. The RMSD of a set of atoms at time  $t$ , with respect to a reference conformation, for example the initial conformation, reads:

$$(Eq. 5.16) \quad RMSD(t) = \sqrt{\frac{\sum_{i=0}^N |\vec{r}_i(t) - \vec{r}_i^0|^2}{N}}$$

## CAPÍTULO 5

Where  $|\vec{r}_i(t) - \vec{r}_i^0|$  is the displacement of the  $i^{\text{th}}$  atom at time  $t$  from the reference position  $\vec{r}_i^0$ . An increase of the RMSD indicates that the protein moves to a conformation different from initial structure.

Similar to the previous analysis is the calculation of the fluctuation of an atom or a group of atoms, the Root Mean Square Fluctuations (RMSF). This is calculated according:

$$\text{(Eq. 5.17)} \quad \text{RMSF}_i = \sqrt{\langle (\vec{r}_i - \langle \vec{r}_i \rangle)^2 \rangle}$$

where  $\vec{r}_i$  is the position vector of the  $i^{\text{th}}$  atom and the  $\langle \vec{r}_i \rangle$  is the average position of this atom along the trajectory. This analysis gives clues about the most mobile regions of the protein.

To evaluate if a water molecule is establishing a hydrogen bond with the key residues of each hydration spot we used the external program HBplus [116], setting the criteria to a maximum cutoff distance of 3.5 Å between donor and acceptor atoms and a minimum angle of 90°.

### 5.6 Metadynamics for cAbl tandem.

The metadynamics method was applied to the structures SH32, SH32-3GLY, NCAP-SH32 and SH32-C2Q as is detailed in the Chapter 4 (Table 4.8) to analyse the role of Ncap, linker and connector regions in the conformational equilibrium of the modular domains SH3-Sh2 tandem of cAbl. All metadynamics simulations were performed using the CHARMM27 force field for protein description and GROMACS simulation package [224] compiled with PLUMED plugin [225].

The initial structures used for each metadynamics were the product after the equilibration step of the NAMD protocol in the section 5.3. We added

an extra equilibration step of 4 ns in order to relax the structure and to avoid problems from the change of MD package. The virtual hydrogen scheme implemented in GROMACS was used allowing the use of a large integration time-step, 4 fs. The equilibration step and the metadynamics were carried out at constant temperature of 298 K and 1 atm of pressure. To control the temperature and the pressure the stochastic velocity-rescaling methodology was used.

For each metadynamics simulation, Bias Exchange methodology was performed using 32 replicas, two collective variables (S4 and S5) and three path collective variables (S1, S2 and S3) were chosen. Each walker or replica was bias in two different CVs; 16 replicas simultaneously bias with S1 and S2 CVs, 11 replicas bias with S1 and S3 CVs, 4 replicas bias with S4 and S5 CVs and one unbiased.

### 5.6.1 Collective Variables.

The preliminary identification of a set of CVs to adequately describe the conformational equilibrium was performed through the analysis of standard MD simulations previously carried out and described in the section 5.3. From this analysis we obtain two main descriptors, the torsional dihedral angle between the centres of mass of the backbone atoms of the residues PRO137, LEU141, TRP146 and PRO150 (D1) and the TYR134, PRO137, LEU141 and TRP146 (D2) Figure 4.10. Both dihedrals are located in the linker sequence connecting the SH2 and SH3 domain in the kinase, and are a good descriptor of the relative positions of the domains as is shown in the Figure 4.11.

Without knowing *a priori* which are the most important degrees of freedom in the conformational equilibrium of the cAbl tandem, it is advisable to use more CVs to achieve the best results in the shortest time. The path

## CAPÍTULO 5

collective variables (PCV) allow the exploration of hidden degrees of freedom, leading to the prediction of unexpected intermediates and metastable states but reduce the space to explore.

The PCVs are extremely powerful whenever one wants to study a transition between two states A and B [211]. As state A we chose the X-ray conformation of cAbl tandem in the open state (PDB ID code 2ABL), while state B represents the close form (PDB ID code 2FO0). Now, we would expect the reactive trajectories to be bundled in a narrow tube around the path between the state A and B. To trace this path, we follow the procedure of Branduardi et al. introducing the two variables  $S_{PATH}$  and  $Z_{PATH}$ :

Defined as  $S_{PATH}$  then measures the position along the path, and is calculated as

$$(Eq. 18) \quad S_{PATH} = \frac{\sum_{i=1}^N i e^{-\lambda d(X_i, X(t))}}{\sum_{i=1}^N e^{-\lambda d(X_i, X(t))}}$$

In this scheme one defines a path as a set of  $N$  reference conformations that define the path in configuration space  $X$  from some initial state to some final state. Where  $X(t)$  is the configuration of the system at any given time,  $d(X_i, X(t))$  is the metric used to calculate distances on the configuration space (the root mean square deviation (RMSD) in our case) and  $\lambda$  is a normalization prefactor and it is proportional to the inverse of the mean square displacement between successive frames.

The  $Z_{PATH}$  measure the orthogonal distance of a given configuration  $X(t)$  to the reference path. Defined with  $Z_{PATH}$  is calculated as

$$(Eq. 5.19) \quad Z_{PATH} = -\lambda^{-1} \log \sum_{i=1}^N e^{-\lambda d(X_i, X(t))}$$

We described the transition between the closed and open conformation of the tandem of cAbl only using two frames, being the closed (2FO0) and the

open conformation (2ABL). The PCVs S1 and S3, were the SPATH for the initial and final states, more specifically, S1 describes how far is the whole structure from the crystallographic references and S3 is centred only in the C $\alpha$  carbons of the connector region, from THR136 to SER145. The  $\lambda$  value was set to 4 and 30, respectably. The PCV S2, by contrast, was set up as the Z<sub>PATH</sub> for the whole structure of the tandem and  $\lambda$  value was set to 4.

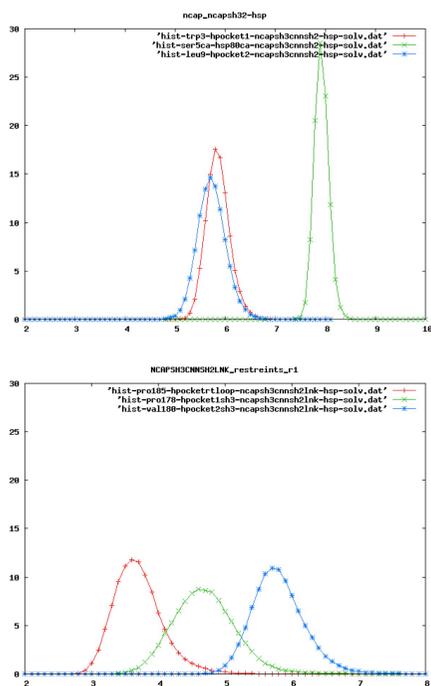
To limit the sampling between the conformations of interest, a repulsive quadratic potential was applied on the path collective variables (S1, S2, S3) in the values at the corresponding reference structures. This was done to keep the sampling between conformations between the ensemble and disassembled structures.

All metadynamics included a series of preparatory phases, during the first 100 ns of simulation time, in which the height of the biasing-Gaussians was progressively reduced from 0.5 kJ·mol<sup>-1</sup> to 0.1 kJ·mol<sup>-1</sup>. This phase is called as filling time. The Biasing-Gaussians are added every 500 steps. The widths of the biasing-Gaussians applied to CVs S1, S2, S3, were 0.5, 0.7 and 0.15 Å, respectively, and the widths of the biasing-Gaussians applied to CVs S4 and S5 were 0.3 radians.

### 5.6.2 Restraints

For the metadynamics where the Ncap or linker (C2Q) regions were present, restraints were applied in order to keep these regions bound to the tandem during all the CVs exploration. We applied a distance restraint with a quadratic potential between the centres of mass of key residues. In order to apply a coherent threshold, first we analysed the MD of section 4.2. Figure 5.6

## CAPÍTULO 5



**Figure 5.6** Distances distribution of the centers of mass between residues during the molecular dynamics simulations. Upper panel: distances histogram for the residues of the Ncap. The center of mass of de sidechains of Lower panel: distances for the linker restrain residues

For SH32-C2Q metadynamic, we chose three of the hydrophobic main interactions, between the linker region and the SH3 domain. The first restraint was applied between PRO149 and TRP118 TRP129, the second between the residues PRO242 and TYR89, TYR134 and finally between VAL244 and PHE91, PRO131. The threshold was 5 Å, 6 Å and 7 Å respectably.

For NCAPSH32, the interactions chosen to keep the Ncap region attached to the connector region were also three. First restraint was between the C $\alpha$  of the phosphoserine 69 and C $\alpha$  of the HIS144, before between the centres of mass of the atoms of the sidechain of the residue TRP67 and residues TRP146, PHE168 and VAL218. Finally, a restraint was applied between the centres of mass of the atoms of the sidechain of the residue

LEU73 and residues VAL218 and TRP146. For those interactions, the threshold was set up to 8.5 Å for the first one, and 7 Å for the others.

### 5.6.3 Analysis of the metadynamics: estimating the Gibbs free energy

A clustering analysis in the space of the CVs 6 and 7 was carried out for each of the metadynamics simulation in order to get the Gibbs free energy surface using the method described in the latter section. Each CV was clustered in bins of 0.3 Å. The biased population of each of these clusters was then reweighted using WHAM method [226, 227], to derive their characteristic free energy.

Finally the CVs S6 and S7 are the root mean square deviation (RMSD) value of the C $\alpha$  of the residues SER78 to PRO237 taking as a reference the X-ray structure of 2ABL and the X-ray structure of 2FO0 respectability. Those collective variables were not biased during the metadynamics as they are used only for analysis purposes.

CAPÍTULO 6

---

**6 RESUMEN Y CONCLUSIONES**

CAPÍTULO 6

## 6.1 RESUMEN

Las proteínas están involucradas en el funcionamiento de la mayoría de los procesos celulares. Dentro del contexto funcional, las proteínas establecen una compleja red de interacciones que es crucial para el correcto funcionamiento celular. Cualquier distorsión de esta red de interacciones puede traducirse en serias alteraciones de las funciones celulares que originan el desarrollo de muchas enfermedades humanas. Entre otras, uno de los principales tipos de proteínas implicadas en los procesos de transducción de señales son las quinasas. Los cuales están encargadas de la fosforilación específica de residuos para la modificación de la actividad de proteínas diana. Como es evidente, en este proceso, juega un papel importante el reconocimiento específico de secuencias, llevado a cabo principalmente por los dominios modulares. Dentro de las familias de dominios modulares destaca los dominios de reconocimiento de secuencias ricas en prolina (MRP). Esta familia de dominios se encuentra compuesta por numerosos grupos como son los dominios SH3, WW, GYF, etc. Los dominios modulares están implicados en el desarrollo de numerosas patologías humanas convirtiendo estos dominios en atractivas dianas para el diseño de inhibidores específicos con el objetivo de desarrollar nuevos fármacos. Para ello es esencial comprender en detalle el mecanismo de reconocimiento de los dominios MRP, pero a pesar de los numerosos esfuerzos por comprender como su funcionamiento, a día de hoy, aun existen numerosas incógnitas, algo inesperado si tenemos en cuenta el pequeño tamaño y simplicidad del sitio de unión. Por ejemplo, la huella termodinámica típica de la unión de ligandos ricos en prolina no es totalmente opuesta a lo que cabria esperar si tenemos en cuenta la hidrofobicidad de los residuos implicados en la interacción. Esto sugiere la participación de otros factores en el mecanismo de unión, como la cooperatividad y la mediación de

## CAPÍTULO 6

interacciones por moléculas de agua. Conocer con mayor detalle este segundo nivel del el mecanismo de unión es esencial para el diseño eficaz de ligandos.

