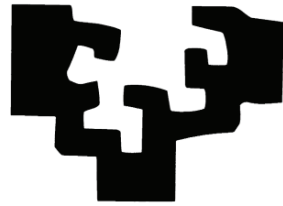


eman ta zabal zazu



Universidad del  
País Vasco

Euskal Herriko  
Unibertsitatea

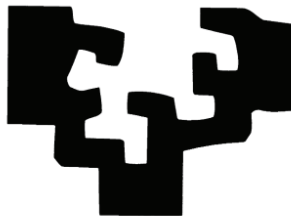
# CARACTERIZACIÓN ESTRUCTURAL DEL SUPRESOR TUMORAL ING5

**Jhon Alexander Rodríguez Buitrago**

**TESIS DOCTORAL. Leioa, 2014**



eman ta zabal zazu



Universidad del  
País Vasco

Euskal Herriko  
Unibertsitatea

# CARACTERIZACIÓN ESTRUCTURAL DEL SUPRESOR TUMORAL ING5

**Departamento de Bioquímica y Biología Molecular  
Facultad de Ciencia y Tecnología  
Universidad del País Vasco/Euskal Herriko Unibertsitatea  
Unidad de Biofísica (CSIC-UPV/EHU)**

**Memoria presentada por  
Jhon Alexander Rodriguez Buitrago  
para optar al grado de Doctor  
por la Universidad del País Vasco/Euskal Herriko Unibertsitatea**

**Trabajo dirigido por  
Dr. Francisco José Blanco Gutiérrez**





## **AGRADECIMIENTOS**

Me gustaría agradecer al Dr. Francisco Blanco por darme la oportunidad de trabajar bajo su dirección en el proyecto de mi tesis doctoral, por transmitirme su vocación investigadora y orientarme constantemente en todos los aspectos de la tesis. A los compañeros del laboratorio Nekane, Maider y Georgina por su colaboración y consejos durante mi estancia en CICbioGUNE. Igualmente a Alain y Alfredo por guiarme en el fascinante mundo de la RMN, por su colaboración en los experimentos y darme sus puntos de vista sobre el trabajo científico. Quiero agradecer a Adriana por introducirme en la cristalografía de proteínas, por el esfuerzo y sus ganas de enseñarme siempre algo nuevo acerca del tema. En general quiero darle las gracias a todo el gran equipo del CICbioGUNE por hacer de mi estancia algo inolvidable: a Félix por hacerme republicano, a Ibon por darme siempre los valores de masas que le solicitaba, a Iraide por su amabilidad, Anita, Igor, Guillermo, Sandra, Paula, Ander, Iker, June, Maria, Fredj, Idoia, Shiva, Ben y sus clases de ingles Britanico, a los de la vieja guardia, Xavi, David, Melisa, Marina, el Gabo, Liza, Valeria, Rafael, Eugenia, Mis compañeros de la UIS: Porras, Ludwin, El mago, javi jerez JJ, Adriancho, El capitán, Quechito, la Nonis, el cuentero del Mendez, los aprecio mucho. Por último pero no menos importante a Daniel el ufólogo por su amistad y estar en el momento justo y al “Master” Pietro Roversi por ser más que un instructor mi hermano.



*A mi abuela Betzabe por su amor constante e inagotable, por  
todo el esfuerzo realizado para que yo estudiara.  
A mi esposa Adriana, por soñar conmigo, ser mi apoyo y darme tanta  
felicidad.*



# ÍNDICE GENERAL

Página

<b>1</b>	<b><u>INTRODUCCIÓN</u></b>	<b><u>1</u></b>
<b>1.1</b>	<b>LA FAMILIA DE SUPRESORES DE TUMORES ING.....</b>	<b>3</b>
1.1.1	ESTRUCTURA DE LAS PROTEÍNAS ING .....	4
1.1.2	DOMINIO N-TERMINAL.....	6
1.1.3	REGIÓN CENTRAL (NLS) .....	7
1.1.4	DOMINIO C-TERMINAL (PHD) .....	9
<b>1.2</b>	<b>OTROS MOTIVOS DE SECUENCIA EN LAS PROTEÍNAS ING.....</b>	<b>10</b>
1.1.5	MOTIVO DE INTERACCIÓN CON PCNA, CAJA PIP .....	10
1.1.6	MOTIVO DE INTERACCIÓN CON LA PROTEÍNA SAP30 (SAID) .....	11
1.1.7	DOMINIO DE INTERACCIÓN CON LAMÍN (LID) .....	11
1.1.8	REGIÓN POLIBÁSICA (PBR) .....	12
1.1.9	INTERACCIÓN CON OTRAS PROTEÍNAS .....	12
<b>1.3</b>	<b>EL CÓDIGO DE HISTONAS Y LA EXPRESIÓN GÉNICA.....</b>	<b>15</b>
<b>1.4</b>	<b>RECONOCIMIENTO DE H3K4ME3 POR PARTE DE LAS PROTEÍNAS ING .....</b>	<b>18</b>
1.1.10	COMPLEJOS DE ACETILACIÓN/DESACETILACIÓN DE HISTONAS EN LOS QUE PARTICIPAN LAS PROTEÍNAS ING....	21
1.1.11	COMPLEJOS DE ACETILACIÓN DE HISTONAS .....	21
1.1.12	COMPLEJO HAT MOZ/MORF .....	22
1.1.13	COMPLEJO HAT HBO1 .....	22
1.1.14	LA PROTEÍNA ING5 .....	24
<b>2</b>	<b><u>OBJETIVOS</u></b>	<b><u>27</u></b>
<b>3</b>	<b><u>MATERIALES Y MÉTODOS</u></b>	<b><u>31</u></b>
<b>1.5</b>	<b>CLONES UTILIZADOS Y MANIPULACIONES MEDIANTE TÉCNICAS DE BIOLOGÍA MOLECULAR .....</b>	<b>33</b>
1.1.15	CLONES DE ING5 .....	33
1.1.16	OTROS CLONES .....	36



<b>1.6</b>	<b>EXPRESIÓN GÉNICA Y PRODUCCIÓN DE LAS PROTEÍNAS.</b>	<b>37</b>
1.1.17	TRANSFORMACIÓN Y PRUEBAS DE EXPRESIÓN DE PROTEÍNAS EN BACTERIAS.	37
1.1.18	PRODUCCIÓN DE PROTEÍNAS EN SISTEMAS LIBRES DE CÉLULAS.	38
1.1.19	EXPRESIÓN DE ING5 COMPLETA.	39
1.1.20	EXPRESIÓN DE ING5 <sub>1-105</sub> Y MUTANTES PUNTUALES Q33R, I68V Y C75R	39
1.1.21	EXPRESIÓN DE ING5 <sub>1-105</sub> EN MEDIO SUPLEMENTADO CON SELENOMETIONINA	40
1.1.22	EXPRESIÓN DE ING5 Y DE ING5 <sub>1-105</sub> MARCADA ISOTÓPICAMENTE EN <sup>15</sup> N	40
1.1.23	EXPRESIÓN DE ING5 <sub>1-105</sub> MARCADA ISOTÓPICAMENTE EN <sup>2</sup> H, <sup>13</sup> C, <sup>15</sup> N	41
1.1.24	EXPRESIÓN DE GST-JADE1L_DOMINIO IIb	42
1.1.25	EXPRESIÓN DE LA PROTEASA TEV ( <i>TOBACCO ETCH VIRUS</i> )	42
<b>1.7</b>	<b>PURIFICACIÓN DE PROTEÍNAS</b>	<b>43</b>
1.1.26	PURIFICACIÓN DE LA PROTEASA TEV ( <i>TOBACCO ETCH VIRUS</i> )	44
1.1.27	PURIFICACIÓN DE ING5 COMPLETA	44
1.1.28	PURIFICACIÓN DE ING5 <sub>1-105</sub> Y MUTANTES Q33R, I68V, C75R	47
1.1.29	PURIFICACIÓN DE ING4 COMPLETA/ING4 <sub>2-107</sub>	48
1.1.30	PURIFICACIÓN DE GST-JADE1L_DOMINIO IIb	49
<b>1.8</b>	<b>FORMACIÓN DE HETERODÍMEROS ING4/ING5</b>	<b>51</b>
<b>1.9</b>	<b>CUANTIFICACIÓN DE PROTEÍNAS EN DISOLUCIÓN POR ABSORBANCIA ULTRAVIOLETA.</b>	<b>52</b>
<b>1.10</b>	<b>ESPECTROSCOPIA DE DICROÍSMO CIRCULAR.</b>	<b>52</b>
<b>1.11</b>	<b>DISPERSIÓN DE LUZ A MÚLTIPLES ÁNGULOS MALLS (MULTI-ANGLE LASER LIGHT SCATTERING)</b>	<b>54</b>
<b>1.12</b>	<b>CRISTALES DE PROTEÍNAS.</b>	<b>55</b>
1.1.31	PRINCIPIOS BÁSICOS DE LA CRISTALIZACIÓN.	55
1.1.32	TÉCNICAS DE CRISTALIZACIÓN.	57
<b>1.13</b>	<b>CRISTALIZACIÓN DE ING5<sub>1-105</sub></b>	<b>58</b>
1.1.33	CRIOPROTECCIÓN DE LOS CRISTALES.	61
<b>1.14</b>	<b>DIFRACCIÓN DE RAYOS X</b>	<b>61</b>
1.1.34	COLECCIÓN DE DATOS CRISTALOGRAFICOS.	63
<b>1.15</b>	<b>ANÁLISIS DE LOS DATOS</b>	<b>63</b>
1.1.35	INDEXACIÓN, ESCALADO E INTEGRACIÓN DE DATOS.	63

<b>1.16</b>	<b>EL PROBLEMA DE LA FASE .....</b>	<b>65</b>
1.17	<b>COLECCIÓN Y ANÁLISIS DE LOS DATOS DE DIFRACCIÓN DE CRISTALES DE ING5<sub>1-105</sub> .....</b>	<b>66</b>
<b>1.19</b>	<b>DETERMINACIÓN DEL GRUPO ESPACIAL .....</b>	<b>69</b>
<b>1.20</b>	<b>REEMPLAZAMIENTO MOLECULAR .....</b>	<b>71</b>
1.1.36	CONSTRUCCIÓN DEL MODELO USADO PARA EL REEMPLAZAMIENTO MOLECULAR .....	71
1.1.37	ESTRATEGIA DE REEMPLAZAMIENTO MOLECULAR .....	72
1.1.38	CONSTRUCCIÓN, REFINAMIENTO Y VALIDACIÓN DEL MODELO DE ING5 <sub>1-105</sub> .....	73
<b>1.21</b>	<b>ANÁLISIS DE LA ESTRUCTURA CRISTALOGRÁFICA DE ING5<sub>1-105</sub> .....</b>	<b>74</b>
<b>1.22</b>	<b>COMPLEMENTARIEDAD DE LA RESONANCIA MAGNÉTICA NUCLEAR Y CRISTALOGRAFÍA APLICADAS AL ESTUDIO ESTRUCTURAL DE PROTEÍNAS .....</b>	<b>75</b>
<b>1.23</b>	<b>RESONANCIA MAGNÉTICA NUCLEAR DE PROTEÍNAS. ....</b>	<b>77</b>
1.1.39	EL EXPERIMENTO <sup>1</sup> H- <sup>15</sup> N-HSQC Y LA ESPECTROSCOPIA TROSY .....	79
1.1.40	EXPERIMENTOS DE TRIPLE RESONANCIA PARA LA ASIGNACIÓN DEL ESPECTRO DE RMN DE ING5 <sub>1-105</sub> .....	82
<b>1.24</b>	<b>ANÁLISIS DE LOS ESPECTROS DE RMN.....</b>	<b>84</b>
<b>1.25</b>	<b>PREPARACIÓN DE MUESTRAS PARAMAGNÉTICAS DE ING5<sub>1-105</sub> .....</b>	<b>86</b>
<b>1.26</b>	<b>ANÁLISIS DE INTERACCIONES DE ING5 MEDIANTE RMN.....</b>	<b>87</b>

#### **4 RESULTADOS 91**

---

<b>1.27</b>	<b>OBTENCIÓN DE PROTEÍNAS ING ADECUADAS PARA ESTUDIOS ESTRUCTURALES .....</b>	<b>93</b>
<b>1.28</b>	<b>CARACTERIZACIÓN ESTRUCTURAL DE ING5 COMPLETA .....</b>	<b>94</b>
<b>1.29</b>	<b>HETERODIMERIZACIÓN DE ING4 E ING5 .....</b>	<b>97</b>
<b>1.30</b>	<b>ANÁLISIS ESTRUCTURAL DEL DOMINIO N-TERMINAL DE ING5 (ING5<sub>1-105</sub>) .....</b>	<b>98</b>
<b>1.31</b>	<b>HETERODIMERIZACIÓN DE STREP-ING4<sub>2-107</sub> /HIS-ING5<sub>1-105</sub> ANÁLISIS DE MUTACIONES DE ING5<sub>1-105</sub> DETECTADAS EN CÁNCER Y SU EFECTO EN LA DIMERIZACIÓN. ....</b>	<b>101</b>
<b>1.32</b>	<b>ANÁLISIS DE MUTACIONES DE ING5<sub>1-105</sub> DETECTADAS EN CÁNCER Y SU EFECTO EN LA DIMERIZACIÓN.....</b>	<b>101</b>
1.33	<b>ESTRUCTURA CRISTALOGRÁFICA DE ING5<sub>1-105</sub> .....</b>	<b>103</b>
1.1.41	DESCRIPCIÓN DE LAS CADENAS A Y B DE ING5 <sub>1-105</sub> .....	107
1.34	<b>DESCRIPCIÓN Y ANÁLISIS DE LA INTERFAZ DEL HOMODÍMERO DE ING5<sub>1-105</sub> .....</b>	<b>111</b>

1.35	ASIMETRÍA ESTRUCTURAL DEL DÍMERO DE ING5 <sub>1-105</sub> .....	114
1.36	EFFECTO DE LA RELAJACIÓN PARAMAGNÉTICA EN EL ESPECTRO <sup>1</sup> H- <sup>15</sup> N- HSQC DE ING5 <sub>1-105</sub> MARCADO CON MTSL EN C19 Y C75. ....	121
1.37	EFFECTO DE DE LAS MUTACIONES PUNTUALES EN LA ESTABILIDAD DE LA ESTRUCTURA DE ING5 <sub>1-105</sub> .....	122
1.38	INTERACCIÓN ING5 <sub>1-105</sub> Y GST-JADE1L_DOMINIO IIb EXPERIMENTO <sup>1</sup> H- <sup>15</sup> N-HSQC .....	123
1.39	INTERACCIÓN DE ING5 CON EL PÉPTIDO DE HISTONA H3K4ME3 .....	124

<u>5</u>	<u>DISCUSIÓN</u>	<u>131</u>
<u>6</u>	<u>CONCLUSIONES</u>	<u>147</u>
<u>7</u>	<u>REFERENCIAS</u>	<u>151</u>
<u>8</u>	<u>ANEXOS</u>	<u>169</u>

1.40	ANEXO 1 SECUENCIA DEL GEN SINTÉTICO DE ING5 CON CODONES OPTIMIZADOS PARA SU EXPRESIÓN EN E. COLI. 171	
1.41	ANEXO 2 DESPLAZAMIENTOS QUÍMICOS DE LOS NÚCLEOS ASIGNADOS DE LA CADENA PRINCIPAL Y LOS <sup>13</sup> C <sub>B</sub> DE ING5 <sub>1-105</sub> .....	172
1.42	ANEXO 3. ÁNGULOS DIEDROS DEL ESQUELETO DE ING5 <sub>1-105</sub> DERIVADOS DE TALOS+.....	175
1.43	ANEXO 4 CONSTANTES DE DISOCIACIÓN (K <sub>D</sub> ) DE ING5 POR EL PÉPTIDO DE HISTONA H3K4ME3 CALCULADAS A PARTIR DE LA TITULACIÓN POR RMN.....	179
1.44	ANEXO 5 CONSTANTES DE DISOCIACIÓN (K <sub>D</sub> ) DEL DOMINIO PHD DE ING5 POR EL PÉPTIDO DE HISTONA H3K4ME3 CALCULADAS A PARTIR DE LA TITULACIÓN POR RMN. . .	180
1.45	ANEXO 6 CONSTANTES DE DISOCIACIÓN (K <sub>D</sub> ) DEL DOMINIO PHD DE ING4 POR EL PÉPTIDO DE HISTONA H3K4ME3 CALCULADAS A PARTIR DE LA TITULACIÓN POR RMN. . .	182

<u>9</u>	<u>ABREVIATURAS</u>	<u>185</u>
<u>10</u>	<u>ÍNDICES DE FIGURAS Y TABLAS</u>	<u>191</u>

10.1	ÍNDICE DE FIGURAS .....	193
10.2	ÍNDICE DE TABLAS .....	195



# **1 Introducción**





### 1.1 La familia de supresores de tumores ING

Los genes de la familia de ING (*INhibitors of Growth*) se clasifican como supresores de tumores en base a estudios donde se muestra la existencia de pérdida, desregulación y disminución de expresión en diferentes tipos de cáncer (Piche & Li, 2010). En varios estudios se ha propuesto que todos los miembros de la familia ING cooperan con p53 para inducir apoptosis y senescencia (Cheung & Li, 2002; Jafarnejad & Li, ; Nagashima et al, 2003; Shiseki et al, 2003). Esta familia está formada por cinco genes, que mediante diversos procesos de empalme alternativo (*alternative splicing*), pueden dar lugar a al menos 17 isoformas de las correspondientes proteínas. En general las INGs actúan en la regulación de la expresión génica modulada por factores epigenéticos, y de esta forma están implicadas directa o indirectamente, en reparación del ADN (Cheung & Li, 2002; Garkavtsev et al, 1996; Scott et al, 2001b), remodelación de la cromatina (Feng et al, 2002; Vieyra et al, 2002), control del ciclo celular (Abad et al, 2007; Campos et al, 2004a), angiogénesis (Garkavtsev et al, 2004), inducción de senescencia celular (Garkavtsev et al, 1998; Garkavtsev et al, 1996; Ythier et al, 2008) y apoptosis (Helbing et al, 1997). Los genes INGs están presentes en una gran variedad de organismos que abarcan desde hongos hasta mamíferos (Russell et al, 2006; Soliman & Riabowol, 2007). De los análisis filogenéticos de la secuencias de las proteínas ING se infiere que su distribución está restringida a eucariotas del grupo unikonta (células con un solo flagelo). El primer miembro de la familia ING en ser descubierto fue ING1, mediante un experimento diseñado para identificar supresores tumorales expresados diferencialmente en células del epitelio mamario en comparación con líneas celulares de cáncer de mama. El gen que codifica para ING1 fue clonado y sobreexpresado observándose su capacidad para inhibir el crecimiento celular, mientras que la expresión de un ARN antisentido de ING1 promovía la transformación celular, con lo cual se propuso que ING1 era un supresor tumoral (Garkavtsev et al, 1996). Los otros miembros de la familia (ING2-5) fueron descubiertos mediante búsqueda basada en homología en secuencia con ING1 (Soliman & Riabowol, 2007). Los genes supresores de tumores son generalmente recesivos y es necesario que los dos alelos estén alterados para desactivarlos y que pueda progresar el tumor. En el contexto de la progresión del cáncer, la causa más frecuente de su desactivación parece ser la

disminución o pérdida total de su expresión en células cancerosas (Ythier et al, 2008). El hecho de que los genes de ING1,2,4 y 5 estén localizados en los extremos de los cromosomas donde se encuentran, muy cerca de las regiones teloméricas, sugiere la posibilidad de que la alteración de sus niveles de expresión observada en muchos tipos de cáncer se deba a la erosión del telómero que tiene lugar con el envejecimiento celular y durante la tumorigénesis (He et al, 2005). El análisis filogenético de las secuencias de la familia ING, permite agrupar sus miembros en tres subfamilias: ING3, ING1-2, e ING4-5 (Figura 8). ING3 muestra la mayor divergencia de secuencia entre los miembros de la familia, la ubicación central de ING3 en el cromosoma donde se encuentra es diferente a la de los otros miembros de la familia, los cuales se ubican en la región telomérica. Dentro del contexto filogenético de las proteínas INGs solo hay evidencia de genes parálogos en plantas en la subfamilia de proteínas ING3. Estas singularidades podrían evidenciar que el gen de ING3 es el más ancestral de la familia ING (He et al, 2005).

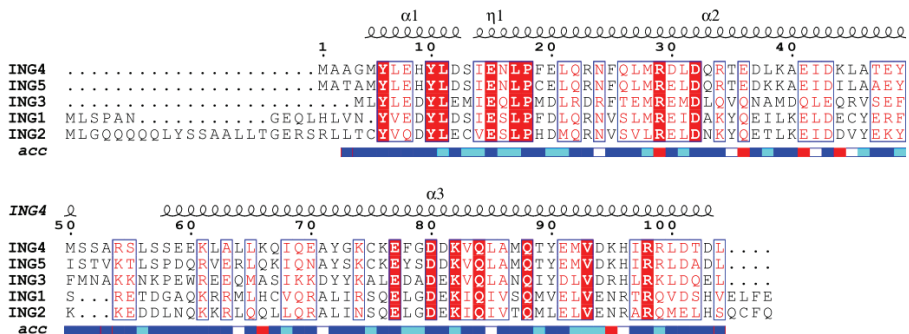
### **1.1.1 Estructura de las proteínas ING**

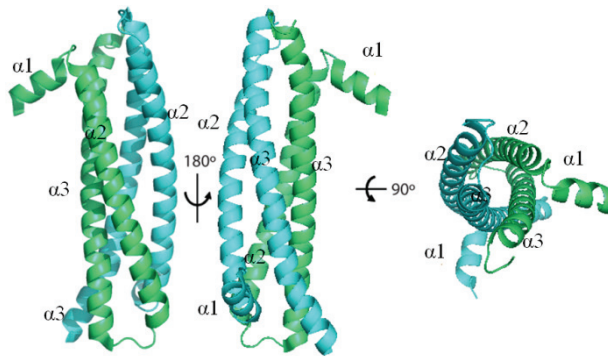
El alineamiento de las secuencias de aminoácidos de las proteínas ING revela la existencia de dos dominios conservados unidos por un dominio central no conservado. Los estudios estructurales experimentales disponibles de la proteína ING4 muestran que los dos dominios conservados en los extremos N- y C-terminal están estructurados mientras que la región central no lo está y se comporta como una cadena flexible y desordenada (Palacios et al, 2010).



### 1.1.2 Dominio N-terminal

El análisis de las secuencia N-terminales de las proteínas ING sugería que este dominio podría adoptar una estructura de hélices enrolladas (*coiled-coil*). Esta predicción se basaba en la presencia de residuos de leucina/isoleucina espaciados cada siete residuos en la región LZL, y que se encuentra frecuentemente en estructuras de hélices enrolladas. Además se sugirió que esto podría dar lugar a homodímeros e incluso heterodímeros (He et al, 2005). Más tarde, mediante RMN y otras técnicas, se comprobó experimentalmente que el dominio N-terminal de ING4 adopta una estructura dimérica de hélices enrolladas (Palacios et al, 2010). La estructura del fragmento ING4<sub>2-105</sub> determinada mediante cristalografía de rayos X código PDB:4AFL (Figura 2B), demostró que esta región se pliega en un homodímero con ensamblaje antiparalelo donde cada protómero forma una estructura elongada de tipo hélice-bucle-hélice más una hélice N-terminal de corta longitud (Culurgioni et al, 2012). En la estructura de este dominio de ING4 los residuos de leucina e isoleucina forman una estructura típica de cremallera de leucinas que contribuyen al empaquetamiento de las cadenas y a la formación del núcleo hidrofóbico de la estructura que resulta estabilizada por la formación de pares iónicos y enlaces de hidrógeno intermoleculares.





**Figura 2** (Arriba) Alineamiento de las secuencias N-terminales de las proteínas de la familia ING usando la numeración de ING4 y con un esquema de los elementos de estructura secundaria determinados en la estructura cristalográfica del fragmento 2-105 de ING4. La línea “acc” indica la accesibilidad al solvente de los residuos considerando la estructura de un protómero (azul indica residuo accesible, cian de accesibilidad intermedia, y blanco enterrado). El alineamiento fue realizado con ESPript (Gouet et al, 1999) mediante el servidor: <http://endscript.ibcp.fr> (Debajo) Representación de la estructura cristalográfica de ING4<sub>2-105</sub> en distintas orientaciones con cada uno de los dos protómeros coloreados en cian y verde.

En base a esta información y al alto grado de homología en secuencia entre los dominios N-terminales de las proteínas INGS, se ha propuesto que la región N-terminal permite la homodimerización de los diferentes miembros de la familia ING (Culurgioni et al, 2012; He et al, 2005). Se ha reportado que el dominio N-terminal de las proteínas ING está implicado en varias de las interacciones proteína-proteína detectadas para las INGS, funcionando como un sitio de unión a distintos componentes nucleares y a complejos de remodelación de la cromatina (Aguissa-Toure et al, 2011). Un análisis estructural y funcional de todos los dominios N terminal de las proteínas ING permitiría entender mejor los mecanismos de acción en la supresión de tumores y a definir mejor las diferencias existentes entre sus miembros.

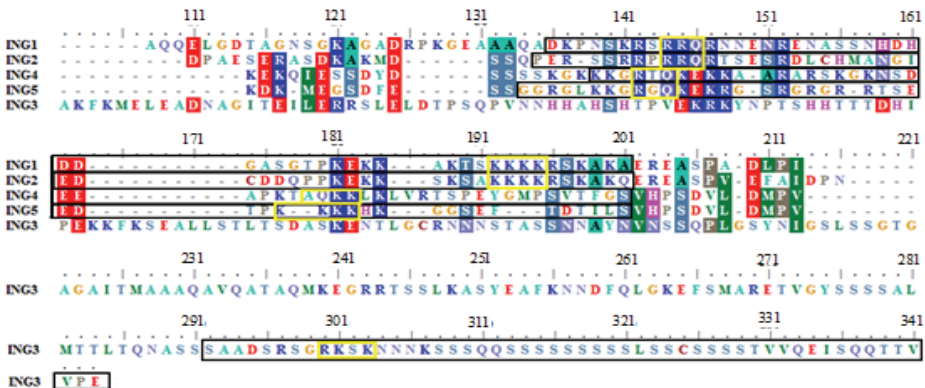
### 1.1.3 Región Central (NLS)

Esta región, que se encuentra desestructurada en ING4, según se ha observado por RMN (Palacios et al, 2010) y muy probablemente también lo esté en el resto de proteínas ING (Figura 3). Es la menos conservada, pero en todos los casos contiene la señal de



## Introducción

localización nuclear (NLS), y por eso se le suele denominar región NLS aunque sea bastante más larga que la secuencia NLS estricta. Esta región contiene además varias secuencias descritas como de translocación nucleolar (NTS, *nucleolar translocation sequence*). En el caso de ING1 se ha visto que estas NTSs dirigen la acumulación de proteína en el nucléolo en condiciones de estrés celular (Scott et al, 2001a). La NLS está involucrada en el transporte de las proteínas ING del citoplasma al núcleo, y es crucial para las funciones que las ING, desempeñan en el núcleo celular. En el caso de ING1 algunos tipos de cáncer (mama, ovario, gástrico y oral) están asociados a mutaciones ubicadas en los motivos de secuencia NLS/NTS (Gong et al, 2005). La delección de la NLS produce una acumulación de proteínas ING en el citoplasma (Ha et al, 2002). Pero también se ha descrito para ING4 una función en el citoplasma, donde se une a la proteína liprina  $\alpha 1$  precisamente a través de la región NLS, regulando la propagación y la migración celular (Shen et al, 2007). Se ha reportado que ING4 interacciona con p53 a través de la región NLS, según se observa en experimentos de *pull down* con fragmentos de la proteína ING4 fusionados a la proteína GST (Scott et al, 2001a). Sin embargo, en un estudio posterior por RMN con las proteínas en disolución no se pudo confirmar esta interacción, indicando que si existe es muy débil, o no es directa y está mediada por otros factores (Palacios et al, 2010).



**Figura 3** Alineamiento de secuencias de la región central de la familia de proteínas ING. Se observan las señales de localización nuclear de la familia ING. En cajas negras están las secuencias NLS, en las cajas amarillas las secuencias de localización nucleolar NTS. El alineamiento fue realizado con Clustal W 2.0. Los colores de los residuos indican: azul: con carga positiva, rojo: con carga negativa, verde: hidrofóbico, gris: polar no cargados. La numeración del alineamiento se basa en la secuencia de ING3.