En el trabajo previo realizado en nuestro grupo de investigación se describió por primera vez la presencias de moléculas de agua enterradas en el sitio de unión del dominio SH3 de cAbl unido al ligando de alta afinidad p41 y se determinó su implicación en la energía de unión. Un posterior análisis de las estructuras cristalográficas de alta resolución disponibles en las bases de datos de la familia de dominios SH3 permitió confirmar la existencia de moléculas de agua enterradas en todos los sitios de unión de la familia de dominios SH3. Este nuevo paradigma sitúa a las moléculas de agua enterradas en el sitio de unión como parte formal del sitio de unión, algo que debemos tener en cuenta en el desarrollo de nuevos ligandos de alta afinidad y especificidad para los dominios SH3. Por otro lado los estudios desarrollados en nuestro grupo de investigación, también sugieren que no todas las moléculas de agua podrían tener la misma implicación en el sitio de unión, ya que muestran diferentes frecuencias de conservación y diferentes propiedades dinámicas. Poseer una herramienta capaz de predecir la posición y las propiedades dinámicas de las moléculas de agua para un dominio dado, seria de gran utilidad en el diseño racional de ligandos. Conocer las propiedades dinámicas de las moléculas de agua es igual de vital que conocer su posición, dado que en el proceso de diseño de ligandos, y que determina la estrategia a seguir para rentabilizar de la forma mas óptima la presencia de dichas moléculas de agua. Pero para todo ello es necesario adquirir una mayor cantidad de información con diferentes complejos y dominios. En esta tesis, hemos desarrollado un algoritmo capaz de registrar las propiedades dinámicas mas relevantes a lo largo de las trayectorias de dinámica molecular con el objetivo de comprender mejor las propiedades de estas moléculas de

agua y su implicación en el mecanismo de unión. Este algoritmo, una vez validado nos permite realizar un estudio a mayor escala dentro de la familia de dominios SH3 y permitiendo describir las diferencias que existen en el patrón de distribución de moléculas de agua entre miembros de diferentes subfamilias de dominios SH3, como por ejemplo la comparación de patrón de moléculas de agua entre el dominio SH3 de la quinasa de tirosina Fyn y el de la quinasa de tirosina de cAbl.

Por otra parte, y también en nuestro grupo de investigación, mediante el examen comparativo de los parámetros termodinámicos de unión de los dominios de SH3 cSrc, Fyn y cYes (todos pertenecientes a la subfamilia de Src y con una alta homología de secuencia) con el ligando de alta afinidad VSL12, se detectó una importante diferencia en la entalpía de unión (más de  $20\text{kJ}\cdot\text{mol}^{-1}$ ). Con la intención de describir cuáles eran los residuos implicados en esta gran diferencia entálpica, se realizó un completo estudio de mutagénesis dirigida sobre el dominio cSrc. Sorprendentemente este estudio mostró que los residuos con mayor implicación en las diferencias en la entalpía de unión observadas, se encontraban a más de  $10\text{ \AA}$  del sitio de unión, siendo especialmente importante la variación al mutar el residuo de la posición 128 de cSrc por el residuo que presentan cYes y Fyn. En esta tesis, mediante el profundo análisis de trayectorias de dinámica molecular hemos caracterizado en detalle la organización diferencial de las interacciones intramoleculares e intermoleculares (enlaces de hidrógeno y puentes salinos) de los mutantes estudiados por calorimetría, ya que los resultados del estudio calorimétrico sugerían la existencia de una alta cooperatividad del dominio, cuya dinámica era importante en el proceso de unión a ligando.

Por otro lado, los dominios modulares no son solo importantes para el reconocimiento específico, sino que también se encuentran implicados en la

## CAPÍTULO 6

regulación de la actividad de algunas quinasas, como por ejemplo la quinasa de tirosina de cAbl donde las interacciones intramoleculares específicas, entre el tándem de los dominios modulares SH3-SH2 y el dominio catalítico, mantienen la quinasa de tirosina autoinhibida. Dichas interacciones se encuentran moduladas mediante un complejo mecanismo alostérico que implica numerosas regiones de cAbl. Esta quinasa de tirosina es un proto-oncogen, de forma que la desregulación de cAbl por la interrupción de uno o varios elementos, conduce a la activación constitutiva de la actividad quinasa de cAbl, lo que tiene un efecto desastroso en la célula que acaba por devenir cancerígena.

En los últimos años se ha comprobado como el mecanismo de regulación de esta quinasa está íntimamente ligado a un gran cambio conformacional, en el que participan casi todas las regiones de cAbl. Este complejo mecanismo respondería a la necesidad de cAbl de responder de forma específica a diferentes estímulos. Pero a pesar de la numerosa información aun se desconoce como funciona el mecanismo de regulación o si existen en efecto diferentes estados activos. En esta tesis, mediante técnicas de simulación de dinámica molecular y simulaciones de metadinámica hemos estudiado principalmente el equilibrio conformacional del tándem SH3-SH2 y como éste es modulado por las diferentes partes de cAbl, con el objetivo de detectar los residuos clave que responsables del cambio de conformación del tándem, así como aportar nueva información al papel que juegan algunas de estas regiones de cAbl, como es el extremo N-terminal denominado Ncap y la importancia de la fosforilación de la serina 69. Y así proponer como cAbl podría adoptar diferentes estados activos .

## 6.2 CONCLUSIONES

Del estudio comparativo de la dinámica estructural de los diferentes mutantes del dominio SH3-cSrc, recogidos en el Capítulo 2 se extraen las siguientes conclusiones.

- 1- Del examen comparativo de las trayectorias de dinámica molecular para cada modelo, libre y en complejo con el ligando VSL12 hemos observado como el comportamiento del modelo R128-cSrc es diferente al de los modelos para los mutantes K128-cYes y E128-Fyn, siendo la presencia de la arginina en la posición 128 el origen de una sensible reorganización de los enlaces de hidrógeno y puentes de salinos tanto intermoleculares como intramoleculares. Aunque es difícil cuantificar la contribución de esta variación, todo apuntaría a que estaría en buena correlación con las diferencias medidas experimentalmente.

Del trabajo realizado en el Capítulo 3 con la propiedades dinámicas de las moléculas de agua de los dominios SH3.

- 1- Hemos desarrollado y validado un algoritmo que es capaz de registrar las principales características de las moléculas de agua enterradas en la interfaz de interacción de proteínas, a lo largo de las trayectorias de dinámica molecular (ZAHORÍ)
- 2- Utilizando el programa ZAHORÍ, se compararon las propiedades dinámicas de las moléculas de agua enterradas para los complejos SH3-cAbl/p41 y SH3-Fyn/VSL12. Ambos

## CAPÍTULO 6

dominios aunque pertenecen a diferentes subfamilias comparten una alta homología y sus ligandos son altamente específicos. Los resultados obtenidos por nuestro algoritmo nos permitieron determinar la existencia de diferentes patrones de distribución en las moléculas de agua para cada dominio, en especial en los lazos n-Src y RT, que son las regiones encargadas de la especificidad.

Gracias al estudio computacional llevado a cabo en el Capítulo 4, hemos podido extraer las siguientes conclusiones.

- 1- Comprobar que el tándem en ausencia de interacciones con otras regiones de la quinasa cAbl, se encuentra explorando mayoritariamente dos conformaciones que coinciden con las descritas en el análisis de las estructuras cristalográficas.
- 2- El análisis estructural y dinámico de las diferentes constructos nos ha permitido comprobar que la región que conecta los dominios SH3 y SH2 es la principal responsable de modular este equilibrio conformacional del tándem, otorgando suficiente flexibilidad para cambiar de conformación, pero a la vez limitando las conformaciones posibles. En esta región, los residuos con una mayor contribución serían los que se encuentran en las posiciones 139, 140 y 141.
- 3- Todos los datos recopilados mediante dinámicas moleculares y simulaciones de metadinámica apuntan a que la región denominada como Ncap actúa como interruptor del cambio

conformacional del tándem SH3-SH2. En este mecanismo, la fosfoserina que se encuentra en esta región sería indispensable para mantener las interacciones entre el Ncap y el tándem, cuya pérdida desencadena un rápido cambio de conformación. Por otro lado, los resultados obtenidos de las metadinámicas indicarían que la región Ncap también modularía el equilibrio conformacional del tándem SH32. Las interacciones intermoleculares de Ncap reorganizarían la distribución conformacional entre la conformación cerrada y una tercera conformación que podría corresponder a un estado activado o un intermedio del mecanismo allostérico de regulación.



---

**7 SUMMARY AND CONCLUSIONS**

CAPÍTULO 7

## 7.1 SUMMARY

Proteins are involved in most cellular processes. In order to keep the cellular homeostasis the proteins establish a complex network of interactions that is critical for the correct regulation of the cellular activity. Any change in this network may result in serious distortions of cellular functions leading to the development of many human diseases, such as cancer. The family of protein kinases plays a particularly relevant role in signal transduction. These enzymes are responsible for the phosphorylation of specific residues leading to the modification of the activity of the target protein. The specific recognition of amino acid sequences plays an important role in this process. This is carried out mainly by modular domains. The modular domains are implicated in the development of several human pathologies becoming, in many cases, attractive targets for the design of specific inhibitors of therapeutic interest. Among the different families of protein interaction modules, the family of polyproline recognition domains are especially relevant. This family is composed of different members, such as SH3, WW, GYF, UEV and EVH domains. A fine understanding of the recognition mechanism of proline rich sequences by these domains is essential for the rational design of efficient and specific inhibitors. Nonetheless, in spite the small size and apparent simplicity of the binding sites of these domains, a clear understanding of the molecular basis of binding affinity and specificity in these systems remains elusive. For example, even though the binding process implicates mostly hydrophobic residues, the thermodynamic fingerprint of the binding of proline rich motif ligands to their target domains is that typical of polar interactions and, consequently, cannot be explained considering exclusively the properties of the residues directly implicated in the interaction. This suggests that the contribution of other factors to the binding

## CAPÍTULO 7

mechanism, such as cooperative effects on the protein and/or the ligand and interactions mediated by water molecules might be relevant elements in the recognition process. Understanding how these elements modulate the binding mechanism is essential for the successful design of ligands.