### 1.1.4 Dominio C-terminal (PHD)

Todas las proteínas ING poseen un dominio PHD en la región C-terminal, el cual es muy común en proteínas reguladoras de la cromatina (Pascual et al, 2000) (Wysocka et al, 2006). El dominio PHD comprende una región de aproximadamente 60 aminoácidos y es el dominio evolutivamente más conservado entre los miembros de la familia ING, con un nivel de identidad mayor del 78% (Aguissa-Toure et al, 2011). Este dominio contiene un motivo de secuencia C<sub>4</sub>HC<sub>3</sub> (cuatro cisteínas, una histidina, tres cisteínas) el cual coordina dos cationes Zn<sup>+2</sup> manteniendo de esta forma la estabilidad del dominio (Champagne et al, 2008). La unión de los cationes de zinc es esencial para la integridad estructural del dominio que posee muy pocos elementos de estructura secundaria y varios bucles de longitud variable, que son importantes para la especificidad de reconocimiento de los diferentes PHD (Aasland et al, 1995; Bienz, 2006; Schindler et al, 1993). El dominio PHD de las proteínas ING reconoce y se une a la cola N-terminal de histona H3 con mayor o menor afinidad dependiendo del estado de metilación de la lisina K4 (no metilada, mono-, di- o trimetilada). En el caso de los PHDs de ING4 e ING5 la afinidad es mayor por la forma trimetilada hacia la marca de histona H3K4me<sub>3</sub>. Se han determinado por cristalografía las estructuras de los dominios PHD de todas las proteínas ING unidos a fragmentos de la cola N-terminal de histona H3 con la marca H3K4me<sub>3</sub>, salvo en el caso de ING3 que solo se ha determinado en su forma libre. En el dominio PHD de ING1 el péptido de histona H3K4me<sub>3</sub> encaja en un bolsillo hidrofóbico formado por los residuos aromáticos Y212 y W235 y las cadenas laterales de S219 y M223 (Peña et al, 2008). Análogamente en ING2 el bolsillo lo forman los residuos Y215, W238, S222 y M226 (Peña et al, 2006), en ING4 lo forman Y128, W221, S205 y M209 (Palacios et al, 2008), y en ING5 lo forman Y188, W211, S195 y M199 (Champagne et al, 2008). La estructura de los cinco dominios es muy similar (Figura 4), con un RMSD de 0.13 Å para los átomos C $\alpha$  de los 52 residuos estructurados y comunes a los cinco dominios superpuestos calculado en el programa Theseus 2.0.11 (Theobald & Wuttke, 2006). El correcto funcionamiento del dominio PHD ha demostrado ser crítico para la capacidad de algunas de las proteínas ING para actuar como supresores de tumores. Se han descrito mutaciones en el dominio PHD asociadas a



reclutamiento de enzimas de reparación del ADN en sitios dañados del ADN (Berardi et al, 2004). La secuencia canónica de la caja PIP es  $QXXhXXaa$ , donde  $h$  es un residuo hidrofóbico,  $a$  es un residuo aromático típicamente (F o Y) y  $X$  es cualquiera de los 20 aminoácidos proteinogénicos (Warbrick et al, 1998). El análisis de la interacción de PCNA con el fragmento 4-25 de ING1 mediante RMN en disolución muestra que la unión es extremadamente débil, lo cual probablemente se debe a que el motivo PIP (QLHLVNY) difiere bastante del canónico (De Biasio et al, 2012). Esto sugiere que la interacción observada en extractos celulares esta mediada por otros factores. Mediante experimentos tipo *pull-down* entre PCNA y diferentes formas truncadas de ING2 se ha descrito que ING2 también se une a PCNA aunque no tiene motivo PIP. Se determino que la región de interacción con PCNA es la que tiene la secuencia de tipo LID en la región N-terminal de ING2. La interacción PCNA/ING2 parece tener que ver más con la regulación de la cantidad de PCNA que se une al nucleosoma que en la reparación del ADN (Larrieu et al, 2009).

#### **1.1.6 Motivo de interacción con la proteína SAP30 (SAID)**

Este motivo está presente solo en la proteína ING1, también se conoce como dominio PBD debido a su homología de secuencia con los bromo-dominios (Aguissa-Toure et al, 2011). El motivo SAID se une a la proteína SAP30 del complejo de desacetilación de histonas mSin3A-HDAC1 implicando a ING1 en procesos de remodelación de la cromatina como desacetilación y movilización del nucleosoma (Kuzmichev et al, 2002).

#### **1.1.7 Dominio de interacción con Lamín (LID)**

El dominio LID es el segundo motivo más conservado en secuencia y se encuentra en todas las isoformas conocidas de la familia ING (He et al, 2005). Se ha demostrado recientemente que este dominio interactúa específicamente con la proteína nuclear lamina A y por ende esta región que antes se denominaba NCR (*Nueva Región Conservada*) pasó a denominarse LID (*Dominio de Interacción con Lamina*). La proteína lamina A, parece que estabiliza a las proteínas ING dentro del núcleo (Han et al, 2008). Además los motivos KIQI de (ING1/2) / KVQL (ING3/4/5), contenidos en el dominio LID de las INGs parecen estar

implicados en la unión de complejos de acetilación (HAT) y desacetilación (HDAC) de histonas, durante los proceso de regulación de la cromatina y regulación de la expresión génica (Gordon et al, 2008; Kuzmichev et al, 2002; Soliman & Riabowol, 2007).

### **1.1.8 Región polibásica (PBR)**

Tras el dominio PHD, en el extremo C-terminal de las INGs hay un motivo conservado llamado región polibásica (PBR), necesario para el reconocimiento de fosfatos de fosfatidilinositol (PtdInsPs). Este motivo solo está presente como tal en ING1 e ING2. Las correspondientes regiones en el resto de los miembros de la familia son también ricas en residuos básicos pero son considerablemente más cortas, y en el caso de ING4 e ING5 contienen el motivo de secuencia RKKK que podrían funcionar como señales adicionales de localización nuclear (Gozani et al, 2003). Los PtdInsPs son moléculas de señalización que regulan la supervivencia celular, el crecimiento y la proliferación (Cantley, 2002). Los últimos cuatro residuos de la región C-terminal de ING4 (RKKK) otorgan una superficie positiva que se ha propuesto puede ser determinante para la unión a PtdInsPs de forma similar a las regiones positivas encontradas en ING1 e ING2 y que son requeridas para la unión a PtdInsPs (Gozani et al, 2003). Sin embargo experimentos de RMN en solución en los cuales se intentó caracterizar la interacción del dominio PHD de ING4 a diferentes tipos de PtdInsPs muestran que ING4 no es un receptor de PtdInsPs o que si se une a estos lo hace de una forma muy débil (Palacios et al, 2006).

### **1.1.9 Interacción con otras proteínas**

Debido a que la familia de proteínas ING está formada por 17 isoformas y estructuralmente están constituidas por tres dominios, los cuales contienen varios motivos asociados a interacciones proteína-proteína, se han descrito un gran número de interacciones en las que participan proteínas de la familia ING. Una de las interacciones mas estudiadas es la que involucra la proteína p53. En la Tabla 1 se presentan los dominios de la familia ING y los complejos proteicos en los cuales se ha descrito que participan.

**Tabla 1** Interacciones proteína-proteínas descritas para las proteínas de la familia ING

Proteína	Dominio	Proteína/complejo involucrado	Método	Referencia
ING1/ Q9UK53	PHD	ARF	IP	(Gonzalez et al, 2006)
		Brg1, BAF47/53/60/155/170/250	IP, pd	(Kuzmichev et al, 2002)
		DMAP-1	pd, IP	(Xin et al, 2004)
		H3K4me2/3	pd, ChIP, NMR, fluo	(Shi et al, 2006; Peña et al, 2006; Peña et al, 2008)
		HDAC1	IP	(Vieyra et al, 2002)
		mSin3, HDAC1/2, SAP30, RbAp46/48	IP, pd	(Kuzmichev et al, 2002)
	N/D	p300	IP	(Vieyra et al, 2002)
		GADD45	IP	(Cheung et al, 2001)
		HMT	IP, pd	(Goeman et al, 2008)
		hSir2	IP	(Kataoka et al, 2003)
		RBP1	IP, MS	(Skowrya et al, 2001)
		SIRT1	IP, pd	(Binda et al, 2008)
		TRRAP, PCAF, CBP	IP	(Vieyra et al, 2002)
		p15 (PAF)	IP	(Simpson et al, 2006)
		p42, p35	IP, MS	(Skowrya et al, 2001)
	PIP	p53	IP	(Garkavtsev et al, 2004)
		PCNA	IP	(Garkavtsev et al, 2004)
		LID	Lámina A	pd, coIP, (LS/LM)
NLS	Karyopherin a,b	coIP,	(Russell et al, 2008)	
ING2/ Q9H160	PHD	H3K4me2/3	ChIP, pd, NMR, fluo	(Peña et al, 2006; Shi et al, 2006)
		mSin3A, HDAC1/2, RbAp46/48	IP	(Doyon et al, 2006)
		p300, p300/p53	IP	(Pedeux et al, 2005; Wang et al, 2006)
		PtdIns5P	IP	(Gozani et al, 2003)
		SNON	IP	(Sarker et al, 2008)
	N/D	RBPI/RBP1-like	IP	(Doyon et al, 2006)
		SAP30, SAP130, SDS3, BRMS1/BRMS1-like	ChIP	(Smith et al, 2010)
		HMT	IP, pd	(Goeman et al, 2008)
		BAF47/53a/155/170		(Doyon et al, 2006)
		SIRT1	IP, pd	(Binda et al, 2008)
NCR	PCNA	IP	(Larrieu et al, 2009)	
ING3/ Q9NXR	PHD	H3K4me2/3	ChIP, pd, fluo	(Peña et al, 2006; Shi et al, 2006)
		PHD TIP60, p400, TRRAP, Brd8, EPC1/2	IP, pd	(Doyon et al, 2006)
	N/D	GAS41	IP, pd	(Doyon et al, 2006)

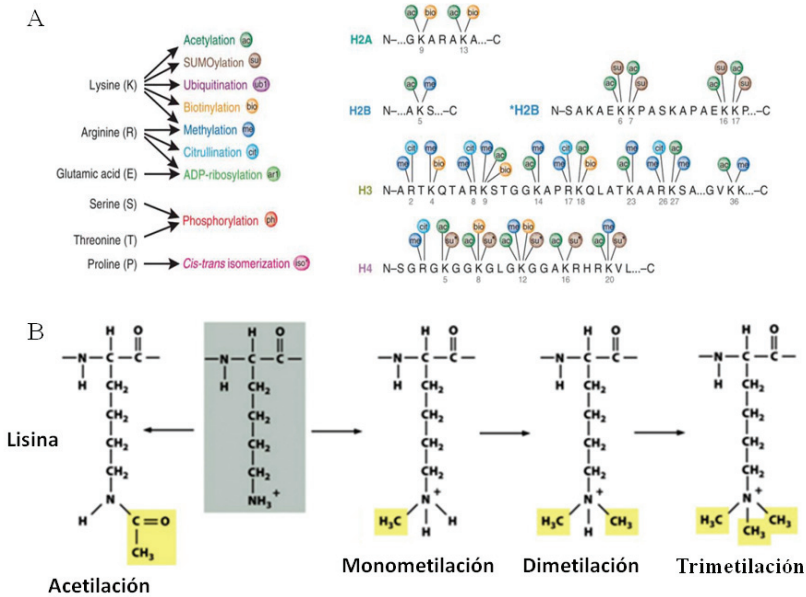
ING4/ Q9UNL	PHD	H3K4me1/2/3	ChIP, pd, fluo	(Peña et al, 2006; Shi et al, 2006)
		HBO1	IP	(Doyon et al, 2006)
		p53	IP	(Shiseki et al, 2003)
		p300	IP	(Shiseki et al, 2003)
	N/D	hEaf6	IP, MS	(Doyon et al, 2006; Ullah et al, 2008)
		HPH-2	y2h, coIP, pd, ChIP	(Ozer et al, 2005)
		JADE1/2/3	IP, pd	(Doyon et al, 2006)
		G3BP2a	IP	(Shen et al, 2007; Unoki et al, 2006)
		NF-jB p65	IP, pd	(Nozell et al, 2008; Xin et al, 2004)
		EBNA	IP, pd	(Saha, 2001)
NCR	Liprin alpha 1	IP	(Unoki et al, 2006)	
ING5/ Q8WYH8	PHD	H3K4me1/2/3	ChIP, IP, pd, fluo	(Champagne et al, 2008; Peña et al, 2006; Shi et al, 2006)
		HBO1	IP, pd	(Doyon et al, 2006)
		MOZ/MORF	IP, pd	(Doyon et al, 2006)
		p53	IP	(Shiseki et al, 2003)
		p300	IP	(Shiseki et al, 2003)
	N/D	BRPF1/2/3	IP, pd	(Doyon et al, 2006)
		JADE1/2/3	MS, IP,	(Doyon et al, 2006)
		MCM2/4/6	IP, pd	(Doyon et al, 2006)
		INCA	IP, pd	(Zhang et al, 2011)
		EBNA	IP, pd	(Saha, 2001)

N/D= No Determinado, y2h= Yeast Two-Hybrid Assay, pd= pull down, IP= immunoprecipitation plus Western blotting, ch= chromatin, MS= mass spectrometry. coIP= coimmunoprecipitation plus Western blotting, (LC/MS)= liquid chromatography-mass spectrometry, fluo= tryptophan fluorescence, ChIP =Chromatin Immunoprecipitation plus Western blotting. Debajo del nombre de cada proteína ING se encuentra su código UNIPROT.

### 1.3 El código de histonas y la expresión génica

La cromatina es un enorme complejo macromolecular de proteínas y ADN cuya unidad básica es el nucleosoma. Este está compuesto por dos copias de cada una de las cuatro histonas (H2A, H2B, H3 y H4) nucleosomales formando un octámero alrededor del cual se enrollan firmemente una doble hélice de 146 pares de bases de ADN. Las histonas son proteínas pequeñas, básicas y cargadas positivamente que interactúan con los grupos fosfato cargados negativamente del ADN (Rizzo, 2003). Estructuralmente las histonas están constituidas por un dominio globular con una estructura altamente conservada de tres hélices alfa, conectadas por dos bucles (Luger et al, 1997). El dominio globular está flanqueado por los extremos N-terminal y C-terminal flexibles y con alta densidad de carga positiva donde tienen lugar la mayoría de modificaciones postraduccionales (MPTs) que afectan más de 60 sitios en las histonas presentes en el nucleosoma (Luger et al, 1997; Talbert & Henikoff)(Figura 5). La cromatina puede adoptar dos formas: la heterocromatina, que se caracteriza por su alto grado de condensación, implica una baja accesibilidad de los genes y una baja o nula transcripción de los genes en los segmentos de ADN empaquetados de esta forma, y la eucromatina que posee un menor grado de condensación, por lo que sus genes están más accesibles, relacionándose con replicación y transcripción activa. Los procesos de remodelación de la cromatina actúan modificando la estructura de la misma, provocando así una mayor o menor accesibilidad al ADN. Los tipos de modificaciones presentes en las histonas incluyen la acetilación de lisinas, metilación de lisina y arginina, fosforilación de serina y treonina, sumoilación de lisina, ubiquitinación de lisina, ADP ribosilación de glutamato y la deiminación de arginina (citrulinación) (Kouzarides, 2007; Krebs et al, 1999; Latham & Dent, 2007).



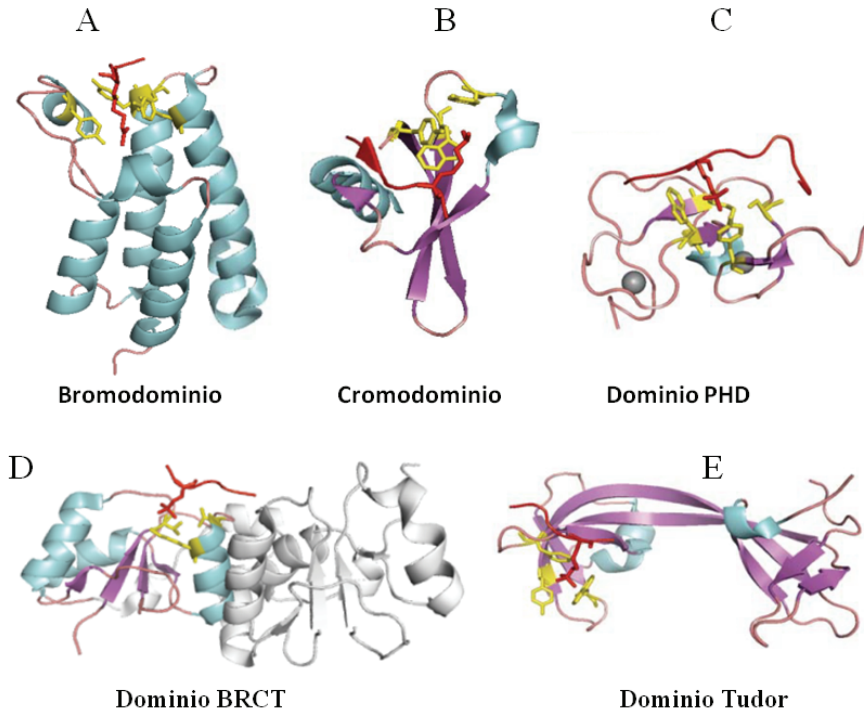


**Figura 5** Modificaciones postraduccionales descritas en la región N-terminal de las cuatro histonas nucleosomales humanas. (A) Las diferentes modificaciones, así como los diferentes residuos que las sufren, se representan con distintos colores especificados en los esquemas. El asterisco indica que la modificación ha sido descrita en *S. cerevisiae*. La figura fue tomada de la referencia (Latham & Dent, 2007). (B) Acetilación y diferentes estados de metilación de lisinas en las colas de histonas. Figura adaptada de Molecular Biology of the Cell, Bruce Alberts et al. 5ª ed. (2008) p222-223.

La acetilación de histonas tiene lugar en los grupos amino de residuos de lisina de las cuatro histonas y está relacionada con activación de la transcripción. Tomando en cuenta los requerimientos electrostáticos necesarios para el ensamblaje del ADN en la cromatina, la acetilación de las histonas en los residuos de lisina produce la neutralización de su carga positiva reduciendo las interacciones electrostáticas entre el ADN y las histonas. Con esto, el equilibrio entre histona libre e histona empaquetada en nucleosomas se desplaza a favor de las histonas libres, de modo que la cromatina adopta, localmente, una estructura menos condensada, facilitando de esta forma la activación de la transcripción (Santos-Rosa & Caldas, 2005); la fosforilación está asociada con diferentes resultados dependiendo del sitio donde se encuentre, por ejemplo la fosforilación que ocurre en la

región N-terminal de la histona 3 introduce carga negativa que está asociada a procesos de activación de la transcripción (Cheung et al, 2000). La metilación de histonas se produce en residuos de lisina y arginina, siendo las histonas 3 y 4 las más frecuentemente metiladas. La arginina puede ser mono- o di-metilada simétrica o asimétricamente en su grupo guanidinio. La lisina puede aceptar uno, dos o tres grupos metilos en su grupo amino, generando las formas mono-, di-, o tri-metilada. Las MPTs cumplen un papel fundamental en la dinámica de la cromatina. Existe una correlación entre metilación y sitios activos en transcripción, aunque su reconocimiento puede tener como consecuencia tanto la activación como la represión de la transcripción (Garkavtsev et al, 1996; Latham & Dent, 2007; Santos-Rosa & Caldas, 2005). Las MPTs son “escritas” y “borradas” por un gran número de complejos enzimáticos especializados. Las combinaciones específicas de diferentes MPTs, junto con la presencia de múltiples dominios “lectores” en una proteína o en un complejo proteico ha inspirado la hipótesis del código de histonas que puede ser definido como un mecanismo de regulación en el cual las diferentes combinaciones de MPTs presentes en las colas de las histonas al interactuar con complejos lectores pueden producir un efecto específico en la dinámica de la cromatina y en la expresión genética, lo cual presenta la posibilidad de ampliar el potencial de información existente en el ADN. Se ha demostrado que las MPTs interactúan entre sí para producir sinergia o antagonismo entre ellas generando un sistema de regulación basado en la afinidad de proteínas al nucleosoma, lo cual influye en procesos celulares fundamentales como el estado de empaquetamiento de la cromatina, la replicación, recombinación y reparación del ADN (Garkavtsev et al, 1996; Latham & Dent, 2007). La “lectura” del código de histonas se realiza de un modo dependiente tanto del residuo y la modificación concreta, como de la secuencia particular de la histona modificada. Se conocen varios tipos de dominios lectores. Los mejor caracterizados incluyen: los bromodominios que reconocen específicamente las histonas H2B, H3 y H4 acetiladas (Khorasanizadeh, 2004; Taverna et al, 2007). Los cromodominios reconocen los estados de metilación en H3K9me2/3 y H3K27me2/3 y el doble cromodominio reconoce los estados de metilación H3K4me1/2/3), además de la fosforilación de los residuos serinas y treoninas (Flanagan et al, 2005; Sims et al, 2005). Los dominios PHD reconocen los estados de metilación H3K4me1/2/3), los dominios

BRCT reconocen específicamente la fosforilación de los residuos serina en la histona H2 (H2AXS139ph) y los dominios Tudor pueden reconocer H3K4me3, H4K20me3 o residuos de arginina dimetilada en las histonas H3 y H4 (Cote & Richard, 2005; Lee et al, 2008; Taverna et al, 2007).



**Figura 6** Estructura de diferentes dominios de reconocimiento de modificaciones postraduccionales de histonas unidos a los correspondientes fragmentos de histonas: (A) Bromodominio de GCN5 unido a H4K16ac (código PDB:IE6I). (B) Cromodominio de HP1 unido a H3K9me3 (código PDB:1KNA). (C) Dominio PHD de ING4 unido a H3K4me3 (código PDB:2VNF). (D) Dominio BRCT unido a H2AXS139ph (código PDB: 2AZM). (E) Dominio Tudor unido a H4K20me3 (código PDB: 2QQS). Las cadenas laterales involucradas en la unión a los péptidos se exponen en color amarillo, los péptidos se exponen en color rojo.

### 1.4 Reconocimiento de H3K4me3 por parte de las proteínas ING

La familia de proteínas INGs está involucrada en la regulación epigenética, mediante el reconocimiento de la metilación en el residuo de lisina 4 de la histona H3 a través de su

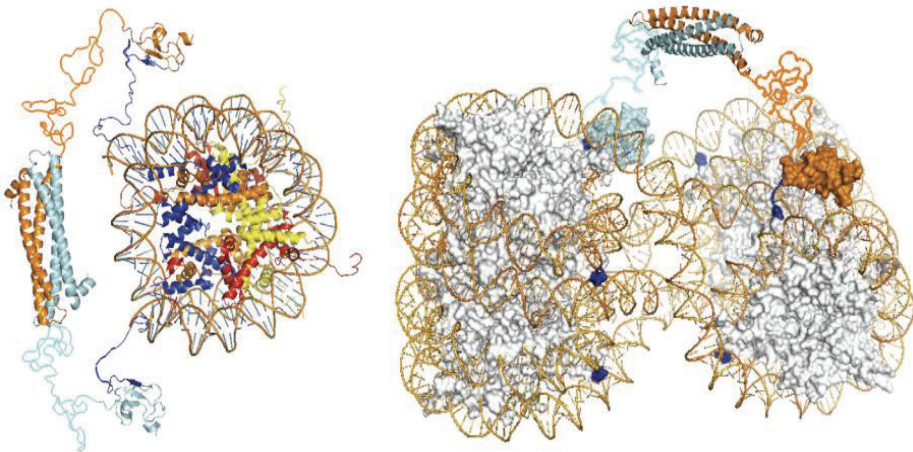
dominio PHD. La afinidad de la interacción entre el dominio PHD de las proteínas ING y péptidos de 10 a 15 aminoácidos de longitud correspondientes al extremo terminal de histona H3 con distintos grados de metilación en K4 (H3K4meX, donde X=0,1,2,3) se ha cuantificado mediante medidas de fluorescencia o RMN (Tabla 2). En general se observa que la unión a H3K4me0 es débil o muy débil y que la afinidad es mayor para H3K4me3. Algunos dominios muestran una selectividad alta por las formas X=1,2, o 3 (como ING1, 2 y 5) mientras que esta selectividad es baja para ING4.

**Tabla 2** Resumen de las constantes de disociación (Kd) medidas para las reacciones de unión de los distintos dominios PHD de las proteínas ING y péptidos correspondientes a la región N-terminal de histona H3 con distintos grados de metilación en el residuo K4.

PHD	Péptido	Kd ( $\mu$ M)	Método	Referencia
ING1	H3K4me0	N/D		
	H3K4me1	419 $\pm$ 71	RMN	(Peña et al, 2008)
	H3K4me2	17.3 $\pm$ 3.9	Fluorescencia	(Peña et al, 2006)
	H3K4me3	3.3 $\pm$ 1.6		
ING2	H3K4me0	2.240 $\pm$ 350	Fluorescencia y RMN	(Peña et al, 2006)
	H3K4me1	208 $\pm$ 80		
	H3K4me2	15 $\pm$ 4		
	H3K4me3	1.5 $\pm$ 1		
ING3	H3K4me0	N/D		
	H3K4me1	N/D		
	H3K4me2	N/D		
	H3K4me3	6.9 $\pm$ 1.1	Fluorescencia	(Peña et al, 2006)
ING4	H3K4me0	370 $\pm$ 20	RMN	(Palacios et al, 2006)
	H3K4me1	1.6 $\pm$ 0.8		
	H3K4me2	1.8 $\pm$ 1		
	H3K4me3	4.0 $\pm$ 0.7	Fluorescencia	(Peña et al, 2006)
	H3K4me3	7.9 $\pm$ 2		
ING5	H3K4me0	261 $\pm$ 34	Fluorescencia	(Champagne et al, 2008)
	H3K4me1	222 $\pm$ 17		
	H3K4me2	16 $\pm$ 1.2		
	H3K4me3	2.4 $\pm$ 1.0		

Desde un contexto estructural se puede sugerir que la familia ING ha evolucionado para facilitar el reclutamiento de complejos de remodelación hacia nucleosomas con esta marca

en la histona H3. Teniendo en cuenta los estudios estructurales y funcionales con las proteínas ING, se puede proponer que la región N terminal de las INGs actúa como sitio de dimerización y de unión a complejos de acetilación (HAT) y desacetilación de histonas (HCDA), el dominio PHD sería el encargado de “anclar” la proteína al nucleosoma por medio del reconocimiento de las diferentes marcas de histonas y el dominio central, está desestructurado y flexible permitiría mantener la dinámica estructural necesaria para que los dominios PHD se unan a las colas de histona H3 en el mismo o en diferentes nucleosomas (Figura 7). En el caso de ING4 se ha comprobado experimentalmente que ING4 es un receptor bivalente ya que es un dímero, y al ser este antiparalelo, sus dominios PHD se ubican espacialmente en posiciones opuestas que teóricamente permitirían abarcar una gran área de interacción con las marcas de histona presentes en los nucleosomas en un amplio radio (Culurgioni et al, 2012).

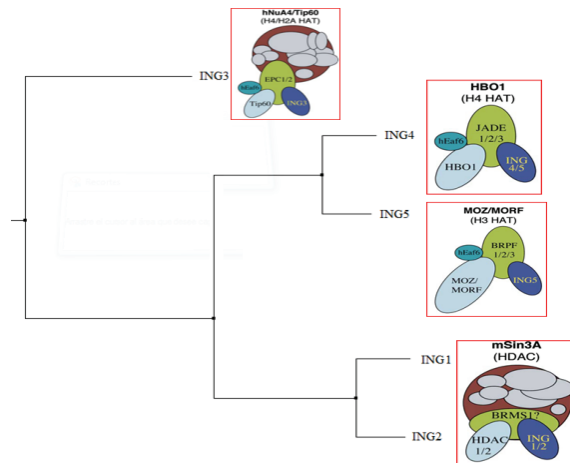


**Figura 7** Modelo de dímero de ING4 unido a nucleosoma. Dada la estructura de homodímero antiparalelo y la flexibilidad de la región central, la molécula de ING4 podrían unirse simultáneamente a dos colas de histona del mismo nucleosoma ó de nucleosomas distintos, también cabe la posibilidad de que cada uno de los dos dominios PHD interaccione con histonas y el otro dominio PHD del dímero lo haga con otra proteína. Figura tomada de (Culurgioni et al, 2012).

### 1.1.10 Complejos de acetilación/desacetilación de histonas en los que participan las proteínas ING

#### 1.1.11 Complejos de acetilación de histonas

Las enzimas que acetilan el grupo amino épsilon de lisinas son proteínas de la familia MYST, que está compuesta por las proteínas humanas HBO1, MOZ, MORF, y Tip60 (Avvakumov & Cote, 2007; Gorrini et al, 2007; Voss & Thomas, 2009; Yang & Ullah, 2007). Estas enzimas forman parte de complejos multiproteicos de acetilación de histonas (HAT) denominados HBO1, MOZ/MORF y hNuA4/Tip60. La acetilación de las lisinas de las histonas es contrarrestada por complejos de desacetilación de histonas HDAC que desacetilan estas lisinas. Mediante experimentos de purificación por afinidad en tándem combinados con análisis de MS de las proteínas separadas en geles (TAP-MS) se ha encontrado que las proteínas ING3,4 y 5 forman parte de diferentes complejos HAT, y que las proteínas ING1 y 2 forma parte del complejo HDAC mSin3A (que puede contener las enzimas HDAC1 o HDAC2).



**Figura 8** Árbol filogenético de la familia de proteínas ING humanas. El árbol fue generado utilizando el alineamiento de las secuencias (utilizando la isoforma canónica o la que se expresa mayoritariamente en tejidos en el caso de ING1b) mediante análisis de distancia y parsimonia, obteniéndose igual filogenia. Los clados presentes en la filogenia indican que existen tres subgrupos dentro de la familia ING: (ING1/2), ING3 y (ING4/5). Al final de las ramas del árbol se encuentra el esquema del complejo HAT o HDAC del cual la proteína forma parte con el fin de exponer el paralelismo existente entre divergencia filogenética y funcional. La proteína ING5 puede formar parte de dos complejos HAT.

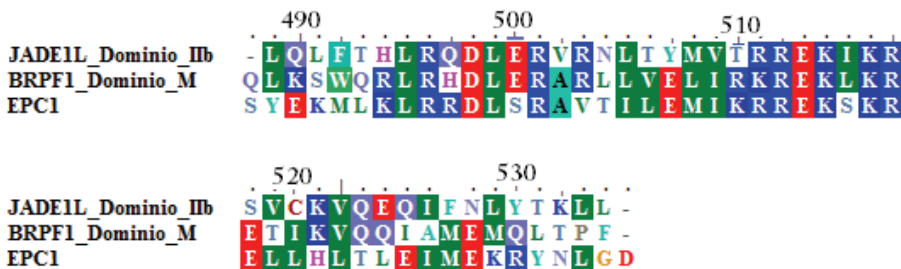
### 1.1.12 Complejo HAT MOZ/MORF

Las acetiltransferasas MOZ y MORF (Dominio de dedos de Zinc de leucemia monolítica y factor relacionado con él) forman complejos con ING5, EAF6 (Proteína ortóloga humana del factor 6 asociado a Esa-1) y BRPF (bromodominio - dominio dedos de zinc PHD-1/2/3). MORF es requerido específicamente para la acetilación de H4K16 (Akhtar & Becker, 2000; Taipale et al, 2005; Thomas et al, 2008), mientras que MOZ tiene una función específica en la acetilación de H3K9 en genes Hox (Voss et al, 2009). La proteína BRPF es la plataforma de ensamblaje de los complejos MOZ/MORF ya que posee en su región N terminal el sitio de interacción con MOZ/MORF (Ullah et al, 2008) y en su región C terminal el dominio M que contiene la región EPc (*Enhancer of PolyComb*), que es el responsable de la interacción directa con ING5 y EAF6 (Avvakumov et al). Un aspecto desconocido es cómo se complementan las funciones de reconocimiento de marcas de histonas de los diferentes dominios PHD presentes en este complejo más interesante al estudiar la estructura de este tipo de complejos donde existen diferentes clases de dominios PHD, es entender cómo actúan en el entorno celular (existen al menos 5 dominios en cada complejo: 2 PHD de MOZ ó MORF, 2 PHD de BRPF y el de ING5) (Ullah et al, 2008).

### 1.1.13 Complejo HAT HBO1

HBO1 fue identificado durante la búsqueda de proteínas que interactúan con ORC1 (*origin recognition complex protein 1*) (Bell & Stillman, 1992; Dutta & Bell, 1997); además de esta interacción HBO1 también se asocia con subunidades de complejo de pre-replicación. El dominio PHD presente en HBO1 puede interactuar con la proteína MCM2 (*minichromosome maintenance*) lo cual sugiere que el complejo está involucrado en procesos de regulación de la replicación de ADN (Saksouk et al, 2009). El complejo acetiltransferasa HBO1 es necesario para la expresión de un gran número de genes y es responsable de la acetilación de la histona H4 en las lisinas K5, K4 y K12 y en la acetilación de la histona H3 en K14. El complejo HBO1 está formado por las proteínas HBO1, JADE (*gene (J) for Apoptosis and Differentiation in Epithelial*) 1/2/3, ING4/ING5 y EAF6, en los cuales los dominios PHD de ING4/5 y JADE1/2/3 reclutan el complejo hacia la cromatina y permiten su interacción mediante la unión a las histonas metiladas en

los residuos lisinas (Foy et al, 2008). Tanto ING4 como ING5 se asocian al complejo HBO1 de forma independiente, las diferencias funcionales entre estos complejos y la posibilidad de la existencia de heterodímeros entre ING4/5 basada en el alto grado de homología en secuencia y la conservación de los residuos que permiten las interacciones necesarias para la heterodimerización, ampliaría el espectro funcional de estos complejos en el nucleosoma. La proteína JADE es la plataforma para el ensamble del complejo HBO1. Existen dos isoformas de JADE1 presentes en los complejos HBO1, que se diferencian en su tamaño y se denominan JAD1L y JADE1S (Doyon et al, 2006). JADE1S forma complejos con HBO1 y EAF6 pero no se asocia con ING4/5, lo cual resulta en un complejo con mayor preferencia por la marca H3K36 que por H3K4 (Saksouk et al, 2009), experimentos de tipo *pull-down* indican que JADE1L interactúa directamente con ING5 a través de su dominio IIb, y con HBO1 a través de su dominio I (Avvakumov et al, 2011).



**Figura 9** Alineamiento de secuencias de dominios de interacción tipo JADE, BRPF y EPC1. El alineamiento expone que existe conservación en secuencia entre proteínas pertenecientes a complejos de acetilación de histonas HAT. Existe evidencia experimental de que estos dominios son las regiones de interacción con proteínas ING. El alineamiento fue realizado con ClustalW 2.0. La numeración del alineamiento se basa en la secuencia que el dominio IIb ocupa en la proteína JADE1L. Los colores de los residuos indican: azul: con carga positiva, rojo: con carga negativa, verde: hidrofóbico, gris: polar no cargados. La numeración del alineamiento se basa en la secuencia de JADE 1L

Por otro lado, experimentos de coimmunoprecipitación muestran que BRPF interactúa con ING5 y hEAF6 a través de su dominio M (Ullah et al, 2008), el cual tiene un alto grado de homología de secuencia con JADE1L\_dominio IIb (Avvakumov et al, 2011). También se ha determinado que el dominio EPcA de la proteína EPC1/2 es el dominio de interacción

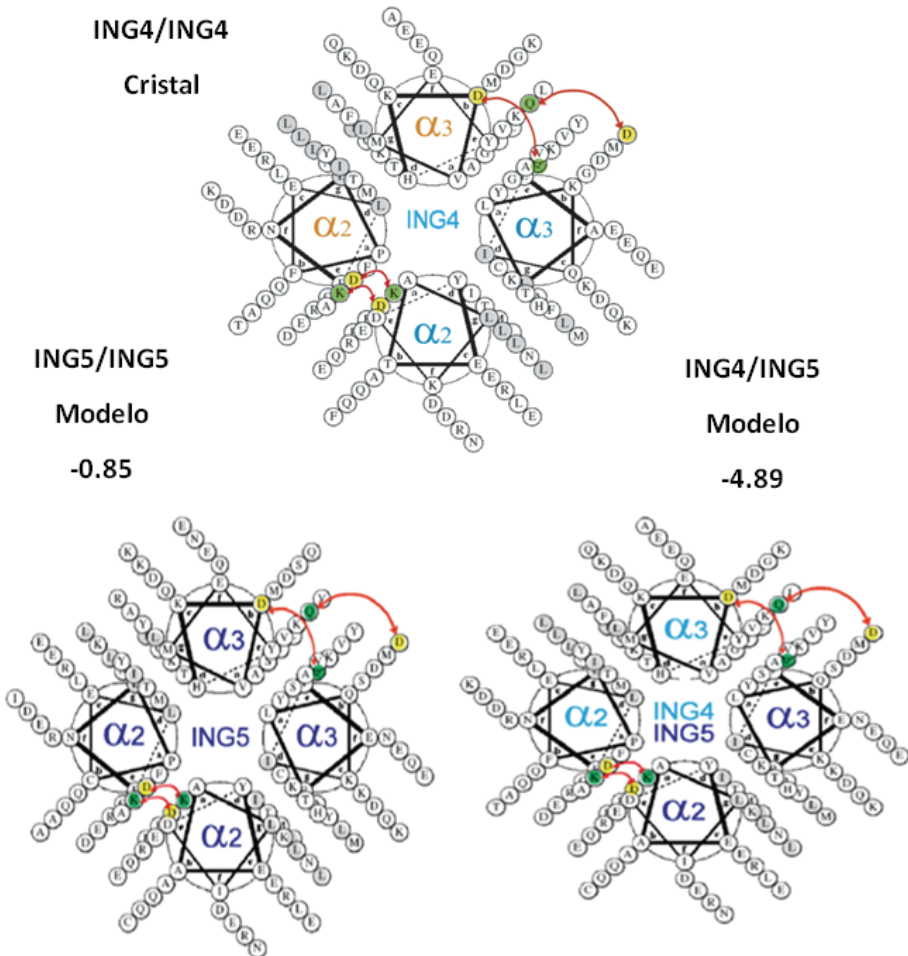


con ING3 en el complejo Tip60, posee también, aunque en menor grado, una homología con los dominios IIB y M de JADE1L y BRPF, como se muestra en la Figura 9. Esto sugiere que los sitios de unión en las proteínas INGs a los complejos MYST-HAT están conservados. Cada complejo HBO1 contiene tres clases de dominios PHD, (JADE-PHD1/PHD2, ING4/5-PHD) los cuales se unen a la histona 3 (H3) con diferentes especificidades que dependen del estado de metilación de la lisina (K4), y son fundamentales en la regulación funcional del complejo HBO1.

### 1.1.14 La proteína ING5

ING5 tiene un alto nivel de homología de secuencia con ING4 y ambas pueden asociarse de forma independiente al complejo HBO1. Probablemente esta homología se mantenga a nivel estructural e ING5 al igual que ING4 esté constituido por tres dominios estructurales: un dominio de dimerización N terminal, un dominio central flexible y en la región C terminal un dominio PHD (Kuzmichev et al, 2002; Scott et al, 2001b). La estructura del dominio PHD de ING5 aislado y unido al péptido H3K4me3 se ha determinado por cristalografía y es muy similar a la correspondiente estructura del PHD de ING4. Sin embargo no hay estudios estructurales de los otros dominios o de la proteína completa. Posiblemente el dominio N-terminal de ING5 forme dímeros e ING5 sea un homodímero y por tanto un receptor bivalente de H3K4me3, como lo es ING4. La alta homología de secuencia entre los dominios N-terminales de las proteínas ING hace posible modelarlos sobre la estructura experimental de ING4 y evaluar su posible estabilidad relativa (Figura 10). Este análisis sugiere que los homodímeros de ING5 son probablemente tan estables como los de ING4 ya que se darían interacciones estabilizantes muy similares a las encontradas en ING4, mientras que en los casos de ING1, 2 y 3 los modelos anticipan una estabilidad mucho menor (Culurgioni et al, 2012). De una forma similar se puede modelar la estructura de un posible heterodímero ING4/ING5, y el resultado es que este heterodímero sería tan estable o más que los correspondientes homodímeros, (Figura 10). Esto, que no deja de ser una predicción basada en unos métodos computacionales con

limitaciones importantes, sería análogo a la formación de heterodímeros confirmada experimentalmente en el caso de los supresores de tumores p63 y p73 (Joerger et al, 2009).



**Figura 10** Distribución de residuos en las estructuras coiled-coil idealizadas de ING4 e ING5 en forma de homodímeros y como heterodímero. Los círculos de color gris claro indican los residuos Leu e Ile que forman la secuencia tipo cremallera de leucina "zipper". Los círculos amarillos y verdes indican los residuos negativamente y positivamente cargados (ó donador de enlace de hidrógeno), respectivamente, que interactúan estabilizando el dímero o los residuos cargados enterrados (unidos mediante líneas rojas). Los números en la parte superior de las gráficas indican la energía libre calculada de la interacción entre los dos protómeros en relación con el correspondiente valor calculado para el homodímero ING4, en kcal/mol. Estos valores se calcularon con el programa FoldX (Van Durme et al, 2011). Cuanto menor sea este número, se predice que es más estable el dímero ó heterodímero en relación con homodímero de ING4.

## Introducción

---

Se han descrito tres mutaciones puntuales en las regiones codificantes del gen ING5 detectadas en cáncer, las cuales dan lugar a los cambios de aminoácido Q33R, I68V, C75R (Cengiz et al, 2010). Las tres se localizan en el dominio N-terminal. Su detección en biopsias no implica necesariamente una relación de causa-efecto. Sin embargo es posible plantear como hipótesis que su posible papel en el desarrollo del cáncer se debe bien a una desestabilización de la estructura de la proteína, o bien a una interferencia con su interacción con otras proteínas a través de su dominio N-terminal.



## **2 Objetivos**



El objetivo general de esta tesis es la caracterización experimental de la estructura de la proteína supresora de tumores ING5. Este objetivo general se desarrolló en los siguientes objetivos particulares:

- Análisis de la organización estructural de la proteína ING5 completa.
  
- Determinación de la estructura tridimensional del dominio N-terminal de ING5.
  
- Evaluar la posible formación de heterodímeros ING4/ING5.
  
- Evaluar el efecto de mutaciones detectadas en cáncer en la estructura de ING5.
  
- Caracterización estructural de la interacción de ING5 con el complejo HBO1 a través de la proteína JADE1L.
  
- Estudiar la interacción de la proteína ING5 completa con el péptido de histona H3K4me3 y comparar con la interacción con el dominio PHD aislado.





### **3 Materiales y Métodos**





## 1.5 Clones utilizados y manipulaciones mediante técnicas de Biología Molecular

### 1.1.15 Clones de ING5

Se realizó un análisis de distribución de codones en el gen humano que codifica para la proteína ING5 (Código Uniprot: Q8WYH8), mediante el servidor [http://gcua.schoedl.de/sequential\\_v2.html](http://gcua.schoedl.de/sequential_v2.html). Este análisis mostró que el gen humano posee un 33% de codones que son poco frecuentes en *E.coli*, lo cual podría ocasionar problemas en la expresión de la proteína en cultivos bacterianos. Por este motivo, en lugar de usar un cDNA para clonar el gen en un vector de expresión en bacterias, se compró un gen sintético de la proteína ING5 humana con la secuencia de ADN optimizada en base al uso preferente de codones para expresión en *E.coli*. Este gen fue sintetizado y clonado en el vector de expresión pET11d (usando los sitios de restricción NcoI/BamHI) por la empresa *Entelechon*. La experiencia previa en el laboratorio con la purificación de la proteína ING4 sin cola de afinidad era que resultaba difícil y muy laboriosa, necesitando numerosos pasos cromatográficos y con rendimientos bajos. Por ello se diseñó un clon que contenía una cola de histidinas en el extremo N-terminal para facilitar su purificación mediante cromatografía de afinidad por metal inmovilizado (IMAC), y un sitio de corte específico para la proteasa TEV para eliminar la cola una vez purificada la proteína. Entre la secuencia de corte TEV y la metionina inicial se intercaló un residuo de alanina para prevenir posibles dificultades en el corte proteolítico que pudieran derivarse de la cercanía de un residuo voluminoso como la metionina. La secuencia de nucleótidos del gen se encuentra en el anexo 1. La secuencia completa de la proteína producida por este clon es la siguiente:

```
MGSSHHHHHH SSGENLYFQG AMATAMYLEH YLDSIENLPC ELQRNFQLMR ELDQRTEDKK
AEIDILAAEY ISTVKTLSPD QRVERLQKIQ NAYSKCKEYS DDKVQLAMQT YEMVDKHIRR
LDADLARFEA DLKDKMEGSD FESSGGRGLK KGRGQKEKRG SRGRGRRTSE EDTPKKKKHK
GGSEFTDTIL SVHPSDVLDM PVDPNEPTYC LCHQVSYGEM IGCNDPDCPI EWFHFACVDL
TTKPKGKWFC PRCVQEKRRK K
```

Donde los residuos de la cola His y los del sitio de corte TEV están indicados con letras rojas y verdes, respectivamente. El corte con la proteasa se produce entre los residuos Q y

G, por lo que la proteína finalmente purificada contiene los residuos GA antes del primer residuo de la proteína ING5 humana.

El clon correspondiente al dominio N-terminal de ING5 (ING5<sub>1-105</sub>) fue obtenido insertando un codón de terminación en el sitio correspondiente al residuo R106, y la secuencia de aminoácidos completa es la siguiente:

MGSSHHHHHH SSGENLYFQG AMATAMYLEH YLDSIENLPC ELQRNFQLMR ELDQRTEDKK  
 AEIDILAAEY ISTVKTLSPD QRVERLQKIQ NAYSKCKEYS DDKVQLAMQT YEMVDKHIRR  
 LDADL

El clon ING5<sub>1-105</sub> fue usado como molde para obtener los diferentes mutantes de ING5<sub>1-105</sub> (Q33R, I68V, C75R) mediante el Kit Quick Change Site-Directed Mutagenesis (*Stratagene*). Los cebadores usados en la mutagénesis dirigida se diseñaron con el programa *QuikChange Primer Design (Agilent Technologies)*. Las colonias obtenidas siguiendo el protocolo del Kit se cultivaron en medio LB y se realizó una preparación del plásmido para la secuenciación del gen correspondiente usando los cebadores estándar T7 directo y reverso (secuenciación realizada por la empresa StabVida).

**Tabla 3** Listado de los cebadores empleados para la obtención de los diferentes clones.

Nombre	Clon generado	Secuencias de los cebadores directo y reverso
ING5-StopR106	ING5 <sub>1-105</sub>	5'- GCTGATCTCGCGTGATTTGAAGCTGAC CTCAAAGAC-3' Reverso 5'- GTCAGCTTCAAATCACGCGAGATCAGC ATCCAGACG-3'
ING5-C75R	ING5 <sub>1-105</sub> C75R	5'- ATCCAGAACGCCTATAGCAAACGCAAA GAATACTCTGAC-3' Reverso 5'- GTCAGAGTATTCTTTGCGTTTGCTATAG

		GCGTTCTGGAT-3'
ING5-Q33R	ING5 <sub>1-105</sub> Q33R	<p>5'- GATGCGTGAAGTGGATCGGCGTACCGA AGACAAAA-3'</p> <p>Reverso 5'- TTTTGTCTTCGGTACGCCGATCCAGTTC ACGCATC-3'</p>
ING5-I68V	ING5 <sub>1-105</sub> I68V	<p>5'- AGCGCCTGCAGAAGGTCCAGAACGCCT AC-3'</p> <p>Reverso 5'- GTAGGCGTTCTGGACCTTCTGCAGGCG CT-3'</p>
Strep-ING4	ING4 con cola Strep	<p>5'- ACTTTAAGAAGGAGATATACCATGTGG AGCCACCCGAGT TCGAAAAAGCCGAGGTATGTACCTGG AA-3'</p> <p>Reverso 5'- TTCCAGGTACATACCTGCGGCTTTTTCG AACTGCGGGTGG CTCCACATGGTATATCTCCTTCTTAAAG T-3'</p>
ING4-StopA108	ING4 <sub>2-107</sub>	<p>5'- ACCGTACTGGCTCGTTTCGAATAGGAT CTGAAAGAGAAACA-3'</p> <p>R Reverso 5'- TGTTTCTCTTTCAGATCCTATTCGAAAC GAGCCAGATCGGT-3'</p>

### 1.1.16 Otros clones

El gen sintético de ING4 humano (Código Uniprot: Q9UNL4-1) con la secuencia de ADN optimizada en base al uso preferente de codones para expresión en *E.coli* se ha descrito previamente (Palacios et al, 2010). Este clon no contenía una secuencia para purificación de la proteína mediante cromatografía de afinidad, con el resultado de que la purificación era poco eficiente y los rendimientos escasos. Este clon fue usado como molde para insertar una cola Strep en la región N-terminal mediante PCR mutagénica. El objetivo era doble: facilitar la purificación e introducir una cola de afinidad ortogonal con la cola de histidinas de ING5 para los ensayos de heterodimerización por cromatografía en tándem. La secuencia de aminoácidos de este clon de la proteína ING4 completa es la siguiente:

```
MWSHPQFEKA  AGMYLEHYLD  SIENLPFELQ  RNFQLMRDLQ  QRTEDLKAEI  DKLATEYMSS
ARSLSSEEKL  ALLKQIQEAY  GKCKEFGDDK  VQLAMQTYEM  VDKHIRRLDT  DLARFEADLK
EKQIESSDYD  SSSSKGKKKG  RTQKEKKAAR  ARSKGKNSDE  EAPKTAQKKL  KLVRTSPEYG
MPSVTFGSVH  PSDVLDMPVD  PNEPTYCLCH  QVSYGEMIGC  DNPDCSIEWF  HFACVGLTTK
PRGKWFCPRC  SQERKKK
```

Donde los residuos de la cola Strep están escritos en rojo. El clon ING4<sub>2-107</sub> se obtuvo insertando un codón de terminación en el sitio correspondiente a la A108 mediante PCR mutagénica. La intención original fue generar el clon ING4<sub>2-105</sub>, análogo al ING5<sub>1-105</sub> pero por un error en el diseño de los cebadores se generó el clon ING4<sub>2-107</sub>. La secuencia de aminoácidos de este clon es la siguiente:

```
MWSHPQFEKA  AGMYLEHYLD  SIENLPFELQ  RNFQLMRDLQ  QRTEDLKAEI  DKLATEYMSS
ARSLSSEEKL  ALLKQIQEAY  GKCKEFGDDK  VQLAMQTYEM  VDKHIRRLDT  DLARFE
```

El clon para la producción de la proteína de fusión GST-JADE1L\_dominio\_IIB (residuos 489-535) en el plásmido pGEX-4T3 nos fue cedido cordialmente por el laboratorio del Dr. Jacques Côté (*Laval University Cancer Research Center, Canadá*). La secuencia de JADE1L\_dominio IIB (residuos 489-535) es la siguiente:

```
LQLFTHLRQD  LERVRNLTYM  VTRREKIKRS  VCKVQEIQFN  LYTKLL
```

Se utilizó un clon para la producción del dominio PHD de ING5 fusionado a GST proporcionado por el laboratorio de la Dra. Tatiana Kutateladze, (University of Colorado, EEUU), que había publicado la estructura cristalina del complejo con H3K4me3. Sin embargo los intentos de purificación fueron infructuosos y al secuenciar el clon comprobamos que la secuencia no era correcta. Hubo un error con el plásmido enviado y no fue posible localizar y recibir el clon correcto. Sin embargo la Dra. Kutateladze ofreció compartir los datos de RMN que tenían registrados con una muestra del dominio PHD de ING5 y que no habían analizado. Dado que nuestros intentos de amplificar y clonar el dominio a partir del clon de ING5 no tuvieron éxito, optamos por abandonar la preparación de la muestra y analizar los datos de RMN registrados por el grupo de Dra. Kutateladze en Colorado.

La secuencia de aminoácidos de la proteína analizada corresponde a los residuos 174-240 de ING5, y es la siguiente:

```
SDVLDMPVDP NEPTYCLCHQ VSYGEMIGCD NPDCPIEFWH FACVDLTTKP KGKWFPCRCV  
QEKRRKKK
```

## **1.6 Expresión génica y producción de las proteínas.**

### **1.1.17 Transformación y pruebas de expresión de proteínas en bacterias.**

Células de *E. coli* BL21(DE3) fueron transformadas con los diferentes plásmidos con los genes que codifican para las proteínas de este estudio, por el método de transformación química mediante choque térmico. Las colonias resultantes de la transformación fueron sujetas a pruebas de expresión para seleccionar las colonias con mejor nivel de expresión. La metodología de la prueba consistió en inocular 2 mL de medio LB (Luria Bertani) con una única colonia y dejar crecer el cultivo hasta una OD<sub>600</sub> entre 0.5-0.6 e inducción con 0.5 mM de IPTG durante 3 h a 37 °C. Una muestra de 1 mL del cultivo fueron centrifugadas durante 5 minutos a 12,000 g y el sedimento fue resuspendido en 0.05 mL de 20 mM fosfato sódico pH 7.3, 150 mM NaCl (PBS), por cada unidad de OD<sub>600</sub> en el cultivo cosechado. El nivel de producción de proteínas se evaluó mediante electroforesis desnaturalizante (SDS) en geles de poliacrilamida (SDS-PAGE) (Laemmli, 1970), utilizando geles del 12% hechos en el laboratorio o geles con gradiente NuPAGE Novex 4-

12% Bis-Tris de Life Technologies. Los geles se corrieron a un voltaje constante de 200 V durante 35-50 minutos a temperatura ambiente, se tiñeron con Coomassie Brilliant Blue al 0.01 % y se destiñeron con una mezcla metanol 30%, agua 60%, acético 10%. En la figura 11A se muestra un gel con la prueba de expresión de la proteína ING5. En algunos casos se confirmó la expresión de las proteínas de interés y la posición de la banda correspondiente en el gel de SDS-PAGE mediante *western blots* con anticuerpos anti-His y anti-Strep. Se seleccionaron las colonias con mayor nivel de expresión y se repitió el proceso con un volumen de cultivo de 5 mL e inducción durante 3 h a 37 °C, o durante 16 h a 15 o 20 °C. Se realizaron estas pruebas a distintas temperaturas para evaluar posteriormente la más adecuada para maximizar la producción de proteína en forma soluble, lo cual suele verse favorecido a bajas temperaturas. Además en el momento de la inducción se suplementó el cultivo celular con 50  $\mu$ M ZnCl<sub>2</sub> para asegurar que la concentración de iones Zn<sup>2+</sup>, necesarios para el correcto plegamiento del dominio PHD, no era limitante. Las células sedimentadas y resuspendidas en PBS se rompieron mediante sonicación, se separó la parte soluble de la insoluble mediante ultracentrifugación y se evaluó la presencia de la proteína de interés mediante SDS-PAGE. Para todos los clones de ING4 e ING5 se encontró que la mayoría de la proteína se producía en forma insoluble a tres temperaturas probadas, siendo los niveles totales mayores a 37 °C, con lo que esta fue la temperatura utilizada en el escalado de la producción. En el caso la proteína GST-JADE1L\_dominio IIB, la temperatura óptima de expresión fue de 15 °C dando lugar a proteína en forma soluble.

### **1.1.18 Producción de proteínas en sistemas libres de células.**

Con el fin de producir las proteínas en forma soluble se utilizó un método de síntesis en extractos de células de *E. coli* adaptado de un protocolo descrito previamente (Pedersen et al, 2011). La principal modificación es que el extracto se hizo con células transformadas con un plásmido de expresión de dos chaperones más comunes en *E. coli*: las proteínas GroEL y GroES (Villate et al, 2012). De esta forma el extracto contenía unos niveles elevados de estas chaperonas y podrían facilitar la solubilización de las proteínas ING. Este sistema se podía utilizar con las proteínas ING dado que en el vector pET11d los genes están bajo el control del promotor del fago T7 y disponemos de RNA polimerasa del

fago T7 purificada en el laboratorio. Esencialmente el método consiste en añadir al extracto el vector con el gen (10 µg/mL), la RNA polimerasa del fago T7 (100 µg/mL) y una los reactivos necesarios para las reacciones enzimáticas de transcripción y traducción. Aunque se consiguió producir las proteínas mayoritariamente en forma soluble los niveles de expresión eran muy bajos. Por esta razón se abandonó esta estrategia y se continuó con la producción en bacterias.