In previous works our research group has investigated the presence of water molecules buried in the binding site of the SH3 domain of c-Abl. These water-mediated interactions were demonstrated to contribute significantly to the anomalous thermodynamic profile of the interaction between the SH3 domain of the cAbl tyrosine kinase and the p41 high affinity peptide ligand. From this analysis, it was established that the binding of proline-rich ligands to Abl-SH3 takes place by a dual binding mechanism, in which, the stacking of hydrophobic interactions is complemented by the establishment of an intricate network of water-mediated hydrogen bonds. The analysis of the structural database of SH3 domains, including complexes, free domains and full-length proteins, confirmed that this dual binding mechanism is a universal feature of SH3 domains, calling for a revision of the currently accepted binding paradigm. Even though water-mediated interactions were observed in all complexes studied, a wide variety of hydration profiles were identified, indicating the existence of different types of hydration sites and pointing towards a possible role of interfacial water molecules in the determination of binding specificity in these systems. Understanding the role and dynamic properties of each water molecule at the SH3-ligand binding interface is crucial for a complete understanding of these interactions and, thus, for the development of potent and specific inhibitors of therapeutic potential.

The properties of interfacial water molecules can be analyzed by molecular dynamics simulations. Nonetheless, these are computationally intensive procedures, the analysis of which requires a considerable amount of

time and effort. In order to perform this type of analysis more efficiently and be able to extend it to a high number of systems, in this thesis we have developed a computational tool to automatically analyze molecular dynamic simulations data in order to extract the dynamic properties of the buried interfacial water molecules and study their interactions with the protein and ligand atoms. This automatic and fast methodology can be applied to the study a large set of different domains and complexes and will provide enough information to define an adequate strategy to efficiently take advantage of the second level of interactions provided by water mediated interactions in the design of inhibitors and constitutes an initial step for the future development of a tool for the prediction of hydration sites in protein complexes. This new methodology has been validated with the study of several systems of interest.

In addition to the analysis of the influence of interfacial water molecules, in this thesis, we have also tackled the study of the impact of cooperative interactions and the reorganization of the conformational distribution of SH3 domains and peptides to the thermodynamic profile of binding. In this sense, our research group has recently performed a complete study of the binding energetics of several SH3 domains from the Src family (cSrc, Fyn and cYes) with a high affinity ligand (VSL12) revealing very large differences in the binding enthalpy between complexes with practically identical binding sites. After a complete mutagenic study of SH3-cSrc domain, it was found that the residues with the largest influence in the binding enthalpy difference are located at the distal loop, more than 10 Å away from the binding site. In this thesis, through deep analysis of molecular dynamics trajectories of these domains and some mutants of interest we have identified a differential reorganization of intramolecular interactions, hydrogen bonds and salt bridges, originated at the mutated position in the distal loop that is

## CAPÍTULO 7

propagated throughout the SH3 domain to affect residues at the binding site. This study clearly indicates that the differential reorganization of the conformational equilibrium of the free SH3 domains upon binding of the peptide ligand contributes very significantly to the highly exothermic character of the interaction, further illustrating the high level of complexity of SH3 interactions.

Finally, in this thesis we have also tackled the study of the effect of contextual factors, such as the presence of additional domains in tandem, to the conformational behaviour of SH3 domains and their functional impact. In this sense, even though most modular domains act as docking sites for the recognition of substrates, in some cases, such as the SH3 domains of Src and Abl tyrosine kinase, they are also implicated in the regulation of the catalytic activity of the proteins that contain them. In the case of Abl, this regulation takes place through the establishment of specific intra-molecular interactions of the tandem of the SH3 and SH2 domains with the catalytic domain of tyrosine kinase, which keep the enzyme inactive. These interactions are modulated by a complex allosteric mechanism involving many regions of cAbl. The activation of cAbl is related to a large conformational change, but this structure is unknown. This tyrosine kinase is a proto-oncogene, such that deregulation of cAbl by disruption of one or several elements, leads to constitutive activation of the kinase activity of cAbl, which have a disastrous effect on the cell that eventually become cancerous. In recent years, it was postulated that this complex mechanism responds to the necessity of having different activation states to different stimuli specified by cAbl. After years of studies, the regulatory mechanism remains evasive to a complete understanding. In this thesis, we have used molecular dynamics and metadynamics simulations to study the conformational equilibrium of the Abl

SH3-SH2 tandem and how this is affected by the different elements of cAbl, such as the connection linkers between domains or the N-Cap region. From this study, key residues responsible of the tandem conformational change were identified, providing new and important information about the role of these regions in the regulation mechanism of cAbl. Deciphering the regulation code of cAbl is of relevance for the development of novel and effective treatments against chronic myelogenous leukaemia. .

## 7.2 CONCLUSIONS

Chapter 2. From the thermodynamic, structural and dynamic analysis performed in chapter 1 we conclude:

1. From the comparative study of the trajectory of molecular dynamics simulation for each model, free and in complex with the high affinity ligand VSL12, we observe how the dynamic behaviour of the model R128-cSrc is different to the observed for the models K128-E128-Fyn CYES and also show how the presence of the arginine in the 128 position of the domain reconfigure the hydrogen bond network and salt bridge of the domain and the ligand. Both behaviours are in good agreement with the thermodynamic results.

## CAPÍTULO 7

Chapter 3. From the analysis of the properties of the buried water molecules at the binding site of the SH3 domains carried out in the third chapter of this thesis, we conclude.:

1. We have developed an algorithm to analyze the dynamic properties of the buried water molecules of the binding site; ZAHORI. This algorithm has been validated for its application to the study of hydration sites in SH3 complexes. This algorithm constitutes a potent tool for the automatic and efficient analysis of interfacial hydration in protein complexes, opening the door to a thorough analysis of different protein families and the development of prediction tools for protein hydration.
2. From the application of ZAHORI to the study of hydration in different SH<sub>3</sub> complexes (SH3-cAbl/p41 and SH3-Fyn/VSL12) we confirm that each domain has a different pattern distribution of hydration sites in the binding site, with different dynamic properties. Establishing these properties is essential for the efficient incorporation of water-mediated interactions into binding site descriptions for rational design or ligand identification. We observe that water molecules have different role and dynamic properties in the binding site as It was proposed previously. It is important to know where are the water molecules and with dynamic properties have to chose the right strategy to take advantage of theses water molecules when we design the inhibitors.

Chapter 4. From the analysis of the conformational equilibrium of the cAbl SH3-SH2 tandem, we conclude:

1. The conformational equilibrium of the isolated cAbl-SH3-SH2 tandem, in the absence of intermolecular interactions with other regions of the cAbl, as C2Q and Ncap regions, is characterized by two major available conformations, characterized by different relative orientations of the two domains originated from the rotation of the two domains around the SH3-SH2 connection linker, which correspond closely to those conformations observed in the available crystal structures.
2. The specific sequence of the SH3-SH2 connector is relevant for the properties of the conformational distribution of the tandem, with residues at position 139, 140 and 141 playing a particularly relevant role. Mutations at these positions significantly alter the population of the different conformations accessible to the tandem, modulating, thus, the regulation mechanism of the kinase. The sequence of the SH3-SH2 linker is optimized to energetically favour the close (inactive) conformation in which the SH3 domain is bound to the kinase module and an open conformation allowing binding of external ligands and activation of the kinase.
3. The Ncap region of Abl also plays a significant role in the regulation mechanism of c-Abl kinase. The elimination of the interactions established between a phosphoserine residue at position 69 in the NCap region and the SH3-SH2 tandem leads to a conformational change, suggesting that this position

## CAPÍTULO 7

might act as a conformational switch mediated by the action of serine kinases and phosphatases. The presence of the NCap modulates the conformational space of the tandem favouring the inactive conformation and stabilizing alternative conformations, not significantly populated in the absence of this region, supporting the hypothesis of the existence of a different activation state.



BIBLIOGRAFÍA

1. Pawson T (1995) Protein modules and signalling networks. *Nature* **373**, 573-580.
2. Manning G, Whyte D, Martinez R & Hunter... T (2002) The protein kinase complement of the human genome. *Science*.
3. Czernilofsky AP, Levinson AD, Varmus HE, Bishop JM, Tischer E & Goodman HM (1980) Nucleotide sequence of an avian sarcoma virus oncogene (src) and proposed amino acid sequence for gene product. *Nature* **287**, 198-203.
4. Smart JE, Oppermann H, Czernilofsky AP, Purchio AF, Erikson RL & Bishop JM (1981) Characterization of sites for tyrosine phosphorylation in the transforming protein of Rous sarcoma virus (pp60v-src) and its normal cellular homologue (pp60c-src). *Proc Natl Acad Sci USA* **78**, 6013-6017.
5. Stehelin D, Fujita DJ, Padgett T, Varmus HE & Bishop JM (1977) Detection and enumeration of transformation-defective strains of avian sarcoma virus with molecular hybridization. *Virology* **76**, 675-684.
6. Gilboa E, Witte O & Baltimore D (1980) Structure of the Abelson murine leukemia virus genome and the homologous cellular gene: studies with cloned viral DNA. *Cell*.
7. Yin Jen Wang J, Ledley F, Goff S, Lee R & Groner... Y (1984) The mouse c-abl locus: molecular cloning and characterization. *Cell*.
8. Groffen J, Stephenson J & Heisterkamp... N (1984) Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. *Cell*.
9. Konopka J & Watanabe... S (1984) An alteration of the human c-abl protein in K562 leukemia cells unmasks associated tyrosine kinase activity. *Cell*.
10. Huse M & Kuriyan J (2002) The conformational plasticity of protein kinases. *Cell* **109**, 275-282.
11. Rosen OM, Herrera R, Olowe Y, Petruzzelli LM & Cobb MH (1983) Phosphorylation activates the insulin receptor tyrosine protein kinase. *Proc Natl Acad Sci USA* **80**, 3237-3240.
12. Hubbard SR & Till JH (2000) Protein tyrosine kinase structure and function. *Annu Rev Biochem* **69**, 373-398.
13. Mohammadi M, Schlessinger J & Hubbard SR (1996) Structure of the FGF receptor tyrosine kinase domain reveals a novel autoinhibitory mechanism. *Cell* **86**, 577-587.
14. Cowan-Jacob S, Ramage P & Stark W (2006) *Structural biology of protein tyrosine kinases*. Human Press.