### 1.1.19 Expresión de ING5 completa

Células de *E.coli* BL21(DE3) seleccionadas de la prueba de expresión fueron utilizadas para inocular 200 mL de medio LB suplementado con 1 µM de ampicilina. El cultivo creció toda la noche a 37 °C. Posteriormente los 200 mL fueron divididos en 4 tubos de 50 mL y centrifugados durante 15 minutos. Cada uno de los sedimentos obtenidos fue resuspendido en 1.5 litros de medio LB suplementado con 1 µM ampicilina a 37 °C e incubados en un agitador orbital a una velocidad de 220 rpm. La expresión de la proteína fue inducida en fase exponencial con 0.5 mM de IPTG cuando la OD<sub>600</sub> estaba aproximadamente entre 0.4 y 0.6. El crecimiento del cultivo fue interrumpido después de 3 h de la inducción a 37 °C. El sedimento bacteriano obtenido por la centrifugación del cultivo, fue resuspendido en tampón 20 mM Tris-HCl pH 8.0, inhibidores de proteasas (1 tableta de *Complete EDTA-free, Roche*) en un volumen en mL definido por la fórmula: 6 x (número de litros de cultivo) x (OD<sub>600</sub> final del cultivo). Las células se lisaron por sonicación durante 3 minutos en series de 20 s de sonicación y 60 s de descanso, manteniendo siempre la muestra en un baño de hielo-agua. Posteriormente, el extracto celular fue ultracentrifugado a 35,000 g durante 1.5 h. Después de la ultracentrifugación se confirmó mediante SDS-PAGE que la proteína se encontraba en la fracción insoluble.

### 1.1.20 Expresión de ING5<sub>1-105</sub> y mutantes puntuales Q33R, I68V y C75R

El procedimiento seguido fue esencialmente el mismo que el usado para la proteína completa.



### 1.1.21 Expresión de ING5<sub>1-105</sub> en medio suplementado con Selenometionina

Para este propósito se utilizaron células de *E.coli* de la estirpe B834, un auxótrofo de metionina. Después de transformar con el plásmido y de comprobar la expresión de la proteína a pequeña escala, se inoculó un volumen de 50 mL de medio LB suplementado con 1  $\mu$ M de ampicilina con una colonia. El cultivo creció durante la noche (unas 16 h) a 37 °C. Posteriormente los 50 mL fueron centrifugados durante 15 minutos. El sedimento obtenido fue resuspendido en 1 litro de medio suplementado selenometionina *SelenoMet*<sup>TM</sup> *Medium (molecular dimensions)*, preparado de la siguiente forma:

Se disolvieron 21.6 g de *SelenoMet Medium Base* en 1 litro de H<sub>2</sub>O destilada, esta mezcla se autoclavó. Aparte de disolvieron 5.1g de *SelenoMet Nutrient Mix* en 50 mL H<sub>2</sub>O destilada, se filtró (tamaño de poro 22  $\mu$ M) y se añadió al *SelenoMet Medium Base* preparado en el primer paso, después se le añadió al medio 1  $\mu$ M ampicilina, el cultivo creció a 37 °C en un agitador orbital a una velocidad de 220 rpm y la expresión de la proteína fue inducida en fase exponencial con 0.5 mM de IPTG cuando la OD<sub>600</sub> estaba aproximadamente entre 0.4-0.6. El resto del procedimiento fue el mismo que el descrito para la producción en medio LB.

### 1.1.22 Expresión de ING5 y de ING5<sub>1-105</sub> marcada isotópicamente en <sup>15</sup>N

Estas proteínas se produjeron con enriquecimiento en <sup>15</sup>N para llevar a cabo los experimentos de interacción con H3K4me3 y con la proteína de fusión GST-JADE1L\_dominio Iib. Partiendo de un sedimento de células obtenido de 3 L de cultivo en medio LB con una OD entre 0.5-0.6, se resuspendió en 1 L de medio mínimo enriquecido en <sup>15</sup>N. Este medio se prepara según se describe a continuación.

A 500 mL H<sub>2</sub>O ultrafiltrada se le añaden las siguientes disoluciones:

100 mL de sales M9 pH 7.2 10X (5 g NaCl, 30 g KH<sub>2</sub>PO<sub>4</sub>, 40 g, Na<sub>2</sub>HPO<sub>4</sub>·2H<sub>2</sub>O en agua)

2g D-glucosa (resuspendido en 25 mL de H<sub>2</sub>O ultrafiltrada)

1g 98% <sup>15</sup>NH<sub>4</sub>Cl (resuspendido en 25 mL H<sub>2</sub>O ultrafiltrada).

1 mL ampicilina 1mM

10 mL de D-Biotina 0.1 g/L

3 mL de Tiamina 1%

2 mL MgCl<sub>2</sub> 1 M

2 mL de solución 500X de elementos traza, la cual se prepara disolviendo en 1 L de agua las siguientes sales:

6 g FeSO<sub>4</sub>·7H<sub>2</sub>O

6 g CaCl<sub>2</sub>·2H<sub>2</sub>O

1.2 g MnCl<sub>2</sub>·4H<sub>2</sub>O

0.8 g CoCl<sub>2</sub>·6H<sub>2</sub>O

0.7 g ZnSO<sub>4</sub>·7H<sub>2</sub>O

0.3 g CuCl<sub>2</sub>·2H<sub>2</sub>O

0.02 g H<sub>3</sub>BO<sub>3</sub>

0.25 g (NH<sub>4</sub>)Mo<sub>7</sub>O<sub>24</sub>·4H<sub>2</sub>O

5.0 g EDTA-Na<sub>2</sub>(H<sub>2</sub>O)<sub>2</sub>

y añadiendo HCl al 37 % para bajar el pH hasta solubilizarlas.

El pH de la mezcla final se ajusta a 7.2 con NaOH, se afora la disolución hasta 1 L con H<sub>2</sub>O ultrafiltrada, y se esteriliza por filtración usando una membrana con tamaño de poro de 22 μM.

El cultivo bacteriano en el medio mínimo se creció a 25 °C en un agitador orbital a una velocidad de 220 rpm. Al cabo de una hora se indujo la expresión de la proteína con 0.5 mM de IPTG y se continuó creciendo el cultivo durante la noche (unas 16 horas). El resto del procedimiento es idéntico al descrito para la proteína producida en LB.

### **1.1.23 Expresión de ING5<sub>1-105</sub> marcada isotópicamente en <sup>2</sup>H, <sup>13</sup>C, <sup>15</sup>N**

Células de *E.coli* BL21 (DE3) seleccionadas de la prueba de expresión fueron utilizadas para inocular 10 mL de medio LB suplementado con 1 μM de ampicilina. El cultivo creció 8 h a 37 °C. Posteriormente los 10 mL fueron centrifugados durante 15 minutos. El sedimento celular obtenido fue resuspendido en 50 mL de medio mínimo M9 enriquecido isotópicamente en <sup>2</sup>H, <sup>13</sup>C y <sup>15</sup>N, y se creció durante la noche a 37 °C. Posteriormente el cultivo fue centrifugado y el sedimento obtenido fue resuspendido en 1 L del mismo medio. La preparación de este medio se realizó de forma similar al medio enriquecido en <sup>15</sup>N, con la diferencia de que se utilizó agua deuterada en todas las disoluciones, que la glucosa

utilizada estaba uniformemente enriquecida en  $^2\text{H}$  (con un nivel del 97 %, además del enriquecimiento en  $^{13}\text{C}$ ) y que se suplementó con 1 g/L de Celtone enriquecido en  $^2\text{H}$ ,  $^{13}\text{C}$  y  $^{15}\text{N}$ . Celtone es un producto comercial que contiene un extracto de algas enriquecido en estos isótopos, y fue comprado a *Cambridge Isotope Laboratories*, como el resto de reactivos con enriquecimiento isotópico.

El resto del protocolo de expresión es el mismo que el descrito para la proteína producida en LB.

### 1.1.24 Expresión de GST-JADE1L\_Dominio IIB

Células de *E.coli* BL21(DE3) fueron transformadas con el plásmido pGEX-4t3 que contienen el gen que codifica la secuencia de JADE1L\_dominio IIB. Con las colonias obtenidas de la transformación se realizaron pruebas de expresión, de las cuales se concluyó que la temperatura óptima de expresión es de 15 °C. La colonia seleccionada fue utilizada para inocular 100 mL de medio LB complementado con 1  $\mu\text{M}$  ampicilina, a 15 °C en un agitador orbital a una velocidad de 220 rpm. La expresión de la proteína fue inducida en fase exponencial con 0.5 mM de IPTG, cuando la  $\text{OD}_{600}$  estaba aproximadamente entre 0.4-0.6). El crecimiento del cultivo fue interrumpido después de 16 h de la inducción. El sedimento bacteriano obtenido por la centrifugación del cultivo, fue resuspendido en tampón 20 mM Tris-HCl pH 8.0, inhibidores de proteasa (1 tableta de *Complete EDTA-frees, Roche*) en un volumen definido por la fórmula: 6 x (número de litros de cultivo) x ( $\text{OD}_{600}$  final del cultivo). Las células se lisaron por sonicación durante 3 minutos en series de 20 s de sonicación y 60 s de descanso, manteniendo siempre la muestra en un baño de hielo-agua. Posteriormente el extracto celular fue ultracentrifugado a 35,000 g durante 1.5 h. Después de la ultracentrifugación se confirmó mediante SDS-PAGE, que la proteína se encontraba en la fracción soluble. Esta muestra se filtró (tamaño de poro de 22  $\mu\text{M}$ ) antes de iniciar el proceso de purificación.

### 1.1.25 Expresión de la proteasa TEV (*Tobacco Etch Virus*)

Esta proteasa se obtiene en el laboratorio a partir de un clon con el gen fusionado al de la proteína MBP (*maltose binding protein*). La secuencia de aminoácidos entre la MBP y la

proteasa TEV tiene un sitio de corte TEV y una cola de histidinas, de tal forma que la proteína de fusión producida (mayoritariamente soluble gracias a la MBP) es proteolizada por la propia TEV, la cual retiene la cola de histidinas en su extremo N-terminal. A partir de una colonia de células de *E.coli* Rosetta(DE3) seleccionada en una prueba de expresión, se inocularon 200 mL de medio LB suplementado con 1  $\mu$ M de ampicilina y 30 mg/L cloranfenicol (a partir de un stock 1000X en etanol). El cultivo creció toda la noche a 37 °C, se dividió en 4 tubos de 50 mL y se cosecharon las células por centrifugación. Cada uno de los sedimentos obtenidos fue resuspendido en 1.5 litros de medio LB con ampicilina y cloranfenicol a 37 °C e incubados en un agitador orbital a una velocidad de 220 rpm. La expresión de la proteína fue inducida en fase exponencial con 0.5 mM de IPTG, cuando la OD<sub>600</sub> estaba aproximadamente entre 0.4-0.6. El crecimiento del cultivo fue interrumpido después de 3 h de la inducción a 37 °C. El sedimento bacteriano obtenido por la centrifugación del cultivo, fue resuspendido en tampón 50 mM fosfato sódico pH 8.0, inhibidores de proteasas (una tableta de *Complete EDTA-free, Roche*) en un volumen en mL definido por la fórmula: 6 x (número de litros de cultivo) x (OD<sub>600</sub> final del cultivo). Las células se lisaron por sonicación durante 3 minutos en series de 20 s de sonicación y 60 s de descanso, manteniendo siempre la muestra en un baño de hielo-agua. Posteriormente el extracto celular fue ultracentrifugado a 40,000 g durante 1.5 h. Al sobrenadante se le añadió polietilenglicol a una concentración final de 0.1 % (peso/vol) para precipitar los ácidos nucleicos presentes en la disolución. La muestra fue agitada manualmente y sometida a una segunda ultracentrifugación a 15,000 g durante 30 minutos. Después de la ultracentrifugación se confirmó mediante SDS-PAGE que la proteína se encontraba en la fracción soluble y esta, se filtró (tamaño de poro de 22  $\mu$ m) antes de iniciar el proceso de purificación por cromatografía.

### 1.7 Purificación de proteínas

La purificación de las proteínas se realizó mediante técnicas cromatográficas de afinidad y exclusión molecular utilizando resinas empaquetadas en columnas y aparatos de cromatografía líquida de baja y media presión AKTA-Prime, AKTA-FPLC y AKTA-Purifier. Todas las columnas y aparatos de cromatografía eran del fabricante *General*

*Electric Healthcare*, salvo las columnas de Strep-Tactin, que eran de *Iba Solutions for Life Sciences*. En general las purificaciones expuestas en este trabajo se realizaron a 4 °C. El grado de pureza de las proteínas purificadas se evaluó por SDS-PAGE cargando exceso de muestra. Un ejemplo representativo se muestra en la Figura 11C. La integridad estructural y, en su caso, el nivel de enriquecimiento isotópico de las proteínas, se evaluó mediante espectrometría de masas (MALDI-TOF).

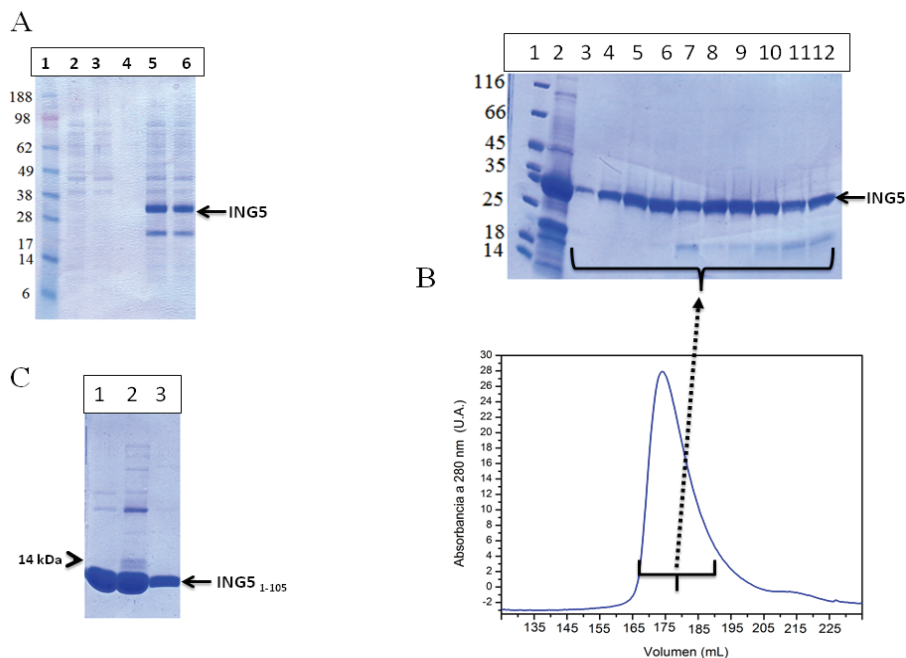
### **1.1.26 Purificación de la proteasa TEV (*Tobacco Etch Virus*)**

La proteína fue purificada mediante cromatografía de afinidad a níquel en columnas precargadas con Ni<sup>2+</sup> HisTrap de 5 mL. La columna fue previamente equilibrada en tampón de carga, (20 mM PO4 pH 8.0 200 mM NaCl, 10% glicerol, 1 mM DTT 25 mM imidazol). Una vez la muestra fue inyectada, la columna fue lavada con entre 6 y 10 volúmenes de columna (CV). Después de este lavado se aplicó un gradiente de elución de 20 CV para llegar al 100% del tampón de elución (el mismo que el de carga pero con 500 mM imidazol). La proteína eluyó alrededor del 30% del gradiente de elución. Las fracciones contenidas en el pico de elución de la cromatografía se juntaron y fueron concentradas hasta 1 g/L. Seguidamente se realizó un paso de tamizado molecular en una columna *Superdex* 200 26/60 en tampón 100 mM Tris-HCl pH 8.0, 500 mM NaCl. Las fracciones contenidas en el pico de elución fueron reunidas y concentradas por ultrafiltración en concentradores Amicon Ultra con membranas con límite de paso de 10 kDa hasta 1 g/L. Se añadió glicerol 50 % (vol/vol, concentración final), se dividió en alícuotas de 500 µL que fueron congeladas en nitrógeno líquido y almacenadas en el congelador a -80 °C. La concentración de la proteína fue determinada por absorbancia a 280 nm usando un coeficiente de extinción  $\epsilon = 32,220 \text{ M}^{-1} \cdot \text{cm}^{-1}$  asumiendo que las cisteínas están reducidas.

### **1.1.27 Purificación de ING5 completa**

La proteína fue extraída de los cuerpos de inclusión mediante solubilización con tampón 20 mM Tris-HCl pH 8.0 1 mM DTT con 8 M urea (1.5 mL de tampón por cada litro de cultivo crecido) y ultracentrifugación a 35,000 g durante 2 h. La proteína del sobrenadante replegó por dilución a 4 °C, gota a gota, en tampón 20 mM Tris-HCl pH 8.0, 300 mM

NaCl, 1mM DTT y 50  $\mu\text{M}$   $\text{ZnCl}_2$ , ya que los iones  $\text{Zn}^{2+}$  son necesarios para el correcto plegamiento del dominio PHD de la proteína. El factor de dilución utilizado fue de 50. La disolución obtenida del proceso de replegamiento, se filtró (tamaño de poro de 22  $\mu\text{m}$ ) y se inyectó en una columna HisTrap de 5 mL precargada con  $\text{Ni}^{2+}$  y previamente equilibrada en tampón 20 mM Tris-HCl pH 8.0 300 mM NaCl, 1 mM DTT. La columna fue lavada con entre 6 y 10 CV de tampón y se aplicó un gradiente de elución de 20 CV para llegar al 100% del tampón de elución (el mismo que el de carga pero con 500 mM imidazol). La proteína eluyó alrededor del 30% del gradiente de elución. Las fracciones contenidas en el pico de elución de la cromatografía se juntaron y fueron concentradas hasta 1 g/L. Seguidamente se realizó un paso de tamizado molecular (columna Superdex 200 26/60) en tampón 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 1mM DTT. En la Figura 11B se muestra el resultado de esta etapa de purificación por tamizado molecular. Las fracciones contenidas en el pico de elución fueron reunidas y concentradas por ultrafiltración en concentradores Amicon Ultra con membranas con límite de paso de 10 kDa hasta 1 g/L. Se dividió en alícuotas de 500  $\mu\text{L}$  que fueron congeladas en nitrógeno líquido y almacenadas en el congelador a  $-80\text{ }^\circ\text{C}$ . La concentración de la proteína fue determinada por absorbancia a 280 nm usando un coeficiente de extinción  $\epsilon = 24,410\text{ M}^{-1}\cdot\text{cm}^{-1}$  asumiendo que las cisteínas están reducidas.



**Figura 11** Resultados representativos de la expresión, purificación y grado de pureza de las proteínas ING5 e ING5<sub>1-105</sub>. (A) Gel de la prueba de expresión de ING5 completa. En la parte izquierda del gel están indicados las masas (en kDa) de los marcadores de peso molecular utilizados Carril 1: marcadores de peso molecular, Carriles 2 y 3: muestras de las colonias cultivadas antes de la inducción con IPTG, Carril 4: carril vacío, Carriles 5 y 6: muestras de los cultivos correspondientes a los carriles 1 y 2 después de inducción con IPTG a 37 °C durante 3 h. La banda sobreexpresada entre los marcadores de 28 y 38 kDa (marcada con una flecha) es la de ING5 (30.1 kDa). (B) Se muestra un fragmento del cromatograma de la purificación de ING5 mediante tamizado molecular en la columna Superdex 200 16/60 con el pico de elución de la proteína, así como un gel SDS-PAGE para analizar los resultados de la purificación. Carril 1: marcadores de peso molecular, Carril 2: muestra de la proteína antes de inyectar en la columna, Carriles 3 a 12: Muestras de fracciones colectadas en el pico de elución. En el gel se observa una banda predominante de ING5 además de un contaminante de bajo peso molecular (alrededor de 20 kDa). Durante el proceso de purificación de ING5 completa se observó precipitación de la muestra. (C) Gel mostrando el grado de pureza de diferentes preparaciones de ING5<sub>1-105</sub>. Carril 1: ING5<sub>1-105</sub> con selenometionina. Carril 2: ING5<sub>1-105</sub>, ING5<sub>1-105</sub> marcado isotópicamente en <sup>2</sup>H <sup>13</sup>C y <sup>15</sup>N. En la parte izquierda del gel se indica el lugar que usualmente ocupa el marcador molecular de 14 kDa en este tipo de gels.

El reporte de análisis MALDI/TOF de una muestra de 10µL del pico de elución de ING5 marcada con <sup>15</sup>N indicó una masa experimental de 30.454 kDa, que está de acuerdo con la masa teórica calculada para ING5 incluyendo la cola de histidinas y el marcaje uniforme en

$^{15}\text{N}$  (30.472 kDa). La diferencia de 18 Da está en el rango del error experimental de la técnica.

### 1.1.28 Purificación de ING5<sub>1-105</sub> y mutantes Q33R, I68V, C75R

Las proteínas mutantes se purificaron de la misma forma que la salvaje, según se describe a continuación. La proteína fue extraída de los cuerpos de inclusión mediante solubilización con tampón 20 mM Tris-HCl 1 mM DTT pH 8.0 con 8 M urea (1.5 mL de tampón por cada litro de cultivo crecido) y ultracentrifugación a 35,000 g durante 2 h. La proteína del sobrenadante se replegó por dilución a 4 °C, gota a gota, en tampón 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 1mM DTT (sin  $\text{ZnCl}_2$ , ya que no es necesario para el plegamiento de este dominio). El factor de dilución utilizado fue de 50. La disolución obtenida del proceso de replegamiento, se filtró (tamaño de poro de 22  $\mu\text{m}$ ) y se inyectó en una columna HisTrap de 5 mL precargada con  $\text{Ni}^{2+}$  y previamente equilibrada en tampón 20 mM Tris-HCl pH 8.0 300 mM NaCl, 1 mM DTT. La columna fue lavada con entre 6 y 10 CV de tampón y se aplicó un gradiente de elución de 20 CV para llegar al 100% del tampón de elución (el mismo que el de carga pero con 500 mM imidazol). La proteína eluyó alrededor del 30% del gradiente de elución. Las fracciones contenidas en el pico de elución de la cromatografía se juntaron, se estimó la concentración mediante absorbancia a 280 nm y seguidamente se realizó la reacción de corte de la cola de histidinas mediante proteasa TEV. La reacción se realizó con un ratio 1:30 TEV/ING5 (peso/peso) a 4 °C durante toda la noche (unas 16 h). Posteriormente se eliminó el exceso de imidazol de la reacción de corte mediante el paso por una columna de desalado (HiPrep 26/10 Desalting) en tampón 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 1mM DTT y de nuevo otro paso por la columna de afinidad a níquel equilibrada en el mismo tampón con el objetivo de separar la fracción cortada de la no cortada y la proteasa TEV. La fracción que no se unió a la columna de afinidad a níquel (mayoritariamente ING5 sin cola de histidina) fue objeto de un cambio de tampón (mediante la columna de desalado utilizada anteriormente) a 20 mM Tris-HCl pH 8.0, 1mM DTT con el fin de disminuir la fuerza iónica para posteriormente ser purificada mediante intercambio aniónico. Este se realizó en una columna Hi-Trap Q-HP de 5 mL equilibrada en el mismo tampón. Una vez la muestra fue cargada la columna se lavó con



entre 6 y 10 CV y se aplico un gradiente de elución de 20 CV para llegar al 100% del tampón de elución (el mismo que el de carga con 1 M NaCl). Las fracciones contenidas en el pico de elución de la cromatografía se juntaron y concentraron por ultrafiltración (Amicon Ultra con membranas de 10 kDa de límite de paso) hasta 5 mL y seguidamente se purificaron por tamizado molecular en una columna Superdex 75 16/60 y en tampón 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 1mM TCEP. Las fracciones contenidas en el pico de elución fueron reunidas y concentradas hasta 2 g/L. Se dividió en alícuotas de 500 µL que fueron congeladas en nitrógeno líquido y almacenadas en el congelador a -80 °C. La concentración de la proteína fue determinada por absorbancia a 280nm usando un coeficiente de extinción  $\epsilon = 8,940 \text{ M}^{-1} \cdot \text{cm}^{-1}$ , asumiendo que las cisteínas están reducidas. La precisión de la reacción de corte con proteasa TEV junto con la eficiencia de incorporación del marcaje isotópico con ( $^{15}\text{N}$  y  $^{13}\text{C}$ ), la incorporación de selenio-metionina fueron evaluadas mediante espectrometría de masas por MALDI-TOF. El volumen teórico de elución del dímero de ING5<sub>1-105</sub> en una columna Superdex 75 16/60 es de 78 mL, pero experimentalmente el volumen de elución es de 62 mL. Por lo cual el dímero de ING5<sub>1-105</sub> se comporta como si fuera una proteína de alrededor de 70 kDa probablemente debido a la forma elongada del dominio N-terminal de ING5<sub>1-105</sub>. El reporte de análisis MALDI/TOF de una muestra de de ING5<sub>1-105</sub> 10µL del pico de elución de la purificación mediante tamizado molecular indicó que la masa determinada experimentalmente (12.564 kDa) está de acuerdo con la masa teórica calculada para ING5<sub>1-105</sub> (12.545 kDa). Estos valores exponen una diferencia de 20 Da entre sí, que está en el rango del error experimental de la técnica.

### **1.1.29 Purificación de ING4 completa/ING4<sub>2-107</sub>**

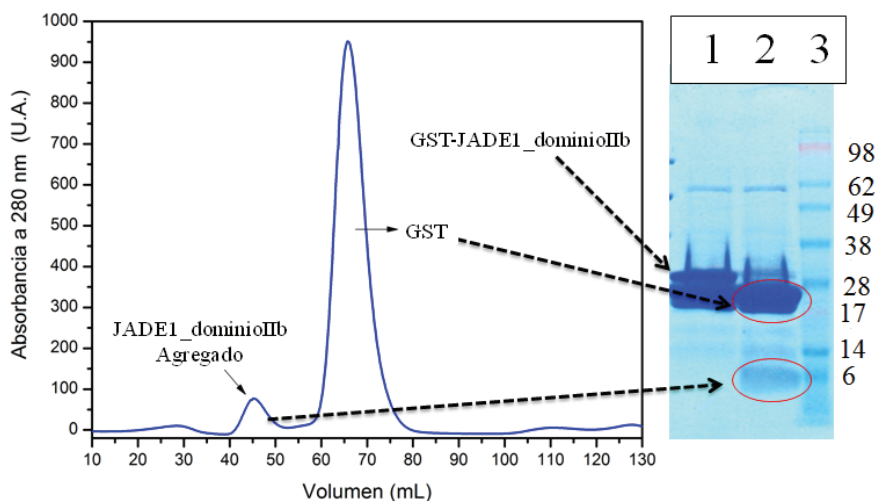
Estas proteínas fueron expresadas de la misma forma sus homólogas ING5 e ING51-105, encontrándose también en la fracción insoluble después de lisado de las células.

Las proteína fueron extraídas de los cuerpos de inclusión mediante solubilización con tampón 20 mM Tris-HCl 1 mM DTT pH 8.0 con 8 M urea (1.5 mL de tampón por cada litro de cultivo crecido) y ultracentrifugación a 35,000 g durante 2 h. La proteína del sobrenadante se replegó por dilución a 4 °C, gota a gota, en tampón 20 mM Tris-HCl pH

8.0, 300 mM NaCl, 1mM DTT (con o sin  $ZnCl_2$  en el caso de ING4 o ING4<sub>2-107</sub>, respectivamente). El factor de dilución utilizado fue de 50. Proteínas fueron purificadas mediante cromatografía de afinidad a Strep usando columnas con resina Strep-Tactin de 5mL. La columna fue previamente equilibrada en tampón de carga (100 mM Tris-HCl, pH 8.0, 150 mM NaCl), una vez la muestra fue cargada la columna fue lavada con 5 CV. Después de este lavado se aplicó un 100% del tampón de elución (100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2.5 mM Destiobiotina). La proteína eluyó alrededor de 1.5 CV. Las fracciones contenidas en el pico de elución de la cromatografía se juntaron se concentraron por ultrafiltración hasta 5 mL para ser purificadas mediante tamizado molecular. ING4 completa en la columna Superdex 200 16/60 e ING4<sub>2-107</sub> en la columna Superdex 75 16/60 en tampón 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 1mM DTT. La concentración de la proteína fue determinada por absorbancia a 280 nm usando un coeficiente de extinción  $\epsilon = 29,910 \text{ M}^{-1} \cdot \text{cm}^{-1}$ . Para ING4 completa y de  $\epsilon = 12,950 \text{ M}^{-1} \cdot \text{cm}^{-1}$  para ING4<sub>2-107</sub>, asumiendo que las cisteínas están reducidas. El reporte de análisis MALDI/TOF de una muestra de 10 $\mu$ L del pico de elución de la purificación de mediante tamizado molecular indicó que la masa determinada experimentalmente (13.769 kDa) está de acuerdo con la masa teórica calculada para ING4<sub>2-107</sub> (13.754 kDa, incluyendo la cola Strep). La diferencia de 15 Da está dentro del rango del error experimental de la técnica.

### 1.1.30 Purificación de GST-JADE1L\_Dominio IIb

La fracción soluble del lisado celular con la proteína GST-JADE1L\_Dominio IIb se inyectó en una columna de afinidad a glutatión GST-Trap de 5 mL equilibrada en 20 mM Tris-HCl 150 mM NaCl pH 8.0. Se lavó al columna con 5 CV y se eluyó la proteína con un cambio de tampón igual al de carga pero con 10 mM glutatión. El pico de elución fue reunido y sometido a una reacción de corte con la proteasa trombina utilizando 2 unidades de la proteasa por mg de proteína de fusión, durante toda la noche a 4 °C. La mezcla de reacción se separó por cromatografía de tamizado molecular en una columna Superdex 75 16/60. La región del cromatograma y el análisis por SDS-PAGE se muestra en la Figura 12.



**Figura 12** Análisis por SDS-PAGE y separación por cromatografía de los productos de la reacción de corte de la proteína de fusión GST-JADE1L\_dominio IIb con trombina. En el carril 1 del gel se muestra la proteína de fusión GST-JADE1L\_dominio IIb antes de la reacción de corte. En el carril 2 se puede observar que la proteína de fusión ha sido cortada, ya que aparecen dos bandas de menor peso molecular que se corresponden con el peso molecular teórico de JADE1L\_dominio IIb (6.7 kDa) y GST (26.9 kDa). El cromatograma de tamizado molecular de la reacción de corte, indica que JADE1L\_dominio IIb se eluyó en el volumen de exclusión de la columna (45 mL), lo cual sugiere que está agregado. El pico principal (65 mL) corresponde a la proteína GST. Carril3: Marcadores de peso molecular (con su peso en kDa indicado a la derecha).

El reporte de análisis MALDI-TOF de una muestra del pico de elución a 45 mL indicó una masa molecular de 6.775 kDa, que concuerda con el peso teórico de 6.770 kDa de JADE1L\_dominio IIb. La diferencia de 5 Da está en el rango del error experimental de la técnica. La muestra del pico de elución a 65 mL dio una masa de 26.96 kDa la cual se corresponde con el peso molecular teórico de GST (26.90 kDa), confirmando que se trata de GST sin el fragmento de JADE1L. Estos resultados y el patrón de bandas observado en el sobrenadante de las células confirman que antes de la purificación la proteína predominante es GST, con una proporción menor de la proteína de fusión. Las causas de esto pueden ser dos, degradación proteolítica dentro de la bacteria o parada prematura de la traducción debido a codones muy poco abundantes en *E.coli*. Los rendimientos extremadamente bajos obtenidos de JADE1L\_dominio IIb, y su estado aparentemente

agregado nos obligó a realizar los experimentos de interacción con ING5<sub>1-105</sub> directamente con la proteína obtenida después de la cromatografía de afinidad a GST, la cual es una mezcla de GST sola y proteína de fusión.

### 1.8 Formación de Heterodímeros ING4/ING5

Con el fin de detectar y purificar el posible heterodímero ING4/ING5, se diseñó un experimento de co-purificación por cromatografía de afinidad secuencial usando colas ortogonales (cola Strep en el extremo N-terminal de ING4 y cola de histidina en el extremo N-terminal de ING). Con este objetivo, muestras de proteínas Strep-ING4 y His-ING5 solubilizadas en urea se mezclaron en cantidad equimolar antes replegarlas. Si se forman los heterodímeros la purificación por cromatografía de afinidad secuencial con colas ortogonales permitirá detectar (en SDS-PAGE y/o por *western blot* con anticuerpos anti-His o anti-Strep) y separar el heterodímero Strep-ING4/His-ING5. Este diseño había funcionado muy bien previamente en el laboratorio para purificar heterodímeros de la meganucleasa I-CreI (Redondo et al, 2008). Con este fin se llevaron a cabo dos tipos de experimentos. Inicialmente los extractos en urea obtenidos de la expresión de las proteínas ING4 e ING5 completas se mezclaron en cantidades aproximadamente equimolares según la intensidad de las bandas correspondientes observadas en SDS-PAGE. Seguidamente la mezcla se replegó por dilución a 4 °C, gota a gota, en tampón 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 1mM DTT y un factor de dilución de 50. La disolución se clarificó por filtración y se procedió a la purificación secuencial mediante columna de afinidad a Ni<sup>2+</sup> y de afinidad a Strep-tactin (o viceversa). En el segundo tipo de experimentos se partió de los dominios purificados His-ING5<sub>1-105</sub> y Strep-ING4<sub>2-107</sub> se mezclaron cantidades equimolares a una concentración de 1 g/L en 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT, y se añadió urea sólida hasta 8 M para desnaturalizar los dominios. La mezcla se replegó (como se ha descrito más arriba) por dilución a 4 °C en volumen 1:50 en tampón (20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT), posteriormente la muestra fue concentrada por ultrafiltración (usando concentradores Amicon Ultra de 3 kDa de límite de paso) hasta 5 mL y se repurificó mediante los dos pasos consecutivos descritos. La presencia de cada una de las dos proteínas en las diferentes fracciones cromatográficas se evaluó mediante SD-PAGE y *western blotting* de cromatografías de afinidad: afinidad a

strep-tactin (primer paso). La elución de esta cromatografía se inyectó en una columna de afinidad a  $\text{Ni}^{2+}$  (segundo paso). Los pasos de repurificación se realizaron también en el orden: Afinidad a afinidad a  $\text{Ni}^{2+}$  (primer paso), afinidad a strep-tactin (segundo paso).

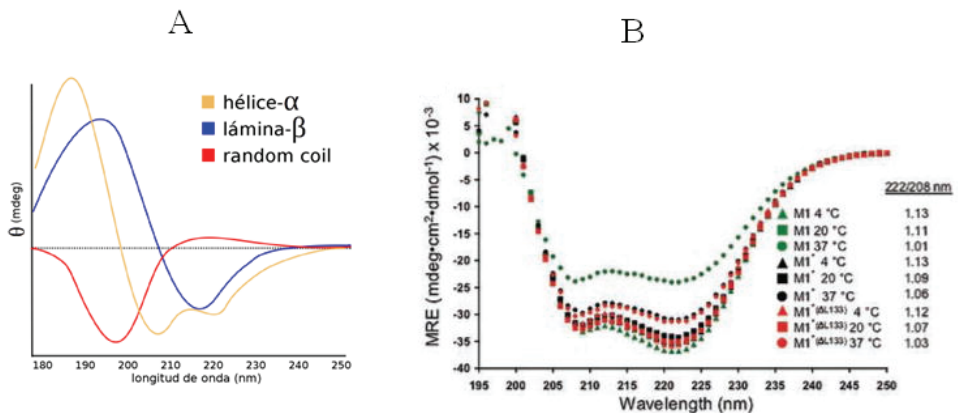
### 1.9 Cuantificación de proteínas en disolución por absorbancia ultravioleta.

Todas las proteínas analizadas contienen en su secuencia triptófanos y/o tirosinas facilitando la medida de la concentración mediante absorción de luz ultravioleta. A partir de la secuencia de aminoácidos de la proteína y usando la herramienta web, se obtuvo el coeficiente de extinción molar a 280 nm, que fue utilizado para el cálculo de la concentración de las disoluciones de proteína a partir de medidas de absorbancia en espectrofotómetros Jasco (usando cubetas de cuarzo de 1 cm de paso óptico) o Nanodrop (sin cubeta, en una gota de 2  $\mu\text{L}$ ). La concentración se calcula a partir de la ley de *Lambert-Beer*:  $A = \epsilon \times c \times l$ , donde  $l$  es el paso óptico de la cubeta en cm,  $c$  la concentración molar de la proteína y  $\epsilon$  el coeficiente de extinción molar calculado.

### 1.10 Espectroscopia de Dicroísmo Circular

Esta técnica permite obtener información estructural de moléculas ópticamente activas (quirales), midiendo la diferencia generada en un experimento de CD entre las ondas de luz polarizadas circularmente hacia la izquierda ( $\epsilon_L$ ) y las ondas de luz polarizadas circularmente hacia la derecha ( $\epsilon_D$ ), donde  $\epsilon$  es el coeficiente de absorptividad molar en ( $\text{M}^{-1}\text{cm}^{-1}$ ), que se originan debido a la asimetría estructural (moléculas quirales). En un experimento de CD se mide la diferencia de absorción entre las dos componentes circularmente polarizadas de la luz, la cual se puede expresar como elipticidad ( $\theta$ , en grados) o, en el caso de proteínas, como elipticidad molar por residuo (MRE o  $[\theta]$ , en  $\text{grados}\cdot\text{cm}^2\text{dmol}^{-1}$ ). En la región del ultravioleta lejano (180-250 nm) son las bandas de absorción de los enlaces peptídicos los que principalmente contribuyen al espectro de dicroísmo circular de las proteínas. Las estructuras secundarias de proteínas determinan la orientación relativa de los enlaces peptídicos, y por esta razón el DC puede ser usado para estimar la estructura secundaria (Norma, 1999). Las estructuras en hélice- $\alpha$  lamina- $\beta$  así como la cadena desestructurada y flexible (*random coil*) tienen espectros característicos de

DC en el UV lejano, como se muestra en la Figura 13A. Las estructuras en hélice alfa presentan espectros de DC con dos mínimos (y elipticidad negativa) a 222 y 208 nm con un ratio  $\theta_{222}/\theta_{208} < 1$ , y con un máximo (y elipticidad positiva) a 193 nm. En el caso de la conformación de hélices enrolladas (*coiled coil*) exponen igualmente estos extremos pero el ratio característico  $\theta_{222}/\theta_{208} > 1$  (McNamara et al, 2008). En general el espectro de DC de una proteína está predominantemente determinado por la suma de los espectros de sus elementos de estructura secundaria que contienen. El espectro de DC de una proteína representativa de estructura de hélices enrolladas (*coiled coil*) se muestra en la Figura 13B. Se trata de la proteína M1 del grupo A de *Streptococcus*, y el espectro muestra el característico ratio  $\theta_{222}/\theta_{208} > 1$  a diferentes temperaturas.



**Figura 13** Espectros de DC de proteínas. (A) Espectros de DC para los dos tipos de estructura secundaria más abundantes en proteínas y para una cadena polipeptídica desestructurada (*random coil*) asumiendo una población del 100 %. (B) Espectros de DC de la proteína M de *Streptococcus* grupo A y algunos mutantes, colectados a diferentes temperaturas. A la derecha de los espectros se muestra el ratio  $\theta_{222}/\theta_{208} > 1$ , indicando que a diferentes temperaturas la proteína M mantiene la estructura *coiled coil* (McNamara et al, 2008).

En los experimentos llevados a cabo en esta tesis los espectros de CD fueron registrados como el promedio de 15 barridos a intervalos de 1 nm y 4 s de tiempo promedio por cada punto, entre 192 y 250 nm a 25 °C en un espectropolarímetro JASCO J-810 equipado con un control termoeléctrico de temperatura. Las muestras fueron preparadas en tampón 20 mM fosfato sódico pH 6.5, 200 mM de NaCl, 1 mM DTT, a una concentración de 50  $\mu$ M

en cubeta de 0.1 mm de paso óptico al registrar los espectros de dicroísmo circular, y a una concentración de 25  $\mu$ M utilizando una cubeta de cuarzo de 2 mm de paso de óptico y tapón de teflón al registrar las curvas de desnaturalización térmica.

### **1.11 Dispersión de luz a múltiples ángulos MALLS (Multi-Angle Laser Light Scattering)**

La dispersión de luz es una técnica que se utiliza principalmente para obtener información sobre el comportamiento hidrodinámico de proteínas en solución y de la cual se puede derivar el peso molecular medio y por tanto su estado de oligomerización. Esta técnica se basa en la medición de la intensidad de la dispersión estática causada por la muestra cuando es irradiada por un rayo de luz láser en función del ángulo, analizando la intensidad de la luz dispersada a diferentes ángulos con respecto al de la luz incidente. Este análisis permite el cálculo de parámetros como el coeficiente de difusión, el radio de giro molecular (distancia media de cada parte de la partícula corregida por su masa al centro de gravedad de esta), el radio hidrodinámico (que representa el radio de la partícula esférica con igual coeficiente de difusión). En este trabajo utilizamos la técnica MALLS porque permite determinar la masa molecular de proteínas en disolución independientemente de su forma molecular (Folta-Stogniew & Williams, 1999), y se utilizó acoplado la salida de una columna de tamizado molecular analítica al detector de MALLS (lo que se conoce como experimento SEC-MALLS). La cromatografía de tamizado molecular se realizó a temperatura ambiente, usando una columna Superdex 200 10/300 GL en 20 mM Tris pH 8.0, 300 mM NaCl 1mM TECP 0.01 %  $\text{NaN}_3$  montada en un cromatógrafo AKTA Explorer. La salida de la columna se acopló a un detector de dispersión de luz DAWN HELEOS II (*WYATT Tech. Corp*). El sistema fue previamente calibrado usando albúmina de suero bovino (BSA) a 1 g/L. Muestras puras de proteínas de ING5<sub>1-105</sub> (salvaje y mutantes) e ING5 completa a las cuales se les cambió el tampón mediante una columna PD10 al tampón usado en la cromatografía se inyectaron a una concentración entre 1 y 2 g/L y se cromatografiaron a un flujo de 0.5 mL/min. La adquisición de datos y sus respectivos análisis fueron realizados usando el programa ASTRA (v 5.3.4.15, *WYATT Tech. Corp*). La masa molecular de las diferentes proteínas fueron calculados usando un

valor del cambio de índice de refracción con la concentración de proteína ( $dn/dc$ ) de 0.183 mL/g.

### **1.12 Cristales de proteínas**

Un cristal se puede considerar como una red imaginaria de unidades elementales idénticas, denominadas cada una de ellas celdilla unidad. La repetición por translación en las tres direcciones del espacio de este elemento estructural genera el volumen del cristal. La celdilla a su vez presenta en su interior una unidad mínima, denominada unidad asimétrica. Mediante la aplicación de ciertas operaciones de simetría a la unidad asimétrica, se genera el contenido de la celdilla unidad y en consecuencia, el de todo el cristal. Dicha celdilla unidad está definida por tres ejes (a, b, c) y tres ángulos inter-axiales ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) que se usan como sistema de referencia para las coordenadas atómicas. La combinación de la celdilla con los elementos de simetría, designan el grupo espacial que describe el patrón de repetición que presenta una o varias moléculas (unidad asimétrica) en el espacio tridimensional. Los cristales de proteína difieren mucho de los cristales de moléculas orgánicas o sales inorgánicas ya que las características de los primeros están condicionadas por la propia naturaleza de estas macromoléculas. Debido al tamaño de las proteínas y a su amplia esfera de solvatación, el volumen de la celdilla unidad suele ser bastante grande. Como las macromoléculas se empaquetan con baja densidad, al ordenarse en el cristal dejan grandes canales constituidos por moléculas de agua desordenadas y el contenido en solvente suele ser elevado (50-80%) (Matthews, 1968). Como consecuencia, este tipo de cristales son generalmente frágiles y sensibles a cambios en el ambiente. Al estar formados por moléculas quirales, se describen exclusivamente con 65 grupos espaciales (de los 230 posibles), es decir, sólo son posibles aquellos en los que no existen las operaciones de repetición que impliquen centros o planos de simetría, planos de deslizamiento o ejes quirales.

#### **1.1.31 Principios básicos de la cristalización**

El punto de partida de un estudio estructural por cristalografía es la obtención de un monocristal y para llegar a este es necesario disponer de una muestra de proteína pura y conformacionalmente homogénea. Durante el proceso de cristalización, las moléculas en

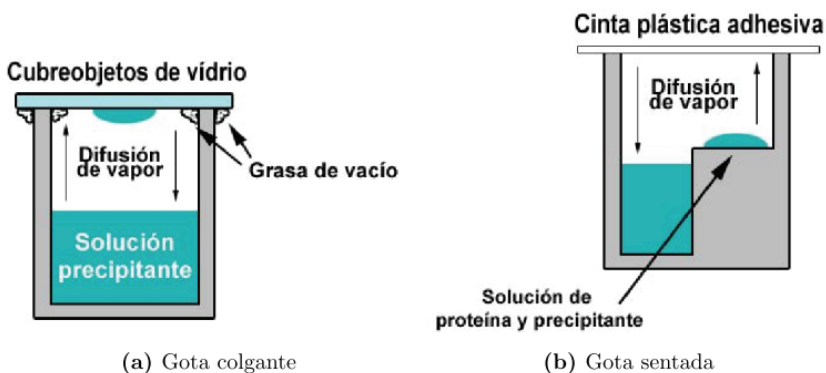


solución, en presencia de agentes precipitantes suaves, son conducidas a un estado de sobresaturación (termodinámicamente inestable) que posteriormente deriva en la aparición de una fase cristalina con las moléculas en equilibrio dinámico entre las fases sólida y líquida (Ducruix, 1992). El proceso de cristalización se puede dividir en dos fases: la fase de nucleación (formación de los primeros agregados ordenados) y la fase de crecimiento del cristal (en donde el conjunto de interacciones atractivas entre las moléculas son máximas, y las repulsivas mínimas). En el caso de las proteínas, la superación excesiva del punto de saturación, sobre todo si sucede rápidamente, conlleva con gran frecuencia la precipitación de la proteína en forma de agregados amorfos, no ordenados, que no son adecuados para los estudios de difracción. Por el contrario, si la sobresaturación es leve y se alcanza un estado metaestable ó zona ideal para el crecimiento de cristales. Por tanto, la formación de cristales requiere alcanzar lentamente un grado limitado de sobresaturación, para que se formen pocos núcleos, de manera que la masa proteica se invierta en el crecimiento de los cristales y así éstos aumenten su tamaño. En la búsqueda de las condiciones de cristalización de proteínas se consideran aspectos importantes como: concentración, pureza, homogeneidad y estabilidad. Dado que hay que alcanzar la sobresaturación, es esencial conseguir que la proteína posea una concentración inicial elevada. Se considera adecuado conseguir concentraciones en el rango 5-20 g/L, aunque no hay que descartar usar concentraciones menores si la solubilidad de la proteína no permite alcanzarlas mayores ya que lo importante es sobresaturar la solución durante el experimento. La pureza de la proteína debe ser tan alta como sea posible, comúnmente que sea al menos del 95 %. Las impurezas pueden limitar el tamaño de los cristales, ya que su interposición puede abortar el crecimiento del cristal, y/o introducir irregularidades que disminuyan la calidad del cristal obtenido. En cualquier caso, no es imposible obtener cristales con purezas relativamente bajas. La heterogeneidad de la muestra es desfavorable para la cristalización. Esta heterogeneidad puede ser tanto intrínseca (isoformas, fosforilación, nitrosilación, grados diversos de glicosilación, formas alternativas de procesamiento, etc), como resultado de la manipulación experimental (grados distintos de oxidación, proteólisis, agregación en grado variable, desnaturalización parcial, etc). Hay que evitar durante la preparación de la proteína someterla a manipulaciones que puedan

introducir heterogeneidad, tales como tratamientos térmicos. Las condiciones de cristalización deben preservar al máximo la estabilidad de la proteína durante el tiempo que lleve el proceso de cristalización. Sin embargo, salvo que la proteína se desnaturalice y precipite muy rápidamente, el conocimiento de que la proteína tiene una estabilidad limitada no es necesariamente una razón para descartar pruebas de cristalización.

### 1.1.32 Técnicas de cristalización

La más común se basa en alcanzar la sobresaturación mediante difusión de vapor. Una pequeña gota que contiene la solución de proteína y la de cristalización, se equilibra contra un reservorio de mucho mayor volumen y que contiene una solución de cristalización a una concentración superior que en la gota. El equilibrio se consigue por difusión de las especies volátiles (agua o solvente orgánico), hasta que las concentraciones de precipitantes en gota y reservorio son iguales. Existen principalmente dos modalidades de esta técnica. En la modalidad de la gota colgante como se muestra en la Figura 14a. La gota se deposita sobre un cubreobjetos siliconizado, que luego se invierte y se utiliza para sellar un pocillo de una placa multipocillo, de modo que la gota quede dentro del recinto sellado, separada de la solución de cristalización presente en el pocillo. En la segunda alternativa expuesta en la Figura 14b. La gota se asienta sobre una elevación plana o cóncava situada por encima del nivel de la solución de cristalización. Esta última aproximación se suele llevar a cabo en placas de 96 pocillos que se sellan con una lámina flexible de plástico adhesivo y transparente.



**Figura 14** Cristalización de proteínas mediante difusión de vapor.

### 1.13 Cristalización de ING5<sub>1-105</sub>

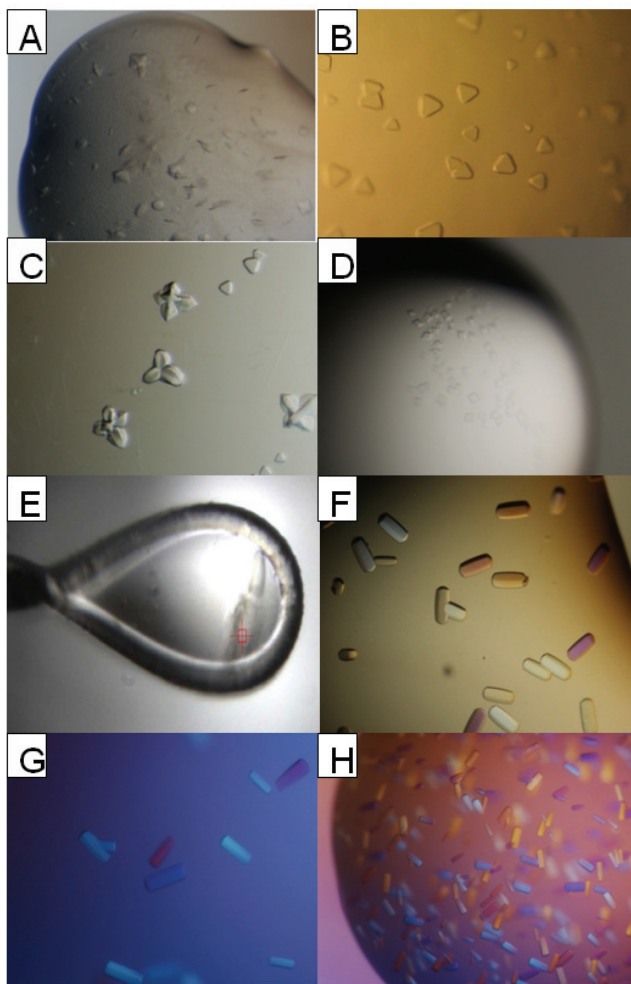
La búsqueda de las condiciones iniciales de cristalización se abordó a 21°C utilizando el método de difusión de vapor de la gota sentada en placas de 96 pocillos MRC (*Molecular Dimensions*) con nanogotas de 0.1 µL mas 0.1 µL de solución de reservorio. Para mezclar las microgotas de proteína y solución de condición de cristalización se utilizó el sistema MOSQUITO. La concentración de ING5<sub>1-105</sub> utilizada fue 20 g/L. Un factor determinante en la obtención de cristales de ING<sub>1-105</sub> fue la presencia de TCEP en la muestra de proteína pura, ya que durante los intentos de reproducir los primeros cristales obtenidos observó que ING5<sub>1-105</sub> no cristalizaba en presencia de DTT, . Todas las formas cristalinas obtenidas provenían de muestras cuyo agente reductor fue TCEP. El TCEP tiene un mayor poder reductor, y es más hidrofílico, y estas propiedades podrían ser las que favoreciesen la cristalización de ING5<sub>1-105</sub>. Los 11 cribados comerciales iniciales fueron: *PACT*, *NR-LBD*, *ProPlex*, *PEG ion 1/2*, *Strutural Screen1/2*, *JCSG-plus*, *JCSGI*, *Index (Molecular dimensions)* y *crystal\_screen (Hampton)*. Los mejores formas cristalinas se obtuvieron en la condición 59 (0.1 M MES pH 5.0, 20% (v/v) MPD) del cribado *JCSGI* (Tabla 4).

**Tabla 4** Resumen de las condiciones de cristalización de ING5<sub>1-105</sub>

Proteína	Condición inicial de cristalización	Condición optimizada	Toma de datos de difracción
ING5 <sub>1-105</sub>	JCSG 1, 59 (0.1MMES pH 5.0, 20% (v/v) MPD)	1. MMES pH 6.5, 22% MPD, 5% isopropanol	Cristal H10 3.3 Å /ID23-2 ESRF, Grenoble
			Cristal H11 3.3 Å /ID23-2 ESRF, Grenoble-

Proteína	Condición inicial de cristalización	Condición optimizada	Toma de datos de difracción
		0.1 M MES pH 6.5, 22% MPD, 5% PEG 3350	3.2 Å/ XALOC AIBA
			3.5 Å/ CICbioGUNE
		22% MPD, 0.1 M MES pH 6.5, 5% Dioxano	No difracción
ING5 <sub>1-105</sub> Selenometionina		0.1 M MES pH 6.5, 28% MPD, 5% isopropanol	No difracción

Los cristales obtenidos presentaban una forma irregular en sus bordes y un tamaño pequeño, por lo cual se decidió optimizar esta condición agregando a estas diferentes tipos de aditivos, de los cuales se obtuvieron diferentes tipos de formas cristalinas. Esta condición fue optimizada a (0.1 M MES pH 6.5, 22% MPD, 5% isopropanol) y se hicieron placas con gotas de 1  $\mu$ L de proteína a 20 g/L y 1  $\mu$ L de solución de reservorio, de esta placa se obtuvieron a los 2 días cristales grandes de forma ovalada. Igualmente con la condición (22% MPD, 0.1 M MES pH 6.5, 5% Dioxano) de esta placa se obtuvieron cristales grandes con formas de pirámide y de flor, como se muestra en la Figura 15B y C. Estos cristales fueron probados en el difractor del CIC bioGUNE X8-Proteum System (*Bruker*), obteniéndose únicamente difracción a 3.5 Å en los cristales de forma ovalada, como se muestra en la Figura 15F. Muestras de cristales de las tres formas fueron colectados de sus respectivas placas con un lazo de nailon (*cryoloop*) y fueron incubados unos pocos segundos en la solución crioprotectora correspondiente (22% MPD, 0.1 M MES pH 6.5, 5% Isopropanol ó Dioxano, 20% Etilenglicol) y posteriormente congelados en nitrógeno líquido para su almacenamiento y transporte. El proceso de cristalización de ING5<sub>1-105</sub> suplementado con selenometionina fue poco exitoso, ya que solo se logró obtener una forma cristalina en la condición 0.1 M MES pH 6.5, 28% MPD, 5% isopropanol, (Figura 15E), y estos cristales no difractaron.



**Figura 15** Representación de las interferencias constructiva (izquierda) y destructiva Formas cristalinas de ING51-105. (A) Cristales obtenidos en la condición de cristalización 59 del cribado JCSG 1. Esta condición se optimizó agregando diferentes tipos de aditivos y se obtuvieron cristales en las siguientes condiciones: (B) formas cristalinas de ING5<sub>1-105</sub> en 0.1 M MES pH 6.5, 20% MPD, Dioxano 3 -10% (C-D) Formas cristalinas de ING5<sub>1-105</sub> obtenidas a partir de la optimización de la solución 59 de JCSG 1. La nueva condición es 0.1M MES pH 6.5, 18 y 24% MPD, 2-6% PEG 3350. (E) Formas cristalinas de ING5<sub>1-105</sub> con selenometionina en 0.1 M MES pH 6.5, 30% MPD. Detalle del cristal de ING5<sub>1-105</sub> con selenometionina antes de ser irradiado con los rayos X (estos cristales no exhibieron difracción). (F) Formas cristalinas de ING5<sub>1-105</sub> obtenidas a partir de la optimización de la solución 59 de JCSG 1. La nueva condición es 0.1M MES pH 6.5, 20% MPD, Isopropanol 4% (se obtuvieron datos de difracción de estos cristales en el sincrotrón ESRF-Grenoble, línea ID23-2). (G-H) Formas cristalinas de ING5<sub>1-105</sub> obtenidas a partir de la optimización de la solución 59 de JCSG 1, la nueva condición es 0.1M MES pH 6.5, MPD 20-24%, PEG3350 2-4%. Se obtuvieron datos de difracción de estos cristales en el sincrotrón ALBA-Barcelona, línea XALOC.