## BIBLIOGRAFÍA

15. Stone JC, Atkinson T, Smith M & Pawson T (1984) Identification of functional regions in the transforming protein of Fujinami sarcoma virus by in-phase insertion mutagenesis. *Cell* **37**, 549-558.
16. Sadowski I, Stone JC & Pawson T (1986) A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Mol Cell Biol* **6**, 4396-4408.
17. Mayer BJ, Hamaguchi M & Hanafusa H (1988) A novel viral oncogene with structural similarity to phospholipase C. *Nature* **332**, 272-275.
18. Yadav S & Miller WT (2008) The Evolutionarily Conserved Arrangement of Domains in Src Family Kinases Is Important for Substrate Recognition†. *Biochemistry* **47**, 10871-10880.
19. Birge RB, Knudsen BS, Besser D & Hanafusa H (1996) SH2 and SH3-containing adaptor proteins: redundant or independent mediators of intracellular signal transduction. *Genes Cells* **1**, 595-613.
20. Mayer BJ & Baltimore D (1993) Signalling through SH2 and SH3 domains. *Trends Cell Biol* **3**, 8-13.
21. Pendergast A (2002) The Abl family kinases: mechanisms of regulation and signaling. *Advances in cancer research*.
22. Barilá D & Superti-Furga G (1998) An intramolecular SH3-domain interaction regulates c-Abl activity. *Nat Genet* **18**, 280-282.
23. Colicelli J (2010) ABL Tyrosine Kinases: Evolution of Function, Regulation, and Specificity. *Science signaling* **3**, re6.
24. Hantschel... O (2004) Regulation of the c-Abl and Bcr–Abl tyrosine kinases. *Nature Reviews Molecular Cell ....*
25. Alvarado JJ, Betts L, Moroco JA, Smithgall TE & Yeh JI (2010) Crystal structure of the Src-family kinase Hck SH3-SH2-linker regulatory region supports an SH3-dominant activation mechanism. *J Biol Chem*.
26. Chen S, Brier S, Smithgall TE & Engen JR (2007) The Abl SH2-kinase linker naturally adopts a conformation competent for SH3 domain binding. *Protein Sci* **16**, 572-581.
27. Iacob RE, Pene-Dumitrescu T, Zhang J, Gray NS, Smithgall TE & Engen JR (2009) Conformational disturbance in Abl kinase upon mutation and deregulation. *Proc Natl Acad Sci USA* **106**, 1386-1391.
28. Sherbenou DW, Hantschel O, Kaube I, Willis S, Bumm T, Turaga LP, Lange T, Dao K-H, Press RD, Druker BJ, Superti-Furga G & Deininger MW (2010) BCR-ABL SH3-SH2 domain mutations in chronic myeloid leukemia patients on imatinib. *Blood*.
29. Nagar B (2007) c-Abl tyrosine kinase and inhibition by the cancer drug imatinib (Gleevec/STI-571). *J Nutr* **137**, 1518S-1523S; discussion 1548S.

30. Miyoshi-Akiyama T, Aleman LM, Smith JM, Adler CE & Mayer BJ (2001) Regulation of Cbl phosphorylation by the Abl tyrosine kinase and the Nck SH2/SH3 adaptor. *Oncogene* **20**, 4058-4069.
31. Barilá D, Mangano R, Gonfloni S, Kretzschmar J, Moro M, Bohmann D & Superti-Furga G (2000) A nuclear tyrosine phosphorylation circuit: c-Jun as an activator and substrate of c-Abl and JNK. *EMBO J* **19**, 273-281.
32. Nagar B, Hantschel O, Seeliger M, Davies JM, Weis WI, Superti-Furga G & Kuriyan J (2006) Organization of the SH3-SH2 unit in active and inactive forms of the c-Abl tyrosine kinase. *Mol Cell* **21**, 787-798.
33. Lee K-Y, Hyeok Yoon JH, Kim M, Roh S, Lee Y-S, Seong B-L & Kim K (2002) A dipalmitoyl peptide that binds SH3 domain, disturbs intracellular signal transduction, and inhibits tumor growth in vivo. *Biochem Biophys Res Commun* **296**, 434-442.
34. Pawson T & Kofler M (2009) Kinome signaling through regulated protein-protein interactions in normal and cancer cells. *Current opinion in cell biology* **21**, 147-153.
35. Cowburn D (1997) Modular peptide recognition domains in eukaryotic signaling. *Annual review of biophysics and ...*
36. Kay BK, Williamson MP & Sudol M (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J* **14**, 231-241.
37. Mayer BJ (2001) SH3 domains: complexity in moderation. *J Cell Sci* **114**, 1253-1263.
38. Zarrinpar A, Bhattacharyya RP & Lim WA (2003) The structure and function of proline recognition domains. *Sci STKE* **2003**, RE8.
39. Zarrinpar A, Park S-H & Lim WA (2003) Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* **426**, 676-680.
40. Musacchio A, Noble M, Pauptit R, Wierenga R & Saraste M (1992) Crystal structure of a Src-homology 3 (SH3) domain. *Nature* **359**, 851-855.
41. Lim WA, Richards FM & Fox RO (1994) Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature* **372**, 375-379.
42. Feng S, Chen J, Yu H, Simon J & Schreiber S (1994) Two binding orientations for peptides to the Src SH3 domain: Development of a general model for SH3-ligand interactions. *Science* **266**, 1241-1247.
43. Li SS-C (2005) Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction. *Biochem J* **390**, 641-653.
44. Wu C, Ma MH, Brown KR, Geisler M, Li L, Tzeng E, Jia CYH, Jurisica I & Li SS-C (2007) Systematic identification of SH3 domain-mediated human

## BIBLIOGRAFÍA

protein-protein interactions by peptide array target screening. *Proteomics* **7**, 1775-1785.

45. Moncalián G, Cárdenes N, Deribe YL, Spínola-Amilibia M, Dikic I & Bravo J (2006) Atypical polyproline recognition by the CMS N-terminal Src homology 3 domain. *J Biol Chem* **281**, 38845-38853.

46. Kato M, Miyazawa K & Kitamura N (2000) A deubiquitinating enzyme UBPY interacts with the Src homology 3 domain of Hrs-binding protein via a novel binding motif PX(V/I)(D/N)RXXKP. *J Biol Chem* **275**, 37481-37487.

47. Pisabarro... M (1996) Rational design of specific high-affinity peptide ligands for the Abl-SH3 domain. *Biochemistry*.

48. Kanner SB, Reynolds AB, Wang HC, Vines RR & Parsons JT (1991) The SH2 and SH3 domains of pp60src direct stable association with tyrosine phosphorylated proteins p130 and p110. *EMBO J* **10**, 1689-1698.

49. Flynn DC, Leu TH, Reynolds AB & Parsons JT (1993) Identification and sequence analysis of cDNAs encoding a 110-kilodalton actin filament-associated pp60src substrate. *Mol Cell Biol* **13**, 7892-7900.

50. Summy JM & Gallick GE (2003) Src family kinases in tumor progression and metastasis. *Cancer Metastasis Rev* **22**, 337-358.

51. Velazquez-Campoy A, Luque I & Freire E (2001) The application of thermodynamic methods in drug design. *Thermochimica acta* **380**, 217-227.

52. Ferreon J & Hilser V (2004) Thermodynamics of binding to SH3 domains: the energetic impact of polyproline II (PII) helix formation. *Biochemistry* **43**, 7787-7797.

53. Palencia A, Cobos E, Mateo P, Martínez J & Luque I (2004) Thermodynamic dissection of the binding energetics of proline-rich peptides to the Abl-SH3 domain: implications for rational ligand design. *J Mol Biol* **336**, 527-537.

54. Renzoni D, Pugh D, Siligardi G, Das P, Morton C, Rossi C, Waterfield M, Campbell I & Ladbury J (1996) Structural and thermodynamic characterization of the interaction of the SH3 domain from Fyn with the proline-rich binding site on the p85 subunit of PI3-kinase. *Biochemistry* **35**, 15646-15653.

55. Casares S, López-Mayorga O, Vega MC, Cámara-Artigas A & Conejero-Lara F (2007) Cooperative propagation of local stability changes from low-stability and high-stability regions in a SH3 domain. *Proteins: Structure, Function and Genetics* **67**, 531-547.

56. Casares S, Ab E, Eshuis H, Lopez-Mayorga O, van Nuland NAJ & Conejero-Lara F (2007) The high-resolution NMR structure of the R21A Spc-SH3:P41 complex: understanding the determinants of binding affinity by comparison with Abl-SH3. *BMC Struct Biol* **7**, 22.

57. Martin-Garcia JM, Ruiz-Sanz J & Luque I (2012) Interfacial water molecules in SH3 interactions: a revised paradigm for polyproline recognition. *Biochem J* **442**, 443-451.
58. Brasher BB, Roumiantsev S & Van Etten RA (2001) Mutational analysis of the regulatory function of the c-Abl Src homology 3 domain. *Oncogene* **20**, 7744-7752.
59. Amanchy R, Zhong J, Molina H, Chaerkady R, Iwahori A, Kalume DE, Grønborg M, Joore J, Cope L & Pandey A (2008) Identification of c-Src tyrosine kinase substrates using mass spectrometry and peptide microarrays. *J Proteome Res* **7**, 3900-3910.
60. Collette Y, Arold S, Picard C, Janvier K, Benichou S, Benarous R, Olive D & Dumas C (2000) HIV-2 and SIV Nef proteins target different Src family SH3 domains than does HIV-1 Nef because of a triple amino acid substitution. *Journal of Biological Chemistry* **275**, 4171-4176.
61. Sudol M (1994) Yes-associated protein (YAP65) is a proline-rich phosphoprotein that binds to the SH3 domain of the Yes proto-oncogene product. *Oncogene* **9**, 2145-2152.
62. Summy J, Guappone A, Sudol M & Flynn D (2000) The SH3 and SH2 domains are capable of directing specificity in protein interactions between the non-receptor tyrosine kinases cSrc and cYes. *Oncogene* **19**, 155-160.
63. Feng S, Kasahara C, Rickles RJ & Schreiber SL (1995) Specific interactions outside the proline-rich core of two classes of Src homology 3 ligands. *Proc Natl Acad Sci USA* **92**, 12408-12415.
64. Cordier F, Wang C, Grzesiek S & Nicholson LK (2000) Ligand-induced strain in hydrogen bonds of the c-Src SH3 domain detected by NMR. *J Mol Biol* **304**, 497-505.
65. Wang W, Lim WA, Jakalian A, Wang J, Wang J, Luo R, Bayly CI & Kollman PA (2001) An analysis of the interactions between the Sem-5 SH3 domain and its ligands using molecular dynamics, free energy calculations, and sequence analysis. *J Am Chem Soc* **123**, 3986-3994.
66. Casares S, Sadqi M, Lopez-Mayorga O, Martinez J & Conejero-Lara F (2003) Structural cooperativity in the SH3 domain studied by site-directed mutagenesis and amide hydrogen exchange. *FEBS letters* **539**, 125-130.
67. Cooper A & Johnson C (1994) Introduction to microcalorimetry and biomolecular energetics. *Methods in molecular biology* **22**, 109-124.
68. Fisher H & Singh N (1995) Calorimetric methods for interpreting protein-ligand interactions. *Methods in Enzymology* **259**, 194-221.
69. Luque I & Freire E (2002) Structural parameterization of the binding enthalpy of small ligands. *Proteins: Structure, Function and Genetics* **49**, 181-190.