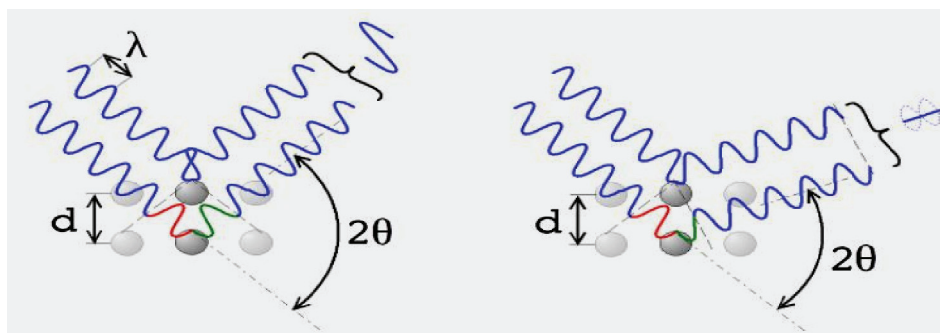
### **1.1.33 Crioprotección de los cristales**

El deterioro de los cristales de proteína al ser expuestos a la radiación se debe a factores como la longitud de onda penetrante y la intensidad del haz, unidas a prolongados tiempos de exposición, provocan fenómenos de absorción que hacen que la muestra se caliente, generándose radicales libres en las moléculas de proteína y solvente, que, con el tiempo, dan lugar a daños irreparables en el cristal que redundan en la pérdida de orden interno. Por este motivo es conveniente enfriar el cristal durante el proceso de difracción (Garman, 2003), mediante un flujo laminar de nitrógeno a una temperatura entre  $-153\text{ }^{\circ}\text{C}$  y  $-173\text{ }^{\circ}\text{C}$ . Para que no se formen cristales de hielo que dañen el cristal, es necesario utilizar sustancias crioprotectoras (McFerrin, 2002), como el glicerol, metilpentanodiol, etilenglicol, etanol, polietilenglicol, sacarosa, etc, que generan un estado vítreo alrededor del cristal. Sin embargo, estas sustancias pueden alterar también el cristal, dañando el orden interno, o disolviéndolo en el crioprotector. Por este motivo, hay que elegir apropiadamente el agente crioprotector, su concentración y el tiempo de permanencia del cristal en la solución crioprotectora. El procedimiento seguido en nuestro caso fue realizar durante unos segundos la inmersión del cristal en la solución crioprotectora antes de colocarlo, a través de un lazo de nailon que se monta mediante un imán en la cabeza goniométrica, bajo el flujo laminar de nitrógeno a baja temperatura. Se realizaron pruebas con diferentes soluciones crioprotectoras (paratona, glucosa 20%, DMSO, etilenglicol 20%, glicerol 20%, sacarosa, MPD, PEG 400, propilenglicol) con el fin de optimizar los experimentos de difracción. La solución crioprotectora seleccionada, debido a que fue en la que se alcanzó la mejor resolución de difracción en el difractor del CICbioGUNE fue la mezcla de solución madre con un 20% de etilenglicol.

### **1.14 Difracción de rayos X**

La difracción de rayos X es el fenómeno físico fundamental a través del cual se manifiesta la interacción de los rayos X con la materia cristalina. Cuando un haz de rayos X incide sobre un cristal, éste interacciona con los electrones de los átomos que componen el cristal, haciéndolos vibrar acopladamente con las variaciones periódicas de su campo eléctrico.

Los electrones oscilantes se convierten, de esta manera, en focos de nueva radiación X que se emite en todas las direcciones del espacio, fenómeno que se denomina “dispersión”. Estos rayos X dispersados por los electrones interfieren entre sí, pudiendo dar interferencias destructivas. Sin embargo, como los átomos en un cristal están ordenados de un modo regular y periódico, también ocurre que en determinadas direcciones, las ondas interfieren constructivamente, reforzándose y dando lugar al fenómeno de dispersión cooperativa o difracción. Este fenómeno se interpreta considerando que los rayos X incidentes son reflejados por una serie de planos paralelos imaginarios que intersectan los diferentes átomos del cristal (Figura 16). Cada familia de planos, separados por una distancia  $d$ , se identifica por tres números enteros ( $hkl$ ) denominados índices de Miller. Cuando el ángulo  $\theta$  de incidencia del haz de rayos X sobre uno de estos planos cumple la Ley de Bragg ( $n\lambda = 2d\sin\theta$ , siendo  $\lambda$  la longitud de onda y  $n$  un número entero), la diferencia de camino recorrido entre los haces reflejados por dos átomos en planos contiguos es un múltiplo entero de la longitud de onda, por lo que salen en fase y se produce una interferencia constructiva.



**Figura 16** Representación de las interferencias constructiva (izquierda) y destructiva (derecha) de acuerdo a la ley de Bragg.

El hecho de utilizar radiación X se debe a que el fenómeno de difracción requiere que la longitud de onda de la radiación electromagnética sea del mismo orden de las distancias entre los focos dispersores, los átomos de cristal. El modo en que la cristalografía de rayos X, llega al conocimiento de la estructura cristalina de cualquier material puede

comprenderse mejor si se compara con el símil que representa la observación de un objeto a través de un microscopio óptico. En éste, la radiación dispersada por el objeto en estudio se recombina de nuevo, a través de un sistema de lentes, para dar lugar a la imagen aumentada del objeto dispersor. Sin embargo, no existe ninguna lente que sea capaz de focalizar los rayos X dispersados por los átomos. Para ello, las técnicas cristalográficas hacen uso de la herramienta matemática Transformación de Fourier, con la que se consigue “focalizar” los rayos X difractados.

### **1.1.34 Colección de datos cristalográficos**

En los experimentos de difracción se hace incidir un haz monocromático de rayos X (con una longitud de onda alrededor de 1 Å) sobre un monocristal de la muestra en estudio y se obtiene un patrón de intensidades correspondientes a los haces difractados por el cristal. La recogida de datos de difracción consiste en la medida de dichas intensidades y su posición espacial respecto al sistema de referencia del laboratorio. El diseño experimental implica la toma de imágenes sucesivas en un detector plano situado en la perpendicular del haz de rayos X, para distintas posiciones angulares del cristal, hasta obtener un conjunto de datos completo. Esto se consigue mediante oscilaciones simples del cristal (giros en  $\Phi$ ) o bien mediante una combinación de giros más complejos. Como se ha mencionado anteriormente los cristales en forma de óvalos del dominio ING5<sub>1-105</sub> obtenidos en la condición (0.1 M MES pH 6.5, 22% MPD, 5% isopropanol), fueron probados en el difractómetro del CICbioGUNE X8-Proteum System (*Bruker*). El resto de los datos fueron colectados usando la línea ID23-2 del *European Synchrotron Radiation Facility* (ESRF) y la línea XALOC del sincrotrón ALBA). Las estadísticas se resumen en la Tabla 5.

### **1.15 Análisis de los datos**

#### **1.1.35 Indexación, escalado e integración de datos**

Del experimento de difracción se obtiene un conjunto de imágenes digitales con información de las intensidades difractadas por el cristal. El análisis de datos sigue la siguiente secuencia procedimental: indexado, integración, escalado y cálculo del módulo de los factores de estructura. Después de obtener la primera imagen de difracción e



inspeccionar distintos parámetros cualitativos como la resolución de las reflexiones, la presencia de anillos de hielo y la extensión de la difracción, si la imagen es buena se puede obtener información (conocer el ángulo inicial requerido para obtener el 100% de las reflexiones con la mínima rotación del cristal) que permite realizar una recogida de datos óptima. Indexar implica asignar índices  $hkl$  a las distintas reflexiones obtenidas del experimento de difracción. Obteniéndose la información de la geometría de la celda unidad. Dependiendo de la calidad de los datos, es posible realizar el indexado del patrón de difracción utilizando una sola imagen (Powell, 1999). En la práctica se suele utilizar más de una imagen para realizar el indexado de datos, debido a que el cristal puede no difractar isotrópicamente por lo cual las reflexiones de diferentes orientaciones aseguran que la difracción anisotrópica no está produciendo pérdida de resolución en ciertas direcciones u orientaciones del cristal. Una vez las reflexiones han sido indexadas, estos deben ser integradas; proceso en el cual se calcula un valor para cada intensidad  $I(hkl)$  de cada una de las reflexiones colectadas y se calcula la desviación estándar asociada  $\sigma I(hkl)$  la intensidad de cada reflexión se estima mediante la integración de las intensidades asociadas a cada elemento de área de cada punto de la imagen de difracción. Para ello hay que realizar estimaciones de las imperfecciones del cristal y tener en cuenta el ruido de fondo de la imagen y la mosaicidad del cristal que hacen que el perfil de los puntos sea diferente (Leslie, 2006; Rossmann, 2006). Las intensidades integradas deben ser corregidas (para que todas sean comparables y se puedan promediar las reflexiones equivalentes, reduciéndolas a la unidad asimétrica) ya que presentan heterogeneidades que se pueden originar por factores como daño del cristal por radiación, variaciones en la sensibilidad del detector, variabilidad en la intensidad e iluminación del cristal, etc. La corrección de las intensidades integradas se denomina escalado, como parte del proceso de escalado se realiza un promedio de las intensidades equivalentes (relacionadas por simetría) y se normalizan calculando la desviación estándar (Evans, 2006; Kabsch, 2006). De este proceso se obtienen una serie de parámetros estadísticos que se utilizan para evaluar la calidad de los datos obtenidos (Evans, 2011; Weiss, 2000) entre los que se encuentran:

$R_{\text{merge}}/R_{\text{sym}}$ : factor de desacuerdo de las intensidades de las reflexiones relacionadas por simetría. Usualmente se asumen como buenos datos aquellos que exponen un  $R_{\text{merge}}$  menor del 10%-15% y de 50%-60% en la última capa de resolución.

$I/\sigma(I)$ : indica la relación de intensidad de la señal con respecto al ruido de las reflexiones promediadas. Valores mayores de 1.5-2 son aceptables (Evans, 2011).

**Multiplicidad o redundancia**: representa el número de observaciones para cada reflexión. Cuanto mayor es este parámetro mejor es la estimación de los errores experimentales y sistemáticos.

**Complejitud**: indica el porcentaje de reflexiones recogidas sobre el total posible. Es necesaria una completitud del orden de 90%-100% y con al menos el 80% en la capa de mayor resolución.

La sinergia existente entre los anteriores parámetros determina la resolución máxima de los datos de difracción, por ende cuanto mejor sea la resolución alcanzada mayor definición tendrán los mapas de densidad electrónica, aumentando la probabilidad de que el proceso de determinar la estructura sea exitoso.

### 1.16 El problema de la fase

La correlación existente entre la estructura cristalina, y los datos de difracción obtenidos experimentalmente (espacio real y recíproco), se establece mediante la transformada inversa de Fourier de los factores de estructura, que da la función de densidad electrónica  $\rho(xyz)$ :

$$\rho(x, y, z) = \frac{1}{V} \sum_{-h}^h \sum_{-k}^k \sum_{-l}^l |F_{hkl}| \cdot e^{-2\pi i(hx+ky+lz-\alpha_{hkl})}$$

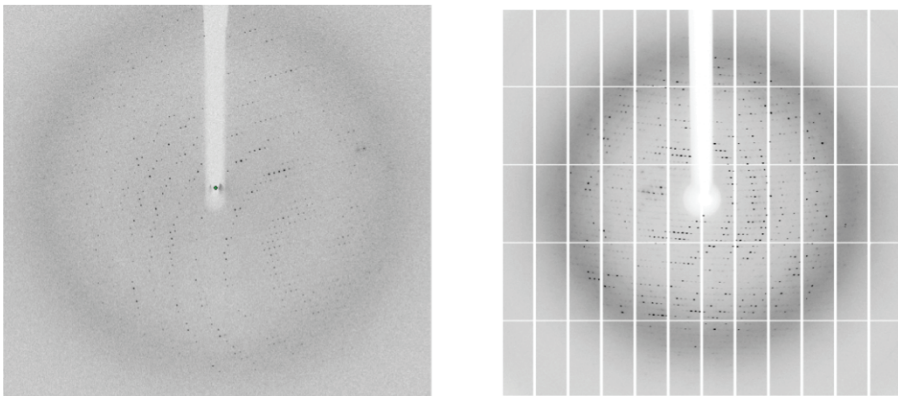
En esta transformación matemática se computan sumatorios en lugar de integrales ya que el espacio recíproco es discreto. La densidad electrónica,  $\rho(xyz)$ , en cada punto  $(xyz)$  del cristal está relacionada con  $F_{hkl}$  (factor de estructura). Los factores de estructura son magnitudes vectoriales que están caracterizados por un módulo  $F_{hkl}$ , obtenido

experimentalmente a partir de las intensidades observadas en el patrón de difracción y una fase  $\alpha_{hkl}$ , cuya información se pierde durante el experimento de difracción, dando lugar al denominado problema de la fase (Rupp, 2009; Taylor, 2003). Afortunadamente, la información sobre las fases puede recuperarse empleando aproximaciones experimentales como el reemplazamiento isomorfo simple/múltiple (SIR/MIR), el cual se basa en introducir en el cristal átomos con alto poder dispersor de los rayos X (es decir átomos con números atómicos elevados), sin que se vea afectada la estructura cristalina, con el objetivo de localizar su posición fácilmente y emplearla para hacer una estimación inicial de las fases comparando entre conjuntos de datos que contengan el átomo pesado y los que no lo contengan. (Carvin D., 2006; Green, 1954). Otro método utilizado para resolver el problema de la fase es la dispersión anómala simple/múltiple (SAD/MAD), la cual se basa en la introducción en el cristal de átomos que sean dispersores anómalos de los rayos X, como el Se en residuos de Met. Para localizar su posición se usa la diferencia en la dispersión que producen estos átomos a la misma (SAD) o diferentes (MAD) longitudes de onda. (Hendrickson, 1985; Hendrickson, 1981; Kahn & Wery, 1985). El método utilizado en esta tesis fue el de Reemplazamiento Molecular (MR), el cual se basa en calcular las fases de la proteína de interés empleando para ello las fases de una proteína homóloga como modelo (Rossmann, 2001).

### 1.17 Colección y análisis de los datos de difracción de cristales de ING5<sub>1-105</sub>

Los experimentos de difracción se llevaron inicialmente a cabo en el CIC bioGUNE donde se evaluó la difracción de los cristales y las diferentes soluciones crioprotectoras. Lamentablemente nunca se obtuvieron datos con una resolución menor de 3.5 Å. Los datos que permitieron resolver la estructura de ING5<sub>1-105</sub> a una resolución mayor fueron tomados en el Sincrotrón en la línea ID23-2 del ESRF y en la línea XALOC del ALBA.. Imágenes representativas de estos datos se muestran en la Figura 17. Las estadísticas de los datos colectados de los cristales de ING5<sub>1-105</sub> obtenidos en las diferentes condiciones se resumen en la Tabla 5. Mediante la indexación de los datos se asignaron los índices hkl a los máximos de difracción y se obtuvo información geométrica de la celda unitaria que sugiere que el retículo existente es primitivo y probablemente posea simetría hexagonal o trigonal.

El indexado, integración y escalado de los datos se ha realizado con el software HKL2000 (Otwinowski et al, 1997). Las intensidades se convierten en módulos de amplitud de factores de estructura, con el programa XDSCONV. Este programa convierte los datos de difracción obtenidos en un formato apropiado para ser utilizados con los programas de la suite de CCP4. Un análisis más profundo de las estadísticas de los datos mediante las ausencias sistemáticas de las reflexiones 00l analizadas en TRUNCATE (French, 19978) sugiere que el grupo espacial podría ser  $P6_122$  o  $P6_522$ , (Figura 18). La gráfica de  $R_{measure}$  en función de la resolución, es expuesta en la Figura 19, y el análisis de la simetría rotacional cristalográfica y no cristalográfica mediante la función de autorotación (Figura 20) el cual indica el grupo puntual es hexagonal  $P622$ . Además se observa que al utilizar una simetría menor como  $P3$  o  $P6_5$ , el programa Xtriage (Zwart, 2005) interpreta el eje binario como eje de mosaicidad; lo cual está en contradicción con dos pruebas (distribución de probabilidad acumulativa y el test Yeates-Padilla) negativas para la presencia de macla en los grupos de simetría puntual 622.



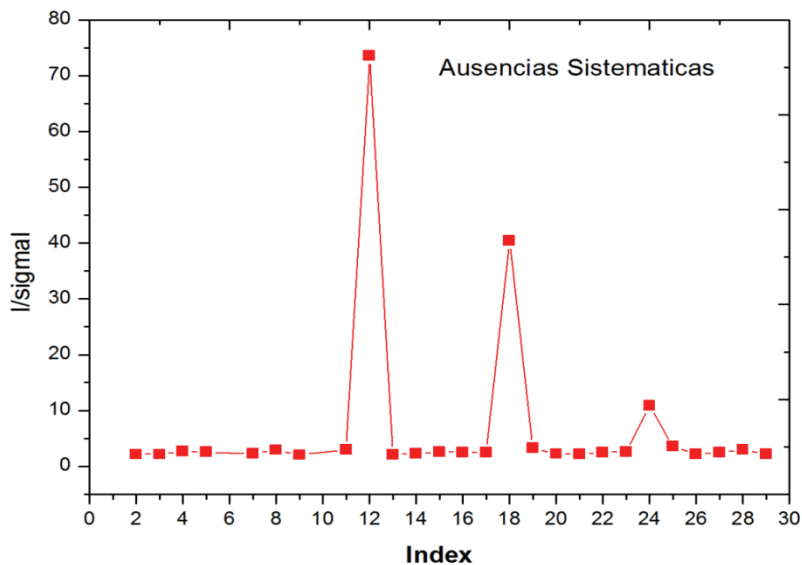
**Figura 17** Imagen de difracción del cristal de  $ING5_{1-105}$ . (Izquierda) Obtenida mediante radiación sincrotrón en la línea ID23-2 de la ESRF (Grenoble). (Derecha) Obtenida mediante radiación sincrotrón en la línea XALOC en el ALBA (Barcelona).

**Tabla 5** Estadística de la colección de datos de difracción de los cristales de ING5<sub>1-105</sub> refinamiento de los diferentes modelos.

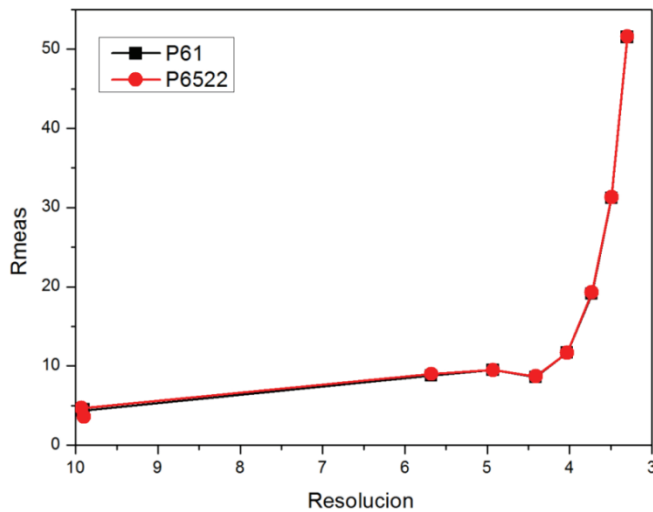
<b>Colección de Datos.</b>	<b>ESFR ID23-2 ESRF</b>	<b>CIC biogune</b>	<b>ALBA XALOC</b>
Longitud de Onda	0.87260	0.77370	0.978861
Distancia del Detector al cristal en mm	373.071	560.210	558.293
Rango de resolución	3.3-45.42	3.50- 21.13	3.2-44.70
Numero de Reflexiones Totales	10946	59837	295960
Numero de Reflexiones únicas	8599	6883	9076
Grupo Espacial	P6 <sub>3</sub> 22	P6 <sub>3</sub> 22	P6 <sub>3</sub> 22
Parámetros de unidad del celda	a = 138.7 Å, b = 138.7 Å, c = 95.9Å $\alpha=90^\circ \beta=90^\circ \gamma=120^\circ$	a = 138.7 Å, b = 138.7 Å, c = 95.9Å $\alpha=90^\circ \beta=90^\circ \gamma=120^\circ$	a = 135.6 Å, b = 135.6 Å, c = 96.68Å $\alpha=90^\circ \beta=90^\circ \gamma=120^\circ$
Multiplicidad	7.6 (7.6)	8.0 (5.1)	7.6(7.2)
Media I/ $\sigma$ (I)	39.86 (3.06)	10.76 (2.72)	47.68 (5.51)
Mosaicidad (°)	0.140	1.770	0.313
Compleitud	99.87	98.96	99.89
R <sub>merge</sub>	0.099 (0.498)	0.089 (0.870)	0.084 (0.850)
R <sub>measure</sub>	0.102 (0.516)	0.090 (0.892)	0.085 (0.862)
<b>Estadísticas del Refinamiento</b>			
R <sub>work</sub>	22.37	22.71	22.69
R <sub>free</sub>	26.87	28.76	26.52
Nº de átomos de proteína	1777	1762	1781
Nº de átomos no proteína	16	26	26
RMSD. ideal longitudes de enlace	0.006	0.010	0.014
RMSD ideal ángulos de enlace	0.58	1.16	1.76
<b>Estadística de Ramachandran</b>			
Nº de Residuos en las regiones más favorables	205	201	201
Nº de Residuos en regiones permitidas	211	5	214
Nº de residuos en regiones no permitidas	1	2	0

### 1.18 Determinación Del Grupo Espacial

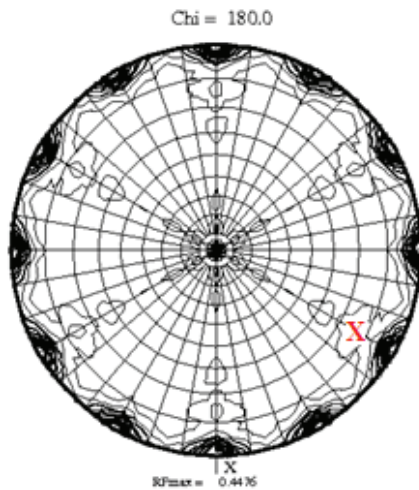
La determinación del grupo espacial se basó en la combinación de información obtenida a partir de estadísticas de los datos de difracción colectados usando la línea ID23-2 del sincrotrón *ESRF*, la información de patrones de simetría derivados de las ausencias sistemáticas, la función de autorotación y preferencias de grupo espacial obtenidas del programa Pointless de CCP4 (Winn et al, 2011). El cual sugirió que el grupo de Laue era  $P6/mmm$  y con un posible grupo espacial  $P6_122$  o  $P6_522$ .



**Figura 18** Representación de  $I/\sigma(I)$  versus Index (reflexiones axiales 00l, donde  $l=Index$ ), correspondientes al cristal de  $ING5_{1-105}$  con escalado en P1. Las ausencias sistemáticas de  $I/\sigma(I)$  para valores de Index que no son múltiplos de 6 indican que existe un eje de simetría  $6_1$  ó  $6_5$ .



**Figura 19** Representación del valor de  $R_{\text{medido}}$  ( $R_{\text{meas}}$ ) versus Resolución en los grupos espaciales  $P6_5$  y  $P6_522$ , utilizando los datos colectados usando las líneas ID23-2 del sincrotrón *ESRF* confirman que el  $R_{\text{medido}}$  no mejora utilizando una simetría más baja, indicando que el grupo Laue de los datos es 622.



**Figura 20** Función de autorotación del cristal de  $ING5_{1-105}$  correspondiente a los datos colectados en la línea ID23-2 del sincrotrón *ESRF* escalado en P1. Esta función muestra que la simetría puntual del cristal es 622. Las coordenadas de el eje binario originado a partir de la simetría no cristalográfica son: ( $\omega=76$ ,  $\phi=60$ ,  $\chi=180$ ) en la figura marcada con una letra X de color rojo.

La función de autorotación, es una herramienta útil para determinar cuántas moléculas están presentes en la unidad asimétrica y como estas están relacionadas mediante SNC rotacional, básicamente la función de auto rotación es el resultado de imponer y rotar la función de *Patterson* sobre sí misma, sin necesidad de buscar un modelo. La función de autorotación siempre genera un gran pico en el origen correspondiente a la autoimposición sin rotación.

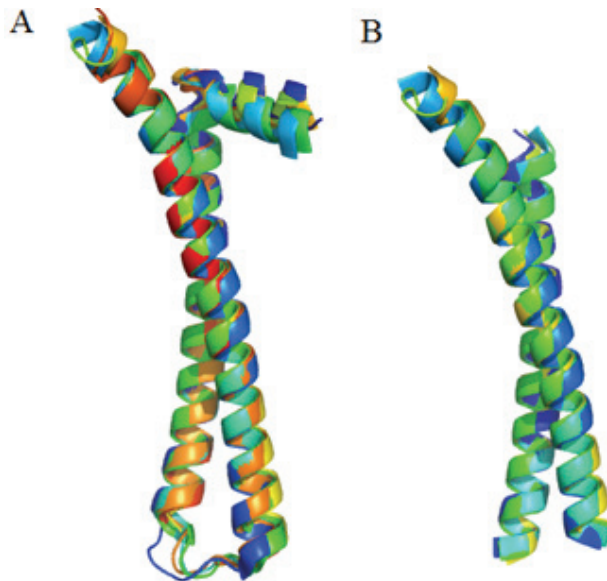
### 1.19 Reemplazamiento Molecular

Se parte de un modelo estructural conocido y un modelo preliminar construido a partir de datos experimentales de secuencia huérfana (con estructura desconocida). La identidad de secuencia entre el modelo conocido y la secuencia huérfana debe ser como mínimo de 25%. El método se basa en el cálculo de una función de rotación R (orientación) y una función de traslación T (posición en la celdilla). El mapa de Paterson del modelo conocido es movido dentro de la celda unidad obtenida experimentalmente hasta obtener la solución con el mejor ajuste entre los datos de difracción calculados a partir del modelo reemplazado y los datos observados de la estructura desconocida. Una vez encontrada la mejor rotación y traslación los factores de estructura pueden ser calculados usando las fases del modelo posicionado en la celdilla y las amplitudes medidas experimentalmente.

#### 1.1.36 Construcción del modelo usado para el reemplazamiento molecular

La estructura de ING5<sub>1-105</sub> se resolvió mediante reemplazamiento molecular utilizando como modelo la estructura de ING4<sub>2-105</sub>, el cual *a priori* parecía ser el mejor modelo posible debido a su alto grado de identidad (75%). Todas las cadenas (A-F) de la estructura de ING4<sub>1-105</sub> se procesaron con el programa CHAINSAW (Stein, 2008), con el fin de obtener un modelo que contenga información a nivel de conservación estructural entre las cadenas laterales de las secuencias de ING4 e ING5 (mediante el alineamiento de secuencias). Posteriormente con el programa COOT (Emsley & Cowtan, 2004) se eliminaron las regiones más flexibles, excluyendo las hélices  $\alpha 1$  y los lazos (*loops*), dejando únicamente las regiones correspondientes a los residuos 14-51 y 62-105 de las cadenas A-F (Figura 21). La superposición de estas cadenas fue finalmente usada como modelo de búsqueda.





**Figura 21** Modelo de búsqueda en MR. (A) Superposición de las cadenas A,B,C,D,E y F de la estructura de ING4. (B) Modelo estructural de la superposición de las cadenas A-F de ING4<sub>2-105</sub> residuos 14-51 y 62-105, el cual fue usado como modelo de búsqueda en las primeras etapas de reemplazamiento molecular con el fin de eliminar las regiones más flexibles de la estructura de ING4<sub>2-105</sub>.