## BIBLIOGRAFÍA

70. Perozzo R, Folkers G & Scapozza L (2004) Thermodynamics of protein-ligand interactions: History, presence, and future aspects. *Journal of Receptors and Signal Transduction* **24**, 1-52.
71. Hyeon C, Lorimer G & Thirumalai D (2006) Dynamics of allosteric transitions in GroEL. *Proc Natl Acad Sci USA* **103**, 18939-18944.
72. Inobe T, Kikushima K, Makio T, Arai M & Kuwajima K (2003) The allosteric transition of GroEL induced by metal fluoride-ADP complexes. *Journal of Molecular Biology* **329**, 121-134.
73. Ma J, Sigler PB, Xu Z & Karplus M (2000) A dynamic model for the allosteric mechanism of GroEL. *Journal of Molecular Biology* **302**, 303-313.
74. Tai K, Shen T, Bjerrsson U, Philippopoulos M & McCammon JA (2001) Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase. *Biophysical Journal* **81**, 715-724.
75. D.A. Case TAD, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, Ross C. Walker, W. Zhang, K.M. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F.Wong, F. Paesani, J. Vanicek, X.Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, and P.A. Kollman (2008) AMBER 10. In: University of California, San Francisco.
76. Forrest LR, Kukol A, Arkin IT, Peter Tieleman D & Sansom MSP (2000) Exploring models of the influenza A M2 channel: MD simulations in a phospholipid bilayer. *Biophysical Journal* **78**, 55-69.
77. Gullingsrud J, Kosztin D & Schulten K (2001) Structural determinants of MscL gating studied by molecular dynamics simulations. *Biophysical Journal* **80**, 2074-2081.
78. Yang L, Song G & Jernigan RL (2007) Comparison of experimental and computed protein anisotropic temperature factors. In *Proceedings - 2007 IEEE International Conference on Bioinformatics and Biomedicine Workshops, BIBMW*, pp. 89-96.
79. Kumar S & Nussinov R (2002) Close-range electrostatic interactions in proteins. *ChemBiochem*.
80. Freire E (2008) Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov Today* **13**, 869-874.
81. Lafont V, Armstrong AA, Ohtaka H, Kiso Y, Mario Amzel L & Freire E (2007) Compensating enthalpic and entropic changes hinder binding affinity optimization. *Chem Biol Drug Des* **69**, 413-422.
82. Tsai CJ, Buyong M, Sham YY, Kumar S & Nussinov R (2001) Structured disorder and conformational selection. *Proteins: Structure, Function and Genetics* **44**, 418-427.

83. Sokalingam S, Raghunathan G, Soundrarajan N & Lee S-G (2012) A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS ONE* **7**, e40410.
84. Nagendra HG, Sukumar N & Vijayan M (1998) Role of water in plasticity, stability, and action of proteins: the crystal structures of lysozyme at very low levels of hydration. *Proteins: Structure, Function and Genetics* **32**, 229-240.
85. Papoian G, Ulander J, Eastwood MP, Luthey-Schulten Z & Wolynes PG (2004) Water in protein structure prediction. *Proc Natl Acad Sci USA* **101**, 3352-3357.
86. Petukhov M, Cregut D, Soares CM & Serrano L (1999) Local water bridges and protein conformational stability. *Protein Sci* **8**, 1982-1989.
87. Scheidig A & Burmester... C (1999) The pre-hydrolysis state of p21ras in complex with GTP: new insights into the role of water molecules in the GTP hydrolysis reaction of ras-like proteins. *Structure*.
88. Westhof E (1988) Water: an integral part of nucleic acid structure. *Annual review of biophysics and biophysical ....*
89. Janin J (1999) Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. *Structure*.
90. Janin J & C. C (1976) Stability and specificity of protein-protein interactions: the case of the trypsin-trypsin inhibitor complexes. *J Mol Biol* **100**, 197-211.
91. Ladbury J (2001) Isothermal titration calorimetry: application to structure-based drug design. *Thermochimica acta*.
92. Levy Y & Onuchic J (2006) Water mediation in protein folding and molecular recognition. *Annu Rev Biophys Biomol Struct* **35**, 389-415.
93. Zafra-Ruano A & Luque I (2012) Interfacial water molecules in SH3 interactions: Getting the full picture on polyproline recognition by protein-protein interaction domains. *FEBS letters*.
94. Palencia A, Camara-Artigas A, Pisabarro MT, Martinez JC & Luque I (2010) Role of interfacial water molecules in proline-rich ligand recognition by the Src homology 3 domain of Abl. *J Biol Chem* **285**, 2823-2833.
95. Halle B (2004) Protein hydration dynamics in solution: a critical survey. *Philos Trans R Soc Lond, B, Biol Sci* **359**, 1207-1223; discussion 1223-1204, 1323-1208.
96. Ohlendorf DH (1994) Accuracy of refined protein structures. II. Comparison of four independently refined models of human interleukin 1beta. *Acta Crystallogr D Biol Crystallogr* **50**, 808-812.
97. Badger J (1997) Modeling and refinement of water molecules and disordered solvent. *Meth Enzymol* **277**, 344-352.

## BIBLIOGRAFÍA

98. Bellissent-Funel M (2000) Hydration in protein dynamics and function. *Journal of Molecular Liquids*.
99. Svergun DI, Richard S, Koch MH, Sayers Z, Kuprin S & Zaccai G (1998) Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc Natl Acad Sci USA* **95**, 2267-2272.
100. Maeda A, Tomson FL, Gennis RB, Ebrey TG & Balashov SP (1999) Chromophore-protein-water interactions in the L intermediate of bacteriorhodopsin: FTIR study of the photoreaction of L at 80 K. *Biochemistry* **38**, 8800-8807.
101. Grdadolnik... J (2001) Bovine serum albumin observed by infrared spectrometry. II. Hydration mechanisms and interaction configurations of embedded H<sub>2</sub>O molecules. *Biopolymers*.
102. Sotomayor C, Aguilar L & Cuevas... F (2000) Modulation of pig kidney Na<sup>+</sup>/K<sup>+</sup>-ATPase activity by cholesterol: role of hydration. *Biochemistry*.
103. Blanch E, Bell A, Hecht L & Day... L (1999) Raman optical activity of filamentous bacteriophages: hydration of [alpha]-helices1. *Journal of molecular ...*
104. Yoshida T, Inagaki H & Kamyra H (2001) Two types of protein hydration measured by dielectric dispersion in the gigahertz region, and effects of anesthetics. *J Phys Chem B* **105**, 256-260.
105. Lounnas V, Lüdemann S & Wade RC (1999) Towards molecular dynamics simulation of large proteins with a hydration shell at constant pressure. *Biophysical chemistry* **78**, 157-182.
106. Walser R & van Gunsteren W (2001) Viscosity dependence of protein dynamics. *Proteins: Structure, Function and Genetics* **42**, 414-421.
107. Samsonov S, Teyra J & Pisabarro MT (2008) A molecular dynamics approach to study the importance of solvent in protein interactions. *Proteins: Structure, Function and Genetics* **73**, 515-525.
108. Park S & Saven JG (2005) Statistical and molecular dynamics studies of buried waters in globular proteins. *Proteins: Structure, Function and Genetics* **60**, 450-463.
109. Stajich JE & Lapp H (2006) Open source tools and toolkits for bioinformatics: significance, and where are we? *Brief Bioinformatics* **7**, 287-296.
110. Moll A, Hildebrandt A, Lenhof H-P & Kohlbacher O (2006) BALLView: a tool for research and education in molecular modeling. *Bioinformatics* **22**, 365-366.
111. Grünberg R, Nilges M & Leckner J (2007) Biskit--a software platform for structural bioinformatics. *Bioinformatics* **23**, 769-770.

112. Vohra S, Hall BA, Holdbrook DA, Khalid S & Biggin PC (2010) Bookshelf: a simple curation system for the storage of biomolecular simulation data. *Database (Oxford)* **2010**, baq033.
113. Fogh RH, Boucher W, Ionides JMC, Vranken WF, Stevens TJ & Laue ED (2010) MEMOPS: data modelling and automatic code generation. *J Integr Bioinform* **7**.
114. Bak A, Polanski J, Stockner T & Kurczyk A (2010) MoStBioDat--molecular and structural bioinformatics database. *Comb Chem High Throughput Screen* **13**, 366-374.
115. Fufezan C & Specht M (2009) p3d--Python module for structural bioinformatics. *BMC Bioinformatics* **10**, 258.
116. McDonald IK & Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. *J Mol Biol* **238**, 777-793.
117. Hubbard SJ & Thornton JM (1993) 'NACCESS', computer program. In *Techreport*.
118. Pawson T & Nash P (2003) Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445-452.
119. Strawn LM & Shawver LK (1998) Tyrosine kinases in disease: overview of kinase inhibitors as therapeutic agents and current drugs in clinical trials. *Expert Opin Investig Drugs* **7**, 553-573.
120. Abelson HT & Rabstein LS (1970) Lymphosarcoma: virus-induced thymic-independent disease in mice. *Cancer Research* **30**, 2213-2222.
121. Witte ON, Dasgupta A & Baltimore D (1980) Abelson murine leukaemia virus protein is phosphorylated in vitro to form phosphotyrosine. *Nature* **283**, 826-831.
122. Sefton BM, Hunter T & Raschke WC (1981) Evidence that the Abelson virus protein functions in vivo as a protein kinase that phosphorylates tyrosine. *Proc Natl Acad Sci USA* **78**, 1552-1556.
123. Ben-Neriah Y, Daley GQ, Mes-Masson AM, Witte ON & Baltimore D (1986) The chronic myelogenous leukemia-specific P210 protein is the product of the bcr/abl hybrid gene. *Science* **233**, 212-214.
124. Bradley WD & Koleske AJ (2009) Regulation of cell migration and morphogenesis by Abl-family kinases: emerging mechanisms and physiological contexts. *J Cell Sci* **122**, 3441-3454.
125. Wang JY (2000) Regulation of cell death by the Abl tyrosine kinase. *Oncogene* **19**, 5643-5650.
126. Woodring PJ, Hunter T & Wang JYJ (2003) Regulation of F-actin-dependent processes by the Abl family of tyrosine kinases. *J Cell Sci* **116**, 2613-2626.
127. Shaul Y & Ben-Yehoyada M (2005) Role of c-Abl in the DNA damage stress response. *Cell Res* **15**, 33-35.

## BIBLIOGRAFÍA

128. Backert S, Feller SM & Wessler S (2008) Emerging roles of Abl family tyrosine kinases in microbial pathogenesis. *Trends Biochem Sci* **33**, 80-90.
129. Gertler FB, Doctor JS & Hoffmann FM (1990) Genetic suppression of mutations in the *Drosophila* *abl* proto-oncogene homolog. *Science* **248**, 857-860.
130. Hurwitz ME, Vanderzalm PJ, Bloom L, Goldman J, Garriga G & Horvitz HR (2009) Abl kinase inhibits the engulfment of apoptotic [corrected] cells in *Caenorhabditis elegans*. *PLoS Biol* **7**, e99.
131. Van Etten RA (1999) Cycling, stressed-out and nervous: cellular functions of c-Abl. *Trends Cell Biol* **9**, 179-186.
132. Smith J & Mayer BJ (2002) Abl: mechanisms of regulation and activation. *Front Biosci* **7**, 31-42.
133. Raitano AB, Whang YE & Sawyers CL (1997) Signal transduction by wild-type and leukemogenic Abl proteins. *Biochim Biophys Acta* **1333**, F201-216.
134. Azam M, Latek RR & Daley GQ (2003) Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL. *Cell* **112**, 831-843.
135. Gorre ME, Ellwood-Yen K, Chiosis G, Rosen N & Sawyers CL (2002) BCR-ABL point mutants isolated from patients with imatinib mesylate-resistant chronic myeloid leukemia remain sensitive to inhibitors of the BCR-ABL chaperone heat shock protein 90. *Blood* **100**, 3041-3044.
136. Ernst T, La Rosée P, Müller MC & Hochhaus A (2011) BCR-ABL mutations in chronic myeloid leukemia. *Hematol Oncol Clin North Am* **25**, 997-1008, v-vi.
137. Dietrich J, Hulme C & Hurley LH (2010) The design, synthesis, and evaluation of 8 hybrid DFG-out allosteric kinase inhibitors: a structural analysis of the binding interactions of Gleevec, Nexavar, and BIRB-796. *Bioorg Med Chem* **18**, 5738-5748.
138. Zhang J, Adrián FJ, Jahnke W, Cowan-Jacob SW, Li AG, Iacob RE, Sim T, Powers J, Dierks C, Sun F, Guo G-R, Ding Q, Okram B, Choi Y, Wojciechowski A, Deng X, Liu G, Fendrich G, Strauss A, Vajpai N, Grzesiek S, Tuntland T, Liu Y, Bursulaya B, Azam M, Manley PW, Engen JR, Daley GQ, Warmuth M & Gray NS (2010) Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. *Nature* **463**, 501-506.
139. Iacob RE, Zhang J, Gray NS & Engen JR (2011) Allosteric interactions between the myristate- and ATP-site of the Abl kinase. *PLoS ONE* **6**, e15929.
140. Hantschel O, Nagar B, Guettler S, Kretzschmar J, Dorey K, Kuriyan J & Superti-Furga G (2003) A myristoyl/phosphotyrosine switch regulates c-Abl. *Cell* **112**, 845-857.