### 1.1.37 Estrategia de reemplazamiento molecular

Se utilizó el programa Phaser (Zwart et al, 2008) de CCP4(Winn et al, 2011) para realizar el reemplazamiento molecular y se encontraron dos copias del ensamblaje de hélices utilizando como modelo la superposición de las cadenas A-F de ING4<sub>2-105</sub> (residuos 14-51 y 62-105) y los datos experimentales de ING5 colectados en la línea ID23-2 del sincrotrón ESRF, procesados en el grupo espacial P6<sub>5</sub>22. Cuando se intentó añadir una tercera copia del ensamblaje de hélices, los valores de los parámetros diagnósticos de calidad del modelo empeoraron, por lo que se decidió no incluir esta tercera molécula en el proceso de refinamiento (Tabla 6).

**Tabla 6** Los valores de LLG (*Logarithm- Likelihood Gain*) y los parámetros RFZ (*Rotation Function Z-score*), TFZ (*Translation Function Z-score*) indican que el modelo no mejora al incluir una tercera copia, por ende dos copias son suficientes para continuar con el proceso de refinamiento.

Numero de Moléculas	RFZ	TFZ <sub>1</sub>	TFZ <sub>2</sub>	TFZ <sub>3</sub>	LLG
1	2.3	6.9	8.5	10	60
2	2.6	8.3	9.4	8.4	136
3	3.4	8.4	4.7	6.1	137

El modelo de ING5<sub>1-105</sub> final se obtuvo a partir de las fases obtenidas con Phaser, después de algunos ciclos de construcción manual con COOT (Emsley & Cowtan, 2004) y autoconstrucción con el programa Buccaneer (Cowtan, 2006), seguidos de varios ciclos de refinamiento con el programa autoBUSTER (Blanc et al, 2004).

### 1.1.38 Construcción, refinamiento y validación del modelo de ING5<sub>1-105</sub>

A partir de las fases obtenidas durante el proceso de resolución estructural se calculó un mapa de densidad electrónica, desde el cual se construyó un modelo inicial de ING5<sub>1-105</sub> basado en los datos colectados en la línea ID23-2 del sincrotrón *ESRF*. Este modelo se mejoró mediante un proceso cíclico entre construcción manual y refinamientos, en los cuales se optimizó la posición de los átomos construidos y la estereoquímica de la molécula de acuerdo al mapa de densidad electrónica. El refinamiento se enfocó en ajustar los diferentes parámetros para optimizar el acuerdo entre los factores de estructura experimentales y los calculados a partir del modelo. Cuando los valores de  $R_{\text{work}}$  y  $R_{\text{free}}$  mejoraron durante el proceso de refinamiento, se decidió incorporar al modelo las hélices  $\alpha 1$  y los *loops* mediante autoconstrucción con el programa Buccaneer (Cowtan, 2006). La última fase del proceso de determinación estructural consistió en la validación del modelo obtenido. Para ello se comprueba que los diferentes parámetros del modelo cumplan con las distintas restricciones geométricas (ángulos y distancias de enlace entre átomos) y estereoquímicas, como los ángulos diedros ( $\phi$  y  $\psi$ ) de la cadena principal (Ramachandran et al, 1963). Esto se realizó con MOLPROBITY (Chen et al, 2010) usando el servidor <http://molprobity.biochem.duke.edu/>. La otra parte de la validación del modelo se basó en

los valores calculados para los factores estadísticos de correlación  $R_{\text{work}}$  y  $R_{\text{free}}$  los cuales son indicadores del acuerdo entre los datos de difracción de rayos X obtenidos experimentalmente ( $F_{\text{obs}}$ ) y los calculados a partir del modelo ( $F_{\text{calc}}$ ). Los valores de  $R_{\text{work}}$  y  $R_{\text{free}}$  deben disminuir a lo largo del proceso de refinamiento. Estos valores son dependientes de la resolución; por ejemplo, con un conjunto de datos a 2.0 Å el valor  $R_{\text{work}}$  debería finalizar alrededor de 20%-22%. Cuando el  $R_{\text{work}}$  disminuye sin la mejora del modelo, implica que existe sesgo en la construcción del modelo o sobrefinamiento. Para evitar este problema se implementó un método de validación cruzada (Brunger, 1992). En este método se calcula un segundo parámetro, el  $R_{\text{free}}$ . Este se calcula de forma similar al  $R_{\text{work}}$ , utilizando alrededor del 5 % de las reflexiones excluidas del proceso de refinamiento, lo cual permite detectar sesgos durante la construcción del modelo. El  $R_{\text{free}}$  actúa por tanto como control contra el sobrefinamiento, pues sólo los cambios que disminuyan el  $R_{\text{work}}$  y el  $R_{\text{free}}$  serán correctos. El refinamiento se da por finalizado cuando el modelo se ajusta al mapa de densidad electrónica y los factores de desacuerdo convergen. En ING5<sub>1-105</sub> el proceso finalizó al obtener un modelo con un residuo en regiones no permitidas en el mapa de Ramachandran (Ile14 de la cadena A) y unos valores aceptables de  $R_{\text{work}}$  y  $R_{\text{free}}$ . Los modelos de ING5-ALBA (que no presenta residuos en regiones no permitidas en el mapa de Ramachandran) y ING5-bioGUNE se obtuvieron realizando refinamientos en AutoBUSTER (Blanc et al, 2004) utilizando el modelo de ING5-ESRF y los respectivos datos colectados en el ALBA y CIC bioGUNE, respectivamente. Como el modelo final ING5-ALBA posee las mejores estadísticas  $R_{\text{work}}$ ,  $R_{\text{free}}$  y no posee residuos en regiones no permitidas en el mapa de Ramachandran, los análisis finales de la estructura se realizaron con este modelo.

### **1.20 Análisis de la estructura cristalográfica de ING5<sub>1-105</sub>**

Los análisis del modelo estructural de ING5<sub>1-105</sub> fueron realizados con los programas y servidores PISA (Krissinel & Henrick, 2007), EPPIC (Duarte et al, 2012), DALI (Duarte et al, 2012), H++ (Anandakrishnan et al, 2004), CoilCheck (Alva et al, 2008), Socket (Walshaw & Woolfson, 2001) y DrawCoil (Grigoryan & Keating, 2008). Las figuras fueron producidas con el programa Pymol (Schrödinger, 2010).

Al comparar los dos protómeros de ING5 y observar diferencias entre estos, en esta tesis se utilizó la simetría matemática para cuantificar la asimetría de la estructura y dar un marco de referencia al análisis de la estructura-función de ING5<sub>1-105</sub>.

El cálculo de la asimetría global (GloA\_Sc) se realizó de la siguiente forma: para un determinado residuo en la cadena A, se calculó la distancia de su C $\alpha$  con cada C $\alpha$  de los residuos de la cadena B. Se realizó un cálculo recíproco con el mismo residuo de la cadena B con todos los residuos de la cadena A. Se calculó el promedio de la diferencia absoluta entre las dos distancias y este fue normalizado por el número de distancias calculadas. Este paso fue repetido para todos los residuos de las dos cadenas (Zwart, 2005) y el valor de GloA\_Sc fue calculado usando la fórmula:

$$GloA\_Sc = \frac{\sum_{i=1}^N \sum_{j=1}^N |AiBj - BiAj|}{NxN}$$

Donde  $i, j$  son el número de residuos,  $N$  es el número total de residuos en las cadenas, y  $A, B$  representan las dos cadenas del homodímero. El valor mínimo que se puede obtener es 0 indicando simetría perfecta (simetría matemática), no hay límite para el valor máximo que se puede obtener, ya que este puede variar dependiendo del tamaño y de la extensión de la asimetría del complejo. El *score 1* de la asimetría de la interfaz (IntA-Sc1) cuantifica el nivel de asimetría en la interfaz de dimerización en base a la fracción de residuos únicos que interactúan entre las dos cadenas en el homodímero. La fórmula usada es:

$$IntA\_Sc1 = \frac{Uir_A + Uir_B}{Tir_A + Tir_B}$$

Donde  $Uir_A$  es el número de residuos únicos de la cadena A interactuando en la interfaz,  $Uir_B$  es el número de residuos únicos de la cadena B interactuando en la interfaz,  $Tir_A$  es el número total de residuos interactuando en la interfaz de la cadena A,  $Tir_B$  es el número total de residuos interactuando en la interfaz de la cadena B.

El rango del valor IntA-Sc1 va de 0 a 1, con 0 indicando simetría perfecta y 1 indicando una interfaz completamente asimétrica.

### 1.21 Complementariedad de la Resonancia Magnética Nuclear y Cristalografía aplicadas al estudio estructural de proteínas

La RMN y la cristalografía son las dos principales técnicas utilizadas en la determinación estructural de proteínas a nivel atómico, la cristalografía con el 87% y la RMN con el 12%

de estructuras depositadas en el PDB (*Protein Data Bank*) respectivamente. La cristalografía de rayos X es una técnica en la cual todas las intensidades del patrón de difracción contienen información sobre la disposición espacial de todos los átomos de la molécula en el cristal, y su análisis global proporciona la densidad electrónica en la cual se modela la estructura molecular. En contraste la RMN provee información de carácter predominantemente local, ya que las frecuencias de resonancia corresponden a un núcleo concreto, aunque la frecuencia se ve modulada por su entorno (los núcleos cercanos). El análisis global de los datos de RMN proporciona un conjunto de restricciones conformacionales (distancias, ángulos, orientaciones relativas) que se usan para construir un modelo compatible con todas ellas, y cuantas más restricciones se tengan mayor será la precisión de este modelo. Sin embargo, la medidas de un conjunto pequeño de restricciones permiten obtener información estructural local (por ejemplo, elementos de estructura de distancias) aún sin conocerse la estructura global. Esto puede ser muy relevante en el caso de que la determinación de la estructura global sea imposible o muy difícil. Especialmente difícil es resolver la estructura por RMN de proteínas grandes (porque su movilidad lenta degrada la calidad de la señal), elongadas y con mucho contenido en hélices (por su movilidad lenta en la dirección perpendicular al eje de elongación y por solapamiento espectral) y con alta simetría (porque los núcleos simétricos son químicamente equivalentes y las frecuencias de resonancia son las mismas). Todas estas características concurren en mayor o menor grado en el caso del dominio N-terminal de ING4, y probablemente también en el caso del dominio análogo de ING5. Por esta razón se han abordado estos estudios por cristalografía. Sin embargo, como se verá más adelante, se utilizó la RMN para obtener información estructural parcial que complementase la cristalográfica. La baja sensibilidad intrínseca de la RMN obliga a trabajar con altas concentraciones de muestra, lo cual conlleva a problemas para encontrar las condiciones óptimas en la preparación de la muestra de estudio, principalmente de solubilidad. Además para la mayoría de los estudios se requiere enriquecimiento isotópico para aumentar la resolución. El tiempo de adquisición de los datos difiere mucho entre las dos técnicas. Mientras que en la RMN la adquisición de datos puede tomar desde minutos a semanas (dependiendo de la información buscada), los datos de difracción de un cristal pueden ser tomados en menos de un minuto usando

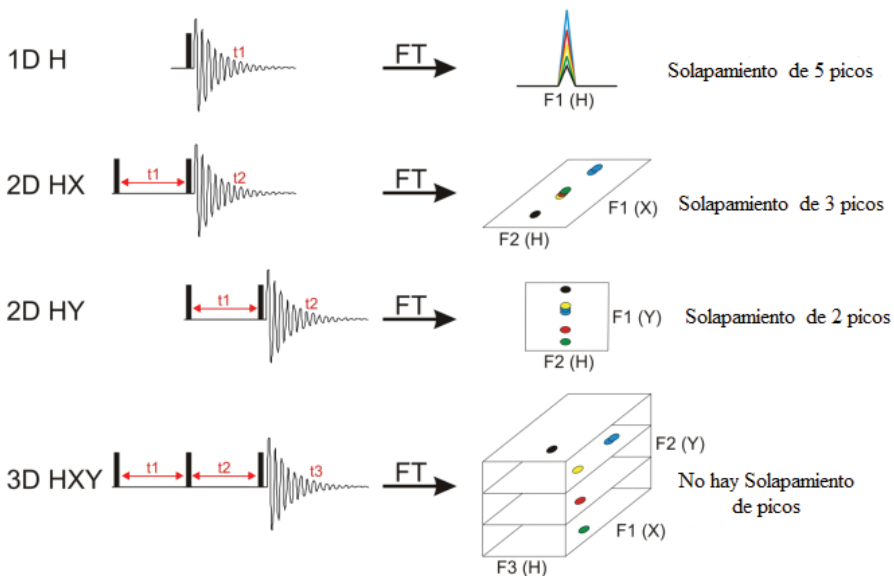
radiación sincrotrón. La RMN permite el estudio de las proteínas en disolución y unas condiciones más cercanas a las fisiológicas, por lo que se pueden llevar a cabo estudios de sus propiedades dinámicas (movimientos internos en varias escalas de tiempo, plegamiento o unión a ligandos). En los últimos años han aumentado el número de estudios donde estas técnicas se complementan siendo la cristalografía utilizada principalmente para determinar la estructura tridimensional de proteínas y complejos proteicos y la RMN enfocada a estudios de su dinámica e interacciones (Yee et al, 2005).

### **1.22 Resonancia magnética nuclear de proteínas.**

Esta técnica espectroscópica se basa en la interacción entre las ondas electromagnéticas y ciertos núcleos atómicos en presencia de un campo magnético. Algunos núcleos exhiben una propiedad mecanocuántica llamada espín y caracterizada por dos números cuánticos. Cuando el número cuántico de espín  $I \neq 0$  el núcleo tiene un momento magnético asociado y en presencia de un campo magnético puede existir en un número limitado de estados discretos de energía. Esto ocurre con núcleos que tienen un número impar de nucleones. Cuando el número cuántico  $I = \frac{1}{2}$ , el número de estados posibles es dos. Los momentos magnéticos nucleares son pequeños y su interacción con los campos magnéticos disponibles (naturales o artificiales) son débiles, por lo que la diferencia de energía entre los dos estados es pequeña y se corresponde con cuantos de luz de muy baja frecuencia (radiofrecuencias del orden de MHz). Los núcleos más comúnmente observados en RMN de proteínas son  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , por sus favorables propiedades magnéticas ( $I = \frac{1}{2}$ ) y su ubicuidad en la cadena polipeptídica. Sin embargo mientras que el núcleo de  $^1\text{H}$ , aparece con una abundancia natural de casi un 99%, la abundancia natural de  $^{13}\text{C}$ ,  $^{15}\text{N}$  es mucho menor (de un 1 y 0.4%, respectivamente). El isótopo más abundante del carbono ( $^{12}\text{C}$ ) tiene un  $I = 0$  /y por tanto no es magnéticamente activo) y el de nitrógeno ( $^{14}\text{N}$ ) tiene un  $I = 1$  que le confiere unas propiedades magnéticas muy desfavorables para su observación. Por esta razón, en la mayoría de los estudios estructurales de proteínas se producen estas con enriquecimiento isotópico en  $^{13}\text{C}$  y  $^{15}\text{N}$ . En el experimento más sencillo de RMN la muestra es irradiada con un pulso intenso de radiofrecuencia de corta duración (de unos pocos  $\mu\text{s}$ ) y potencia constante (de unos pocos vatios). Este pulso es policromático y

cuanto más corto sea es capaz de excitar un mayor rango de radiofrecuencias. Esta clase de pulsos aplicados a una región estrecha de la banda de radiofrecuencias (MHz) afecta a aquellos espines que resuenen en una banda de unos pocos kHz, excitando simultáneamente a los espines de un mismo tipo de núcleo atómico dentro de una molécula, por ejemplo al  $^1\text{H}$ . Los pulsos se transmiten a una muestra de disolución de proteína de 0.5 mL situada en el centro del campo magnético mediante una bobina que la rodea. Después del pulso, y de la interacción de la radiofrecuencia con los núcleos de la muestra, se registra una corriente eléctrica en la bobina inducida por la evolución temporal del momento magnético macroscópico (suma de todos los momentos de cada núcleo). Esta es la señal de RMN, cuya amplitud decae con el tiempo y eventualmente a cero cuando el sistema de espines se ha relajado y vuelve al equilibrio (proceso que tiene lugar en escalas de tiempo de  $\mu\text{s}$ ). Esta intensidad de corriente oscilante en el tiempo es la señal de RMN, y puede convertirse en intensidad de señal a una frecuencia determinada mediante la transformada de Fourier. Las variaciones en las frecuencias de absorción para un mismo tipo de núcleo en los espectros de se deben al distinto apantallamiento que sufre cada núcleo según su entorno químico local. Por ello estas diferencias se denominan desplazamientos químicos, y se miden en partes por millón de una frecuencia de referencia para hacerlos independientes del campo magnético utilizado para medir los espectros. En proteínas el desplazamiento químico de un núcleo de un tipo dado depende fundamentalmente de la arquitectura covalente del aminoácido en el que se encuentre y de la estructura tridimensional de la cadena polipeptídica. También se verá afectado en mayor o menor medida por los cambios de entorno químico local causados por variaciones en la composición del disolvente, pH, temperatura y la unión de ligandos. El experimento descrito más arriba (aplicación de un pulso corto de radiofrecuencia, registro de la señal oscilante y transformada de Fourier) es el más sencillo posible y da lugar a un espectro monodimensional (intensidad frente a una frecuencia). Sin embargo la RMN de proteínas requiere experimentos multidimensionales en los que se obtiene información de intensidad frente a dos o más frecuencias, correspondiendo estas a uno o varios tipos de núcleos. Estos experimentos tienen en común la concatenación de varios pulsos separados por tiempos de duración variable, el registro de series de señales oscilantes (correspondientes a valores discretos de duración de los tiempos entre pulsos). La

aplicación de la transformada de Fourier en la dimensión directa y en las dimensiones indirectas (construidas tomando un punto de cada una de las señales oscilantes de las series) resulta en el espectro multidimensional. De forma muy esquemática esto es lo que se representa en la Figura 22. La RMN multidimensional aumenta la resolución de los espectros facilitando enormemente el análisis de los espectros de polipéptidos, y más aún cuando se combinan varios tipos de núcleos. Habitualmente se miden espectros de proteínas que combinan las frecuencias de  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , por ello se denominan experimentos tridimensionales multinucleares o de triple resonancia.



**Figura 22** Esquema de la espectroscopia de RMN multidimensional y multinuclear de proteínas (donde H representa el núcleo  $^1\text{H}$ , mientras que X e Y pueden ser  $^{13}\text{C}$  o  $^{15}\text{N}$ ). La transformada de Fourier de la señal oscilante en la dimensión temporal directa (representada por la curva) y la de las dimensiones indirectas (no representadas en la figura pero originadas por la repetición de los experimentos con distinta duración de los intervalos de tiempo entre pulsos) da lugar a los espectros multidimensionales que permiten resolver el alto solapamiento de frecuencias del espectro monodimensional.

### 1.1.39 El experimento $^1\text{H}$ - $^{15}\text{N}$ -HSQC y la espectroscopía TROSY

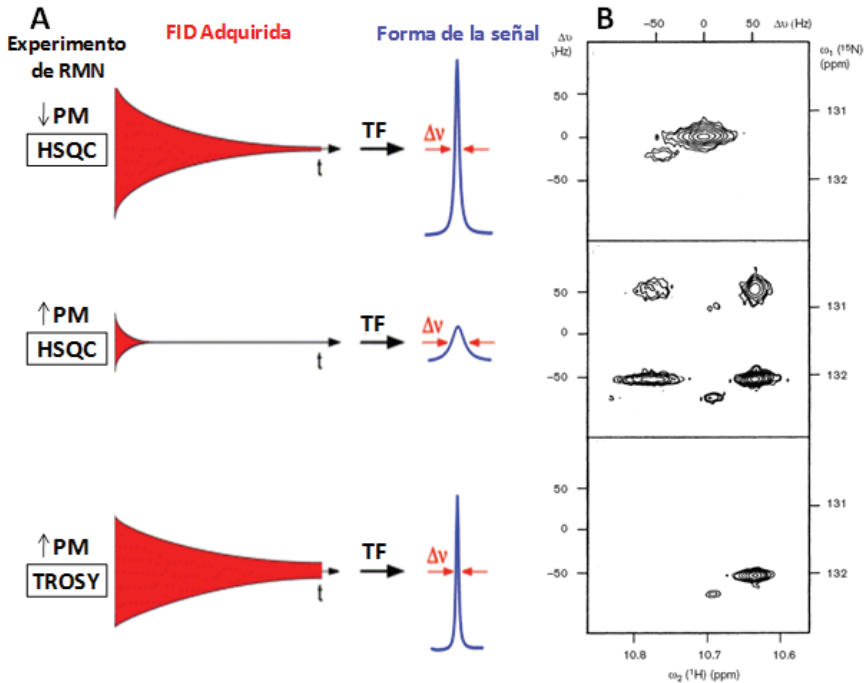
Es frecuente comenzar el estudio por RMN de una proteína preparándola mediante enriquecimiento en  $^{15}\text{N}$  y registrando el experimento HSQC (*Heteronuclear Simple*



*Quantum Correlation*) que correlaciona las frecuencias de núcleos de  $^1\text{H}$  y de  $^{15}\text{N}$ . Las señales observadas en este experimento bidimensional corresponden, en su mayoría, a los grupos N-H de cada aminoácido. Por tanto contando el número de señales y observado su dispersión en frecuencia (que está relacionada con la estructura tridimensional) se puede evaluar la calidad y grado de estructuración de la proteína antes de continuar con la preparación de muestras con esquemas de enriquecimiento más elaborados y el registro de experimentos más sofisticados.

La posibilidad de excitar los diferentes núcleos de la muestra con varios pulsos de radiofrecuencia, cada uno de ellos aplicado con una potencia, duración, frecuencia, forma y fase particulares, e introducirlos en momentos específicos de tiempo durante el experimento de RMN, antes de que el sistema haya regresado al equilibrio por relajación, permite diseñar toda una gama de secuencias de pulsos de las que se puede extraer información molecular muy variada. Aunque en teoría no existe límite en el número de dimensiones de un experimento, experimentalmente hay limitaciones impuestas por la consiguiente pérdida de señal por relajación que conlleva la detección de las distintas dimensiones. De hecho en sistemas donde la relajación es muy rápida (alto peso molecular) incluso los experimentos bidimensionales pueden implicar secuencias de pulsos de duración demasiado larga para empezar el registro de la señal antes de que se haya desvanecido, lo cual disminuye la sensibilidad además de la resolución del experimento. Una estrategia para incrementar la vida media de las señales y facilitar su observación es el uso de la espectroscopia TROSY (*transverse relaxation optimized spectroscopy*). En la espectroscopia TROSY se aprovecha la cancelación por interferencia destructiva de los mecanismos de relajación por acoplamiento dipolar y de anisotropía de desplazamiento químico. Esta cancelación es tanto mayor cuanto menor es la movilidad global de la proteína (mayor tamaño) y más intenso es el campo magnético (al menos en el rango disponible en los espectrómetros comerciales). Mediante la espectroscopia TROSY se puede seleccionar de entre las cuatro componentes que dan lugar a la señal de un NH en un experimento HSQC aquella con la relajación más lenta, y por tanto más estrecha, y detectarla con mayor sensibilidad, y resolución, como se ilustra en la Figura 23. Esta espectroscopia funciona aún mejor cuando la proteína tiene todos sus protones sustituidos por deuterones salvo en el caso de los grupo amino (y en la

práctica también los de otros grupos con átomos de hidrógeno intercambiables con el disolvente).



**Figura 23** Experimento HSQC-TROSY (A) Esquema de la aceleración de la caída libre de la inducción correspondiente a una señal de RMN de una proteína al aumentar el peso molecular, lo cual provoca el ensanchamiento de la señal en el espectro correspondiente. Esta señal se afina con el uso de la espectroscopia TROSY. (B) Detalle del protón indólico del W48 en el espectro de correlación  $^1\text{H}$ - $^{15}\text{N}$  adquirido sobre una muestra 2 mM del homeodominio de *fitz* marcado con  $^{15}\text{N}$  en complejo con un dúplex de ADN de 14 pares de bases (abundancia isotópica natural) en 95%  $\text{H}_2\text{O}$  y 5%  $^2\text{H}_2\text{O}$  a 4 °C, pH 6,0, medido a 17.6 T. Arriba: Espectro HSQC  $^1\text{H}$ - $^{15}\text{N}$  convencional desacoplado. Centro: El mismo espectro que el anterior pero sin desacoplar durante los periodos de evolución y detección. Abajo: Espectro  $^1\text{H}$ - $^{15}\text{N}$  TROSY con selección de la componente del multiplete que relaja más lentamente del multiplete de cuatro componentes, figura adaptada de (Wider & Wuthrich, 1999).

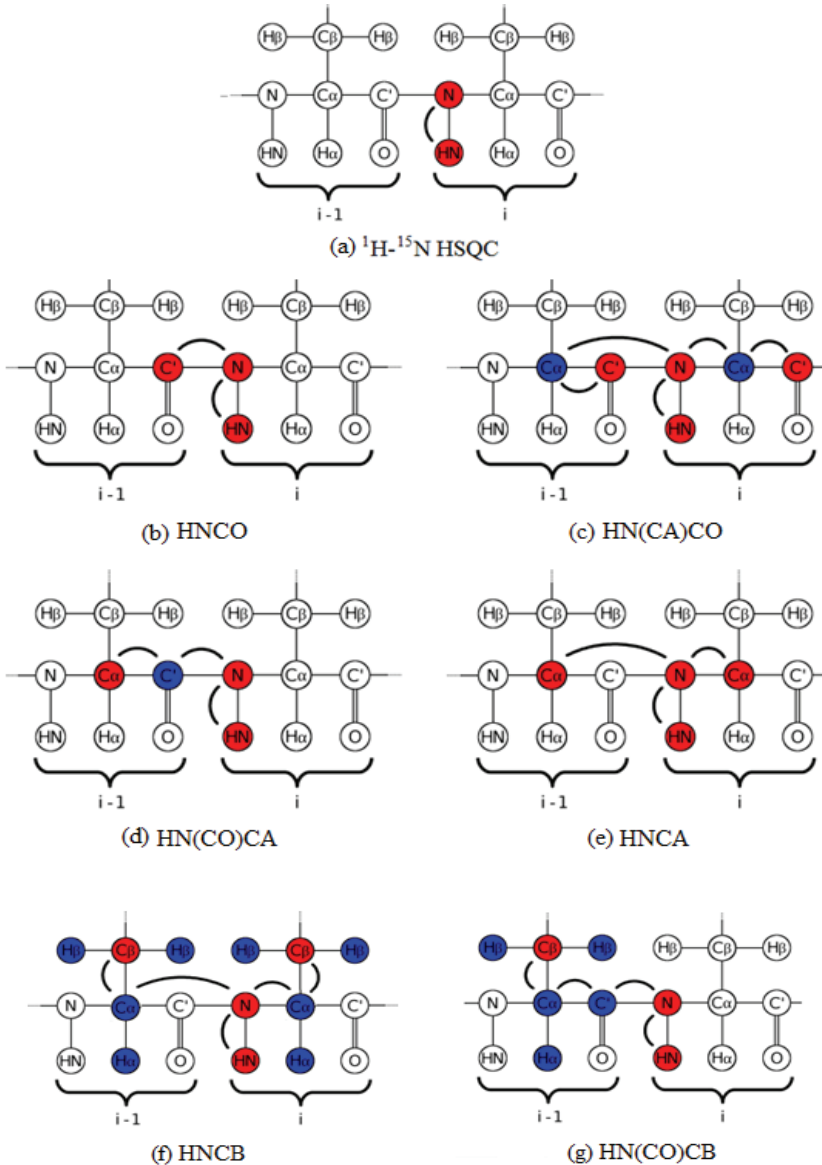
Como se ha comentado más arriba en el caso del dominio N-terminal de ING5 concurren una serie de circunstancias (tamaño, forma y estructura secundaria) que favorecen la relajación rápida de las señales de RMN y como consecuencia la calidad de los espectros con muestras protonadas no era suficientemente alta para un análisis detallado. Por esta

razón fue necesario preparar muestras uniformemente enriquecidas en  $^2\text{H}$ ,  $^{13}\text{C}$  y  $^{15}\text{N}$  y registrar los experimentos utilizando la espectroscopia TROSY en un espectrómetro de alto campo (de 18.8 T, correspondiente a una frecuencia de  $^1\text{H}$  de 800.13 MHz).