141. Shishido T, Akagi T, Ouchi T, Georgescu MM, Langdon WY & Hanafusa H (2000) The kinase-deficient Src acts as a suppressor of the Abl kinase for Cbl phosphorylation. *Proc Natl Acad Sci USA* **97**, 6439-6444.
142. Franz WM, Berger P & Wang JY (1989) Deletion of an N-terminal regulatory domain of the c-abl tyrosine kinase activates its oncogenic potential. *EMBO J* **8**, 137-147.
143. Pluk H, Dorey K & Superti-Furga G (2002) Autoinhibition of c-Abl. *Cell* **108**, 247-259.
144. Boggon TJ & Eck MJ (2004) Structure and regulation of Src family kinases. *Oncogene* **23**, 7918-7927.
145. Brown M & J.A. C (1996) Regulation, substrates and functions of src. *Biochim Biophys Acta* **1287**, 121-149.
146. Chen S, Dumitrescu TP, Smithgall TE & Engen JR (2008) Abl N-terminal cap stabilization of SH3 domain dynamics. *Biochemistry* **47**, 5795-5803.
147. Piana S & Laio A (2007) A bias-exchange approach to protein folding. *The journal of physical chemistry B* **111**, 4553-4559.
148. Alder B & Wainwright T (1959) Studies in molecular dynamics. I. General method. *J Chem Phys* **31**, 459-467.
149. Rahman A (1964) Correlations in the motion of atoms in liquid argon. *Physical Review*.
150. van Gunsteren W & Berendsen H (1977) Algorithms for macromolecular dynamics and constraint dynamics. *Molecular Physics*, 1311-1327.
151. Schneider T & Stoll E (1978) Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Physical Review B* **17**, 1302-1322.
152. F. W, van Gunsteren W & Berendsen H (1990) Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Chem Int Ed Engl* **29**, 992-1023.
153. Streett W, Tildesley D & Saville G (1978) Multiple time-step methods in molecular dynamics. *Molecular Physics* **35**, 639-648.
154. Ryckaert J, Ciccotti G & Berendsen H (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational Physics* **23**, 327-341.
155. Bekker H (1997) Unification of box shapes in molecular simulations. *J Comput Chem*.
156. Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*.
157. Berendsen H, Postma J, van Gunsteren W, Dinola A & Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* **81**.

## BIBLIOGRAFÍA

158. Hoover W (1985) Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*.
159. Nosé S (1984) A unified formulation of the constant temperature molecular dynamics methods. *J Chem Phys*.
160. Bussi G & Donadio... D (2007) Canonical sampling through velocity rescaling. *J Chem Phys*.
161. Loncharich R & Brooks... B (1992) Langevin dynamics of peptides: The frictional dependence of isomerization rates of N - acetylalanyl - N' - methylamide. *Biopolymers*.
162. Izaguirre J, Catarello D, Wozniak J & Skeel RD (2001) Langevin stabilization of molecular dynamics. *J Chem Phys*, 114-123.
163. Morriss G (1983) Isothermal-isobaric molecular dynamics. *Chemical Physics*.
164. Nosé S & Klein ML (1983) Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, 1055-1076.
165. Feller S, Zhang Y & Pastor... R (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. *The Journal of chemical ...*
166. Jorgensen W & Gao P (1988) Cis-trans energy difference for the peptide bond in the gas phase and in aqueous solution. *Journal of the American Chemical Society* **1988**, 4212-4216.
167. Brooks B, Bruccoleri R, Olafson BD, Swaminathan S & Karplus M (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4**, 187-217.
168. Kollman P & Nguyen D (1986) An all atom force field for simulations of proteins and nucleic acids. *Journal of Computational Chemistry* **7**, 230-252.
169. Kirkwood J (1931) The van der Waals forces in gases. *Physical Review*.
170. Loncharich... R (1989) The effects of truncating long - range forces on protein dynamics. *Proteins: Structure*.
171. Klauda J, Wu X & Pastor... R (2007) Long-range Lennard-Jones and electrostatic interactions in interfaces: application of the isotropic periodic sum method. *The Journal of Physical ...*
172. Lagüe P & Pastor... R (2004) Pressure-based long-range correction for Lennard-Jones interactions in molecular dynamics simulations: application to alkanes and interfaces. *The Journal of Physical ...*
173. Cheatham T, Miller J, Fox T, Darden T & Kollman P (1995) Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J Am Chem Soc* **117**, 4193-4194.
174. Gilson M (1995) Theory of electrostatic interactions in macromolecules. *Curr Opin Struct Biol*.

175. Heinz T & Hünenberger P (2005) Combining the lattice-sum and reaction-field approaches for evaluating long-range electrostatic interactions in molecular simulations. *J Chem Phys* **123**.
176. Míguez J, González-Salgado D, Legido J & Piñeiro M (2010) Calculation of interfacial properties using molecular simulation with the reaction field method: Results for different water models. *J Chem Phys* **132**.
177. Darden T, York D & Lee P (1993) Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J Chem Phys* **98**, 10089-10092.
178. Ewald P (1921) Ewald summation. *Ann Phys*.
179. Cieplak P & Kollman P (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry* **21**, 1049-1074.
180. MacKerell Jr A, Bashford D, Bellott M & Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586-3616.
181. Stocker U & van Gunsteren W (2000) Molecular dynamics simulation of hen egg white lysozyme: a test of the GROMOS96 force field against nuclear magnetic resonance data. *Proteins: Structure, Function and Genetics* **40**, 145-153.
182. Jorgensen W & Maxwell... D (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American ...*
183. Mackerell Jr A (2004) Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem*.
184. Teissie J (2007) Biophysical effects of electric fields on membrane water interfaces: a mini review. *European Biophysics Journal*.
185. Zhang L, Wang L, Kao Y & Qiu W (2007) Mapping hydration dynamics around a protein surface. *Proc Natl Acad Sci USA* **104**, 18461-18466.
186. Guillot B (2002) A reappraisal of what we have learnt during three decades of computer simulations on water. *Journal of Molecular Liquids*.
187. Jorgensen W & Tirado-Rives J (2005) Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc Natl Acad Sci USA* **102**, 6665-6670.
188. Wesson L & Eisenberg D (1992) Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci* **1**, 227-235.
189. Gilson MK & Honig B (1988) Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins: Structure, Function and Genetics* **4**, 7-18.

## BIBLIOGRAFÍA

190. Still W, Tempczyk A, Hawley R & Hendrickson RC (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* **112**, 6127-6129.
191. Qiu D, Shenkin P & Hollinger F (1997) The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J Phys Chem A* **101**, 3005-3014.
192. Levy R, Zhang L, Gallicchio E & Felts AK (2003) On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. *J Am Chem Soc* **125**, 9523-9530.
193. Feig M & Brooks CL (2004) Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr Opin Struct Biol* **14**, 217-224.
194. Jorgensen W, Chandrasekhar J & Madura JD (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 927-935.
195. Jorgensen W (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics*.
196. Berendsen H, Postma J, van Gunsteren W & Hermans J (1981) Interaction models for water in relation to protein hydration. *Intermolecular Forces*, 331.342.
197. Berendsen H, Grigera J & Straatsma T (1987) The missing term in effective pair potentials. *J Chem Phys* **91**, 6269-6271.
198. van der Spoel D & van Maaren... P (1998) A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *The Journal of chemical ...*
199. de Azevedo F (2011) Molecular Dynamics Simulations of Protein Targets Identified in Mycobacterium tuberculosis. *Curr Med Chem*.
200. Kollman P (1993) Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews*.
201. Zwanzig R (1954) High - Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J Chem Phys*.
202. Valleau J (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *Journal of Computational Physics*.
203. Pohorille A (2001) Calculating free energies using average force. *The Journal of Chemical Physics*.
204. Laio A & Parrinello M (2002) Escaping free-energy minima. *Proc Natl Acad Sci USA* **99**, 12562-12566.
205. Paci E & Vendruscolo... M (2002) Validity of Go models: Comparison with a solvent-shielded empirical energy decomposition. *Biophys J*.

206. Maragliano L, Fischer A, Vanden-Eijnden E & Ciccotti G (2006) String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J Chem Phys* **125**.
207. Faradjian AK & Elber R (2004) Computing time scales from reaction coordinates by milestoning. *J Chem Phys* **120**, 10880-10889.
208. Voter A (1997) A method for accelerating the molecular dynamics simulation of infrequent events. *Journal of Chemical Physics*.
209. Rodriguez-Gomez D, Darve E & Pohorille A (2004) Assessing the efficiency of free energy calculation methods. *J Chem Phys* **120**, 3563-3578.
210. de Oliveira CAF, Hamelberg D & McCammon JA (2008) Coupling Accelerated Molecular Dynamics Methods with Thermodynamic Integration Simulations. *J Chem Theory Comput* **4**, 1516-1525.
211. Branduardi D, Gervasio FL & Parrinello M (2007) From A to B in free energy space. *J Chem Phys* **126**, 054103.
212. Ensing B, De Vivo M, Liu Z & Klein ML (2006) Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc Chem Res* **39**, 73-81.
213. Laio A & Gervasio F (2008) Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics* **71**, 126601.
214. Leone V, Marinelli F, Carloni P & Parrinello M (2010) Targeting biomolecular flexibility with metadynamics. *Curr Opin Struct Biol* **20**, 148-154.
215. Huo... S (1997) The MaxFlux algorithm for calculating variationally optimized reaction paths for conformational transitions in many body systems at finite temperature. *Journal of Chemical Physics*.
216. Bolhuis PG, Chandler D, Dellago C & Geissler PL (2002) Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu Rev Phys Chem* **53**, 291-318.
217. Maragliano L, Fischer A, Vanden-Eijnden E & Ciccotti G (2006) String method in collective variables: minimum free energy paths and isocommittor surfaces. *J Chem Phys* **125**, 24106.
218. Kumar S, Rosenberg J & Bouzida D (1995) Multidimensional free - energy calculations using the weighted histogram analysis method. *J Comput Chem* **16**, 1339-1350.
219. Marinelli F, Pietrucci F, Laio A & Piana S (2009) A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations. *PLoS Comput Biol* **5**, e1000452.
220. Duan Y, Lee MC, Xiong G, Wu C, Chowdhury S, Luo R, Yang R, Cieplak P, Wang C & Kollman P (2003) A point - charge force field for molecular mechanics simulations of proteins based on condensed - phase quantum mechanical calculations. *Journal of Computational Chemistry* **24**, 1999-2012.

## BIBLIOGRAFÍA

221. Lee, M (2004) Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent. *Proteins: Structure, Function and Genomics*.
222. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L & Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802.
223. Inc AS (2009) Discovery Studio Modeling Environment. In, San Diego.
224. Hess B, Kutzner C, Van Der Spoel D & Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* **4**, 437-447.
225. Bonomi M, Branduardi D, Bussi G, Pietrucci F & Parrinello M (2009) PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer physics communications* **180**, 1961-1972.
226. Kumar S, Rosenberg J & Bouzida D (1992) The weighted histogram analysis method for free - energy calculations on biomolecules. I. The method. *J comp chem* **13**.
227. Souaille M & Roux B (2001) Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer physics communications* **135**, 40-57.

APÉNDICES





### 9.1 Clonado, expresión y purificación de los constructos del tándem cAbl

Las construcciones SH32 y SH32-C2Q han sido expresadas en células *E. Coli BL21-DE3* (NOVAGEN). A la secuencia de dichos constructos se le añadió una cola de histidinas (6xHis tag) en el extremo N-terminal, seguido de un sitio de reconocimiento para la proteasa TEV lo que permite eliminar la cola de histidinas del dominio tras su purificación. Dicha expresión fue inducida con 1mM de IPTG cuando la absorbancia del cultivo celular a 600 nm alcanzó valores entre 0.6 y 0.8 y se dejó crecer durante toda la noche a 37° C. Las construcciones fueron purificadas mediante cromatografía de afinidad Ni-NTA, proceso que se describe brevemente a continuación.

Las células fueron recogidas por centrifugación a 7000 rpm y 4° C durante 10 minutos (centrífuga Hettich modelo Roto Super 40, rotor A6.9) y posteriormente resuspendidas en tampón fosfato sódico 50 mM, cloruro sódico 0.3 M, pH 8.0 (Tampón de Equilibrado de Columna: TEC); después fueron lisadas mediante prensa French (AMICON) a una presión de 1200 psi y posteriormente ultracentrifugadas (BECKMAN Optima LE-80, rotor 45 Ti) a 30.000 rpm y 4 ° C durante 30 minutos. El sobrenadante resultante se pasó por una columna de afinidad con 5 mL de resina Ni-NTA (QUIAGEN) previamente equilibrada con 50 mL de TEC y se llevaron a cabo sucesivos lavados de 50mL con TEC, TEC + 20mM de imidazol y TEC + 50 mM de imidazol. Por último, la proteína fue eluída con 50 mL de TEC + 500 mM de imidazol. Las fracciones que contenían la proteína fueron dializadas frente al buffer de corte de TEV (50mM Tris, 100mM EDTA, 1mM DTT). A continuación se llevó a cabo el corte de la cola de histidinas mediante incubación con TEV a 1mg/mL a temperatura ambiente durante toda la noche. Posteriormente se

dializaron las proteínas en tampón TEC y se sometieron a una segunda etapa de cromatografía de afinidad en Ni-NTA, de modo que la cola de histidinas y la proteína no hidrolizada quedaban retenidas en la columna, mientras que la proteína sin cola de histidinas se eluía en las primeras fracciones, las cuales se concentraron hasta 1 mg·mL<sup>-1</sup> y se almacenaron a -20° C en el mismo tampón.

Las diferentes etapas de la purificación, así como la pureza final, fueron controladas mediante electroforesis en gel de acrilamida (SDS-page) al 16%. Además, la pureza de la proteína (superior al 99%) y su correcto peso molecular fueron comprobados mediante espectrometría de masas.

### 9.1.1 Mutantes

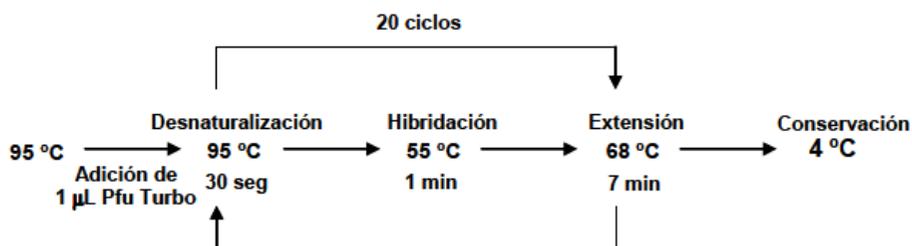
Para la obtención de los mutantes, SH32-C2Q-S140P y SH32-C2Q-3GLY, se utilizó la solución comercial de mutagénesis dirigida *Quikchange Site-Directed Mutagenesis Kit (STRATAGENE)*, que permite realizar mutaciones, tanto puntuales como múltiples, en un plásmido de cadena doble de forma sencilla y con más de un 80% de eficacia. Los oligonucleótidos cebadores se diseñaron de acuerdo con la mutación deseada de forma que fueran complementarios entre sí y con la cadena de doble plásmido, y con longitudes aproximadas de 30 a 40 bases, de modo que su temperatura de fusión ( $T_f$ ) oscilará entre 73° C y 78° C ya que elevadas  $T_f$  favorecen altas temperaturas de anillamiento ( $T_a$ ). La  $T_f$  de los oligonucleótidos en este caso se estimó de acuerdo con la siguiente ecuación empírica:

$$T_f = 81.5 + 0.41 \cdot \left( \% \frac{G}{C} \right) - \frac{675}{N} - (\% \text{ bases mutadas en el oligo})$$

donde N es el número total de bases del oligonucleótido y  $\% \frac{G}{C}$  es el tanto por ciento de Guaninas y Citosinas. Además, se procuró que el triplete

## APÉNDICES

que codifica la mutación estuviese siempre lo más centrado posible en la secuencia del oligonucleótido y que correspondiera con el triplete con mayor propensión en *E. Coli*. La PCR fue diseñada según las recomendaciones del método *Quikchange Site-Directed Mutagenesis Kit* (STRATAGENE), que se ilustra en la Figura 2.1.



**Figura 9.1.** Esquema general de los ciclos programados para la realización de las PCRs llevadas a cabo en esta Memoria.

La purificación de los mutantes procedentes del tándem SH32-C2Q se llevó a cabo siguiendo el protocolo anteriormente descrito para el dominio silvestre.

## 9.2 Calorimetría diferencial de barrido

Las proteínas fueron dializadas en un volumen de 2 L en tampón Pipes 50mM pH7.0. La concentración fue medida a una longitud de onda de absorción de 280nm usando un coeficiente de extinción de 39880 M<sup>-1</sup> cm<sup>-1</sup> para las construcciones SH32-C2Q y los mutantes de Glicina y Prolina y de 31400 M<sup>-1</sup>cm<sup>-1</sup> para el tándem SH32. Para los dominios independientes el coeficiente de extinción es 16894 M<sup>-1</sup> cm<sup>-1</sup> y 17420 M<sup>-1</sup> cm<sup>-1</sup> para el dominio SH3-cAbl y SH2-cAbl respectivamente. Los experimentos de DSC fueron realizados en un microcalorímetro automático VP-DSC (*Capillary cell microcalorimeter de MICROCAL INC.*) usando una velocidad de barrido de 90°

C/h. Las muestras fueron calentadas de 5° C a 100° C para comprobar la reversibilidad del proceso de desplegamiento. La concentración de las muestras de proteína fue de unos 0.6mg/mL. Cada termograma fue corregido con el tiempo de respuesta del calorímetro y con la línea base instrumental obtenida con el correspondiente tampón de diálisis. Después de la normalización con la concentración de la proteína, la capacidad calorífica,  $C_p$  (T), fue calculada con el resultado del termograma asumiendo que el volumen específico parcial es de 0.73 mL/g. Los cálculos fueron hechos usando el programa Origin.

### 9.3 Superfices de energía libre

Con el objetivo de explorar el equilibrio conformacional del tándem SH32 y los constructos SH32-3GLY, Ncap-SH32 y SH32-C2Q, las simulaciones de metadinámica fueron sesgadas en 5 coordenadas de reacción; las tres primeras, S1, S2, S3, son *Path Collective Variables* (PCV) y las dos restantes, D1 y D2, son ángulos diedros.

Como esta detallado en la Sección 5.6.1 las coordenadas de reacción de camino (PVC) nos indican en qué punto se encuentra una configuración dada en el espacio que conecta dos formas A y B definidas con anterioridad. Obteniendo un valor que indicara a cuál de estas dos formas se encuentra más próxima una estructura dada. En este estudio, las coordenadas de reacción S1 y S3 se aproximarán a 1 cuando las estructuras se encuentren cerca de la conformación abierta y 2 en el caso opuesto. Tanto S1 como S3 usaron como referencia la conformación abierta (2ABL) y la cerrada (2F00), pero mientras en S1 evaluamos el RMSD de todo el tándem, en S3 solamente se calculó el RMSD de la región de conexión C3-2, comprendida entre los residuos (238 a

## APÉNDICES

253). Por otro lado, la coordenada de reacción S2 es complementaria a las anteriores, siendo su misión la de impedir la deriva excesiva de las estructuras en el espacio del RMSD que separa ambas estructuras finales, y nos indica la distancia a la que se encuentra una configuración dada del camino que conecta ambos estados finales.

Como esta detallado en la Sección 4.2.2, las coordenadas de reacción D1 y D2 son dos ángulos diedros constituidos por los centro de masas de restos que se encuentran el conector C3-2 y sus proximidades. Estas coordenadas de reacción describen la orientación relativa entre los dos dominios.

En la Figura 9.2 se muestra la superficie bidimensional de energía libre reconstruida en las coordenadas de reacción S1 y S2 para todos los constructos. En la Figura 9.3 se muestra la misma reconstrucción, pero para las coordenadas de reacción S1 y S3. Finalmente en la Figura 9.4 se muestra la superficie de energía libre, proyectada en las coordenadas de reacción D1 y D2.

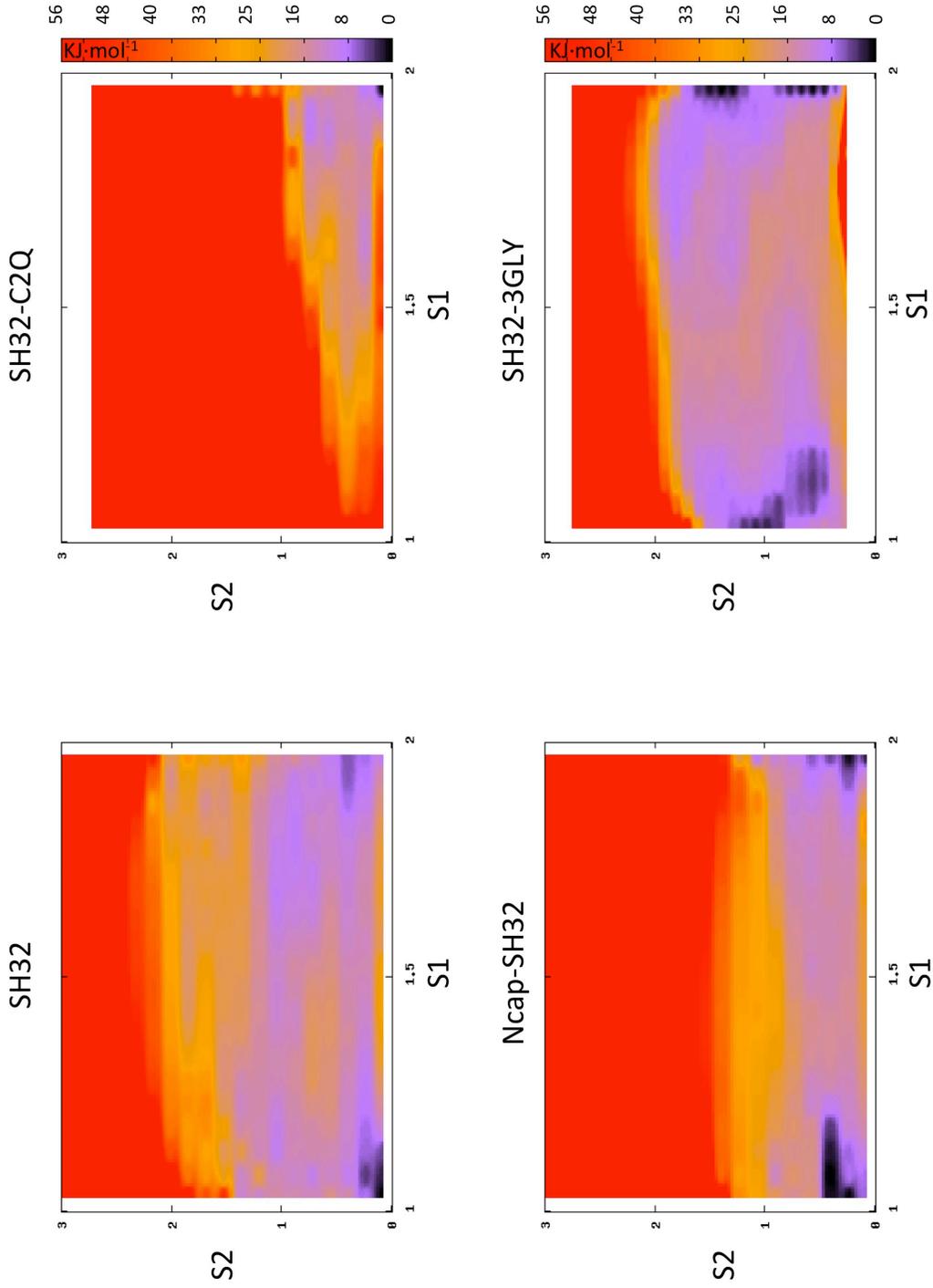


Figura 9.2 Superficie de energía libre reconstruida en función de las coordenadas de reacción  $S_1$  y  $S_2$ .

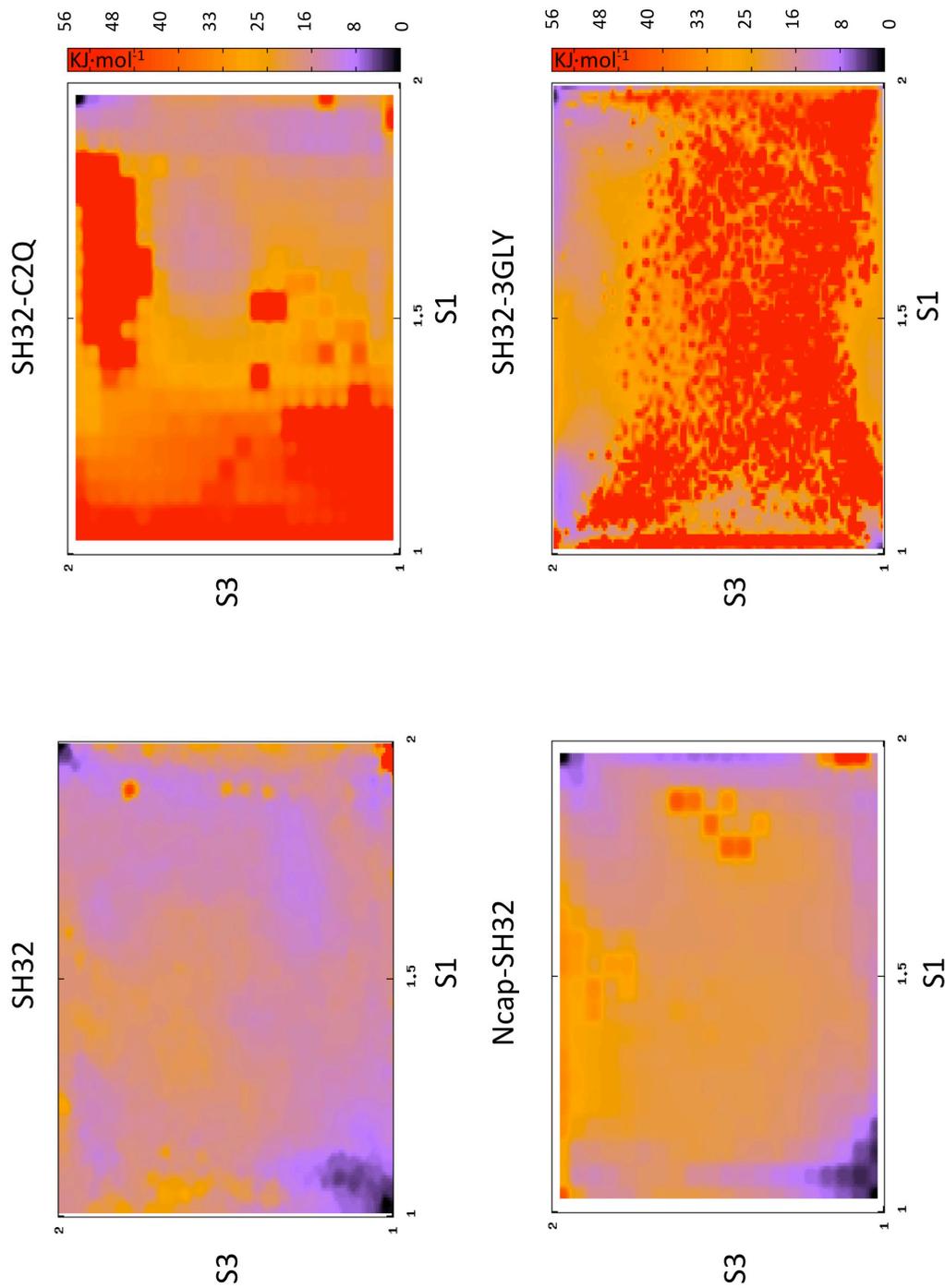


Figura 9.3 Superficie de energía libre reconstruida en función de las coordenadas de reacción S1 y S3.

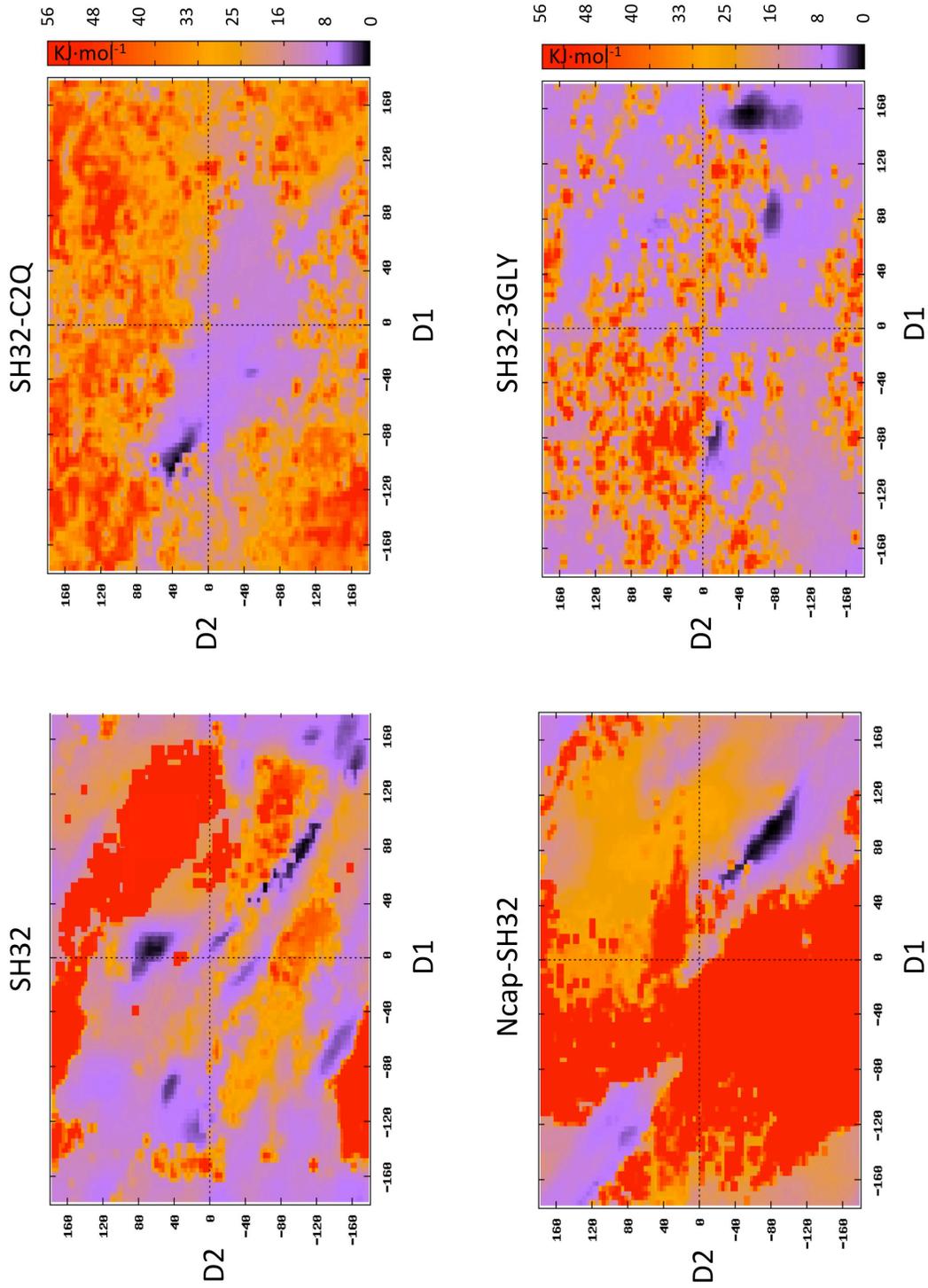


Figura 9.4 Superficie de energía libre reconstruida en función de las coordenadas de reacción  $D1$  y  $D2$ .

APÉNDICES