### **1.1.40 Experimentos de triple resonancia para la asignación del espectro de RMN de ING5<sub>1-105</sub>**

La asignación secuencial es el método por el cual se combina el conocimiento de la secuencia de aminoácidos de la proteína con la conectividad de los núcleos en la cadena polipeptídica (a través de enlaces covalentes y de distancias cortas, inferiores a 5Å) para asignar las frecuencias de resonancia observadas en los espectros a núcleos concretos de la proteína. El método original desarrollado por Kurt Wüthrich en los años 80 {Wüthrich, 1986 #998}, el cual utilizaba tanto la información de distancias como de enlaces covalentes contenida en espectros bidimensionales homonucleares (de  $^1\text{H}$ , en proteínas con abundancia isotópica natural). En los años 90, con la disponibilidad de muestras de proteínas isotópicamente enriquecidas y con el desarrollo de nuevas secuencias de pulsos, se impuso la estrategia basada en espectros de triple resonancia que hace un uso casi exclusivo de la conexión por enlaces covalentes para asignar los núcleos de la proteína {Ikura, 1990 #999}. Esta estrategia es además susceptible de automatización con una tasa de éxito que depende de la calidad de los espectros. La asignación de las resonancias del esqueleto y el primer carbono de las cadenas laterales ( $^{13}\text{C}^\beta$ ) de ING5<sub>1-105</sub> se realizó usando un conjunto de seis experimentos de triple resonancia. Estos experimentos y la información que se obtiene en cada uno de ellos se ilustran en la Figura 24. Los nombres de los experimentos de triple resonancia son muy descriptivos, ya que están formados por los nombres de los núcleos observados y entre paréntesis los núcleos utilizados en la secuencia de pulsos pero no observados, lo cual da una idea del tipo de información que proporcionan. Estos experimentos se analizan por parejas ya que en cada par se observa el mismo tipo de núcleos pero con conexión secuencial complementaria, como se puede observar comparando la información del experimento HNCO (secuencial) y el experimento HN(CA)CO (secuencial e intraresiduo). El análisis conjunto de estos experimentos proporciona la información con la redundancia necesaria para suplir las carencias en un experimento determinado (debido a imperfecciones en las secuencias de pulsos, artefactos,

y características particulares de la secuencia de aminoácidos y/o estructura local) y permiten resolver el “puzzle” de la asignación.



**Figura 24** Diagramas de transferencia de la magnetización en los experimentos utilizados en la asignación de las resonancias del esqueleto y C $\beta$  del espectro de RMN de ING5 $_{1-105}$ . En azul se señalan los núcleos que participan en los procesos de transferencia pero no son observados y en rojo los núcleos cuyas frecuencias sí son observadas y cuyas frecuencias se miden en los respectivos espectros.

La muestra utilizada para adquirir los experimentos de triple resonancia se preparó enriquecida uniformemente en  $^2\text{H}$ ,  $^{13}\text{C}$  y  $^{15}\text{N}$ , aunque durante el proceso de purificación en tampones protonados se intercambian los deuterones de los grupos amino por protones. Una vez terminada la etapa de purificación por tamizado molecular, se seleccionaron las fracciones más limpias y el tampón se cambió a 50 mM MOPS pH 7.0 mediante una columna de desalado PD10. La proteína fue concentrada hasta 250  $\mu\text{M}$  en un volumen de entre 300  $\mu\text{L}$  mediante ultrafiltración con concentradores Amicon con un corte de peso molecular de 3 kDa, se añadió un 10 %  $^2\text{H}_2\text{O}$  (v/v), 0.02 %  $\text{NaN}_3$  se introdujo en tubo Shigemi de 5 mm de diámetro. Los espectros se realizaron a 25 °C (después de evaluar la calidad de los espectros entre 25 °C y 35 °C) en un espectrómetro Bruker Avance III 800 MHz usando una sonda de triple resonancia con capacidad de pulsar en  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  y con gradientes en z. Además la bobina de detección que rodea la muestra, así como otros componentes electrónicos, están enfriados mediante un vapor de helio para aumentar la sensibilidad de las medidas. Algunos de los parámetros de adquisición más relevantes de los diferentes experimentos registrados se muestran en la Tabla 7. Todos los experimentos de triple resonancia se registraron en su versión TROSY. En el caso del par de experimentos HN(CB) y HN(CO)CB, estos se registraron en versión *constant time*, lo cual facilita su análisis.

### 1.23 Análisis de los espectros de RMN

Los desplazamientos químicos de  $^1\text{H}$  fueron calibrados en relación al compuesto de referencia DSS (*Sodium 2,2-dimethylsilapentane-5-sulphonate*) presente en la muestra, y los de  $^{15}\text{N}$  y  $^{13}\text{C}$  fueron calculados indirectamente a partir del calibrado de  $^1\text{H}$  (Wishart et al, 1995). Los espectros obtenidos fueron procesados con el programa TopSpin (Bruker) y analizados con Sparky (Goddard, 2008). Para facilitar la asignación se ha utilizado el programa de asignación automática MARS (Jung YS, 2004). Este programa requiere un listado de los picos (lista de desplazamientos químicos de cada una de las señales de los espectros) y la secuencia de la proteína. Utilizando una base de datos de desplazamientos químicos para cada tipo de aminoácido realiza un análisis global para explicar las

conexiones secuenciales observadas en los distintos experimentos. La asignación proporcionada por MARS tiene distintos niveles de fiabilidad (según la redundancia de información disponible) y se confirmó manualmente en algunos casos.

**Tabla 7** Parámetros de los experimentos de RMN utilizados en la asignación de ING5<sub>1-105</sub>.

Experimento	Duración (h)	Barridos	Puntos de adquisición	Puntos de procesado	Anchura spectral (ppm)	Offset (ppm)
<sup>1</sup> H- <sup>15</sup> N-TROSY	0.25	8	2048 128	2048 512	16 21	4.703 116.5
HNCO	8.5	4	2048 96 80	2048 128 128	16 7 21	4.703 177.7 116.5
HN(CA)CO	31	16	2048 96 96	2048 128 128	16 7 21	4.703 177.7 116.5
HNCA	19	8	2048 96 96	2048 128 128	16 16 21	4.703 60.7 116.5
HN(CO)CA	17	8	2048 96 80	2048 128 128	16 16 21	4.703 60.7 116.5
HN(CB)	68	8	2048 320 96	2048 256 128	16 46 21	4.703 38.7 116.5
HN(CO)CB	46	8	2048 256 80	2048 256 128	16 46 21	4.703 38.7 116.5

A partir de los valores de los desplazamientos químicos de los núcleos del esqueleto se pueden identificar de manera bastante fiable los elementos de estructura secundaria. Los desplazamientos químicos de los residuos en zonas desestructuradas presentan valores característicos y conservados, mientras que cuando el residuo se encuentra en zonas estructuradas (hélice- $\alpha$ , lamina- $\beta$ ) de la proteína los valores de sus desplazamientos

químicos se alejan de dichos valores según un patrón determinado. Esta característica se utiliza para determinar el *Chemical Shift Index* (Wishart & Sykes, 1994) (CSI) o índice de desplazamiento químico como valor de identificación de estructura secundaria, que asigna a un residuo estructura en (hélice- $\alpha$ , lamina- $\beta$  o irregular (*coil*)).

Posteriormente se optimizó este método de análisis teniendo en cuenta los efectos de los residuos vecinos (desplazamientos químicos secundarios) con los cuales se puede correlacionar los desplazamientos químicos con medidas estándares de dinámica de proteínas como el parámetro de orden y estimar cuantitativamente la flexibilidad por residuo del esqueleto de la proteína en lo que se conoce como RCI (*Random Coil Index*) (Berjanskii & Wishart, 2005; Berjanskii & Wishart, 2008). El RCI es un índice sin unidades que se calcula mediante la expresión:

$$RCI = 14,80|\Delta\delta C\alpha| + 4,80|\Delta\delta CO| + |\Delta\delta C\beta| + 3,93|\Delta\delta N| + 5,69|\Delta\delta H\alpha|$$

Donde  $|\Delta\delta C\alpha|$ ,  $|\Delta\delta CO|$ ,  $|\Delta\delta C\beta|$ ,  $|\Delta\delta N|$  y  $|\Delta\delta H\alpha|$  son los valores absolutos de los desplazamientos químicos (en ppm) de las diferencias entre el valor medido para el núcleo correspondiente en la proteína y el valor tabulado para cadenas desestructuradas.

Una alternativa al CSI para identificar elementos de estructura secundaria está implementada en el programa TALOS+ (Shen et al, 2009), el cual correlaciona los desplazamientos químicos y los ángulos dihedros  $\varphi$  y  $\psi$  utilizando una base de datos de proteínas previamente analizadas por RMN y con estructura conocida (Cornilescu et al, 1999). Estos ángulos determinan conjuntamente la estructura secundaria de la cadena polipeptídica.

### 1.24 Preparación de muestras paramagnéticas de ING5<sub>1-105</sub>

El uso de sondas paramagnéticas ha surgido como una técnica de gran importancia para el estudio de la estructura de proteínas, permitiendo, por ejemplo, estudiar interacciones a larga distancia en proteínas intrínsecamente desordenadas (Dyson & Wright, 2005). Una de las sondas de marcaje paramagnético más usada es el MTLS (*1-oxyl-2,2,5,5-tetramethyl-D3-pyrroline-3-methyl methanethiosulphonate*), en el que el átomo de oxígeno unido al pirrol tiene un electrón desapareado, y que se puede unir covalentemente a un grupo tiol de

un residuo de cisteína liberando ácido sulfinico. La reacción introduce un grupo comparable en volumen a las cadenas lateral de una fenilalanina, lo cual, en residuos de cisteínas expuestos, no altera significativamente la estructura y estabilidad de las proteínas (Bolin et al, 1998)El electrón desapareado del MTLs reduce la intensidad de las señales de RMN de los átomos cercanos en el espacio, dentro de un rango hasta unos 20Å (Bomar et al, 2007). Este fenómeno de reducción de intensidad en la señal se denomina PRE (*paramagnetic relaxation enhancement*), y se ha utilizado mucho para estudiar interacciones a larga distancia en proteínas intrínsecamente desordenadas (Dyson & Wright, 2005). En esta tesis el MTLs se utilizó para marcar los residuos cisteínas (C19 y C75) presentes en ING5 y obtener información estructural sobre la estructura oligomérica de ING5<sub>1-105</sub> en solución. Para esto se adaptó un protocolo descrito previamente (Bomar et al, 2007). A una alícuota de ING5<sub>1-105</sub> purificada por tamizado molecular se le cambió el tampón mediante una columna de desalado PD10 a 50 mM MOPS pH 7.0. El cambio de tampón es necesario, debido a que la reacción de marcaje con la sonda MTLs debe estar libre de DTT (para que no reaccionen sus grupos tiol con el MTLs y lo consuman. El marcaje con electrones paramagnéticos de las dos cisteínas: C19 y C75 presentes en ING5<sub>1-105</sub>, se realizó mediante una reacción de 2 h a temperatura ambiente de una muestra de ING5<sub>1-105</sub> 50 µM y MTLs 250 µM en tampón 50 mM MOPS pH 7.0. Para eliminar el exceso de MTLs de la muestra final, se pasó la mezcla de reacción por una columna de desalado PD10 previamente equilibrada en 50 mM MOPS pH 7.0. Posteriormente se confirmó la homogeneidad de la muestra y la presencia de dos moléculas de MTLs unidas covalentemente a ING5<sub>1-105</sub> mediante espectrometría de masas MALDI-TOF (cada molécula de MTLs incrementa la masa de la cadena polipeptídica en 186.3 Da).

### 1.25 Análisis de Interacciones de ING5 mediante RMN

La RMN es una herramienta muy útil para estudiar interacciones entre proteínas y sus ligandos, identificando los sitios de interacción con resolución atómica y en un rango de constantes de disociación de µM a mM. En el caso de proteínas no muy grandes y que se pueden enriquecer en <sup>15</sup>N, el método que con más sencillez proporciona mayor información se basa en la medida de cambios en los desplazamientos químicos de los grupos amida observados en espectros <sup>1</sup>H-<sup>15</sup>N-HSQC registrados en ausencia y en presencia del ligando



no marcado isotópicamente. La interacción con el ligando modifica el entorno químico de las amidas cercanas lo cual produce perturbaciones en sus desplazamientos químicos (CSP). Si se conoce la asignación de estas resonancias a residuos concretos de la proteína (es decir, si el espectro está asignado), entonces es posible no solo detectar la unión sino identificar el sitio de unión (Yang & Ullah, 2007). Las medidas de las perturbaciones de desplazamiento químico (CSP) en función de la concentración de ligando permiten además ajustar los datos a la ecuación que describe el equilibrio de formación del complejo y calcular una constante de disociación. Para estudiar la interacción de ING5<sub>1-105</sub> con JADE1L\_dominio Iib, los tampones de las dos proteínas se cambiaron a tampón 20 mM fosfato sódico pH 7.0, 50 mM NaCl, 1mM DTT utilizando una columna de desalado PD10. Se realizó un experimento <sup>1</sup>H-<sup>15</sup>N-HSQC control de ING5<sub>1-105</sub> a una concentración de 17 μM. Después de este experimento control se preparó otra muestra de ING5<sub>1-105</sub> proveniente del mismo lote a la misma concentración (17 μM) pero en presencia de la proteína GST-JADE1L\_dominio Iib a una concentración final de 85 μM, con un ratio proteína-ligando de 1:5. La interacción de ING5 con el extremo N-terminal de la histona humana H3 trimetilada en el residuo K4 (H3K4me3), se midió usando un péptido con la secuencia ARTKme3QTARKSTGGKAY, que correspondiente a los residuos 1-15 de la histona H3, más una tirosina en el extremo C-terminal para facilitar la medida de la concentración por absorbancia. Este péptido (obtenido por síntesis química, y comprado a la empresa Polypeptide, Strasbourg) se disolvió en el tampón de la medida (10 mM Tris pH 8.0, 300 mM NaCl, 1 mM DT), se ajustó el pH a 8.0 con NaOH (ya que los liofilizados de péptidos suelen elevar el pH al disolverse) y se midió la concentración por absorbancia. A 500 μL de una muestra 10 μM de ING5 en el mismo tampón (suplementado con 0.01% NaN<sub>3</sub> 3% <sup>2</sup>H<sub>2</sub>O y 4 μM DSS) y en un tubo de RMN de 5 mM de diámetro, se le añadieron volúmenes crecientes de la disolución concentrada de péptido (3.8 mM) y se registraron los correspondiente espectros <sup>1</sup>H-<sup>15</sup>N-HSQC a 25 °C en el espectrómetro de 800 MHz con criosonda. La constante de disociación (Kd) se calculó mediante las siguientes ecuaciones:  $\Delta\delta = (Kd + [P]+[L] - \sqrt{((Kd + [P] + [L])^2 - 4*[P]*[L])})/(2*[P])*\Delta\delta_{max}$ , donde [P] es la concentración de la proteína ING5 y [L] es la concentración del ligando (H3K4me3),  $\Delta\delta$  es una medida combinada de la perturbación en los desplazamientos químicos de <sup>1</sup>H y <sup>15</sup>N de

una señal particular y  $\Delta\delta_{\max}$  es la diferencia máxima entre los desplazamientos químicos de la proteína en estado libre y saturada con el ligando.  $\Delta\delta$  se calcula de la ecuación:  $\Delta\delta = \sqrt{((\Delta\delta_H)^2 + (\Delta\delta_N / 5)^2) * 0.5}$  donde  $\Delta\delta_H$  y  $\Delta\delta_N$  son los cambios en los desplazamientos químicos en las resonancias de  $^1\text{H}$  y  $^{15}\text{N}$ , respectivamente. En el caso de la interacción de H3K4me3 con el dominio PHD de ING5, se utilizó una muestra 20  $\mu\text{M}$  y un péptido con la secuencia ARTKme3QTARKSTG en 10 mM Tris pH 6.5, 150 mM NaCl, 1mM DTT y se registraron espectros  $^1\text{H}$ - $^{15}\text{N}$ -HSQC a 25 °C y 500 MHz. Estos experimentos se llevaron a cabo en el laboratorio de la Dra. Kutateladze, Universidad de Colorado, pero fueron analizados en el CICbioGUNE en el marco de esta tesis. En el caso de la interacción del dominio PHD de ING4, se usó una muestra 50  $\mu\text{M}$  en 20 mM fosfato sódico pH 6.5, 50 mM NaCl, 1mM DTT, 10% D<sub>2</sub>O, 0.03% NaN<sub>3</sub> y el mismo péptido que se usó para ING5 completa, registrándose los espectros a 25 °C en un espectrómetro de 600 MHz con criosonda (Palacios et al, 2006). Estos espectros fueron obtenidos antes del comienzo de esta tesis pero solo se había analizado la titulación de una de las señales del espectro (W237). En esta tesis se ha extendido el análisis a todas las señales que experimentaban CSPs significativos.





## **4 Resultados**

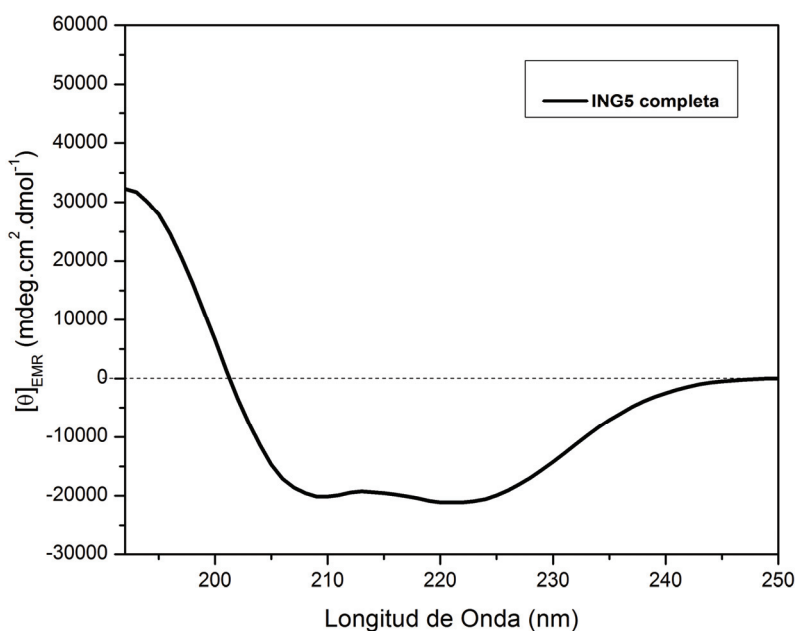


## 1.26 Obtención de proteínas ING adecuadas para estudios estructurales

Todas las proteínas producidas para los estudios estructurales de esta tesis se producían en forma insoluble en los cultivos bacterianos. Se evaluó si era posible producir las proteínas solubles mediante síntesis libre de células en presencia de exceso de chaperones GroEL y GroES, que había dado resultados positivos en el laboratorio para otras proteínas (Villate et al, 2012). Sin embargo, aunque se consiguió producir las proteínas mayoritariamente en forma soluble los niveles de expresión eran muy bajos. Como consecuencia el rendimiento final de proteína pura era también muy pequeño (24  $\mu\text{g/mL}$  de reacción en el caso de ING5<sub>1-105</sub>), y por esta razón se abandonó esta estrategia y se continuó con la producción en bacterias en forma insoluble. Esto obligó a poner a punto protocolos de solubilización en agentes desnaturalizantes y de replegamiento antes de la purificación por cromatografía, como se detalla en la sección de materiales y métodos. Aunque los niveles de expresión de partida eran muy elevados, se producían pérdidas importantes en las distintas etapas (debido a una solubilización solo parcial, precipitación parcial en el proceso de replegamiento y también en ocasiones en los distintos pasos cromatográficos). Los rendimientos típicos de proteína pura obtenidos fueron de 0.7 mg por L de cultivo de ING5, de 11 mg por L de cultivo de ING5<sub>1-105</sub> y de 2 a 4 mg por L de cultivo de los mutantes puntuales de ING5<sub>1-105</sub>. Estos rendimientos fueron menores en el caso de proteínas con enriquecimiento isotópico. El rendimiento obtenido en la purificación de las proteínas ING4 e ING4<sub>2-107</sub> fue de unos 0.4 y 2.8 mg por L de cultivo, respectivamente. Los dos miembros de la familia ING purificados (ING5 e ING4) tanto en la versión completa como los dominios N-terminales tienen un comportamiento particular en el perfil de elución en la purificación mediante tamizado molecular. En general estas proteínas presentan volúmenes de elución menores a los esperados teóricamente para proteínas globulares con el peso molecular del monómero o incluso del dímero. En el caso de ING4, que se había ya determinado experimentalmente que forma homodímeros, este comportamiento anómalo se puede atribuir a la forma elongada de la molécula. El mismo comportamiento que tienen las proteínas ING5 permite anticipar un estructura oligomérica elongada, y que la masa molar de las especies no se puede determinar con fiabilidad a partir de una columna de tamizado molecular calibrada.

### 1.27 Caracterización estructural de ING5 completa

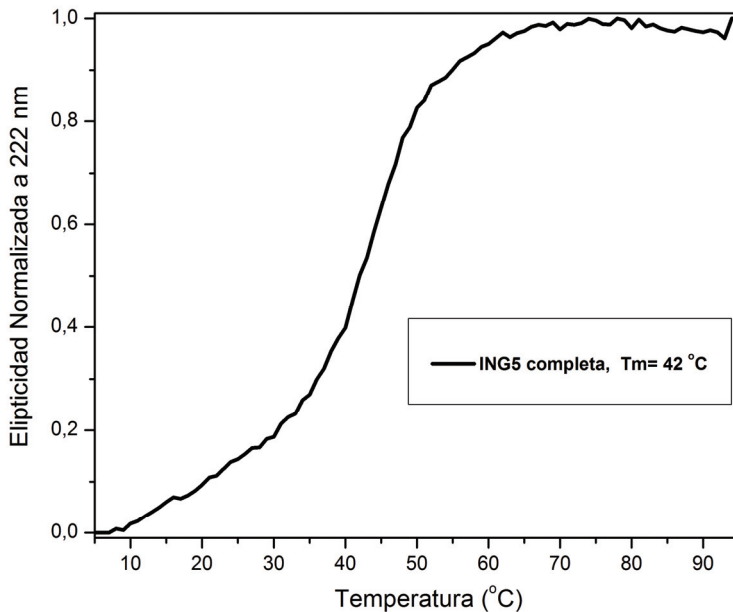
La purificación de esta proteína requirió una planificación especial ya que se debía iniciar y terminar el mismo día antes de analizarla o congelarla, ya que de otro modo se perdía mucha muestra por precipitación. La eliminación de la cola de histidinas mediante proteólisis con la proteasa TEV, no se pudo llevar a cabo ya que la mayor parte de la proteína precipitaba durante el proceso. Por ello todo el análisis de esta proteína se ha realizado con una cadena polipeptídica con 21 aminoácidos extra en su extremo N-terminal (ver secuencia en la sección 3.1.1).



**Figura 25** Estructura secundaria de ING5. Espectro de dicroísmo circular de ING5 registrado a 25 °C con una muestra a una concentración de 50  $\mu$ M en 20 mM fosfato sódico pH 6.5, 200 mM NaCl, 1 mM DTT.

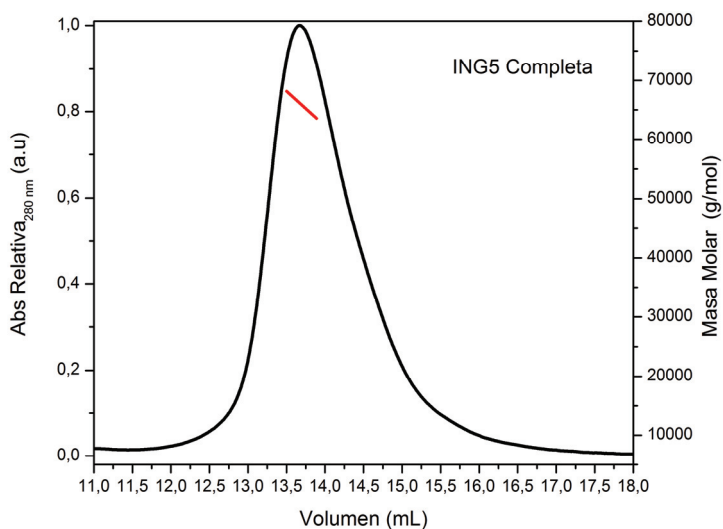
El espectro de dicroísmo circular de ING5 (Figura 25) muestra dos mínimos a 222 y 208 nm con un ratio  $\theta_{222}/\theta_{208} = 1.0$ , esto es consistente con una contribución mayoritaria de estructura helicoidal. La relación de elipticidades no permite confirmar que exista un ensamblaje en *coiled-coil* aunque posiblemente sea así y la relación no llegar a ser mayor que uno debido a otras contribuciones al espectro. De hecho la intensidad del máximo a 191

nm debería ser mayor en relación a la intensidad absoluta de los mínimos a 222 y 208 nm, lo cual sugiere que hay contribuciones de otras estructuras con elipticidad negativa en esa región, que podrían ser de una región central desordenada (*random coil*) y del dominio PHD. Este comportamiento es el mismo que el observado para la proteína homóloga ING4 (Palacios et al, 2010). La desnaturalización térmica observada por dicroísmo circular a 222 nm (Figura 26) muestra una curva con forma sigmoide con un punto de inflexión en 42 °C. Esta desnaturalización cooperativa indica que la proteína, o al menos una parte de ella tiene una estructura terciaria definida. La desnaturalización térmica es irreversible (al menos en las condiciones utilizadas) observándose la proteína parcialmente precipitada en la cubeta de cuarzo. Por tanto el punto de inflexión de la curva no es necesariamente el punto medio de desnaturalización de ING5.



**Figura 26** Estabilidad de ING5. Desnaturalización térmica de ING5 medida por dicroísmo circular a 222 nm en una muestra a una concentración de 25  $\mu$ M en 20 mM fosfato sódico pH 6.5, 200 mM NaCl, 1 mM DTT.



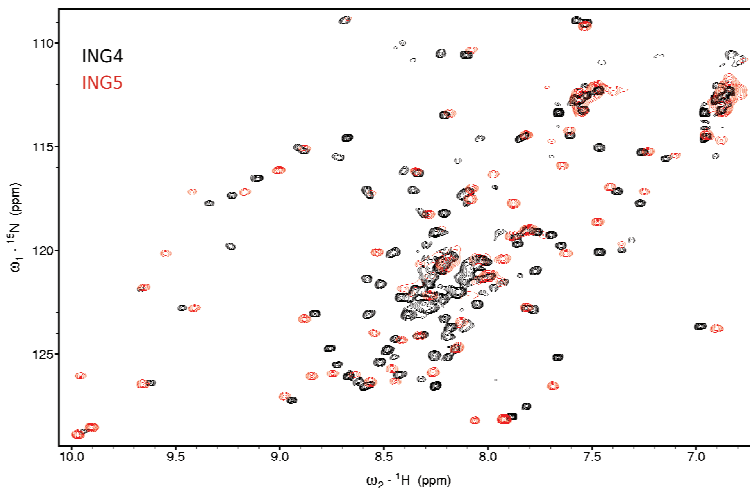


**Figura 27** Oligomerización de ING5. Experimento de SEC-MALLS registrado a temperatura ambiente en 20 mM Tris pH 8.0, 300 mM NaCl 1mM TECP 0.01 % NaN<sub>3</sub>. En negro se representa la elución de la proteína de la columna de tamizado molecular (absorbancia a 280 nm, escala en la izquierda) y en rojo la masa molar calculada a partir de los datos de dispersión de luz (escala de la derecha). Se tomó el parámetro *Mn* (promedio de la masa molar en el pico) calculado por el programa ASTRA como la masa molar experimental de la proteína.

Las medidas de SEC-MALLS (Figura 27) indican que ING5 posee una masa molar de de 62.0 kDa, muy similar a la masa de dos moléculas de ING5 (60.2 kDa) con lo cual ING5 se comporta como un dímero en solución.

El espectro de RMN <sup>1</sup>H-<sup>15</sup>N-HSQC de ING5 (Figura 28) muestra un conjunto de señales dispersas en la dimensión de <sup>1</sup>H (en un rango de 10 a 7 ppm) y algunas señales intensas entre 7.8 y 8.4 ppm (el grupo de señales a un desplazamiento químico de <sup>15</sup>N en torno a 113 ppm no son señales de grupos NH del esqueleto sino de cadenas laterales de residuos de Asn y Gln). Este patrón es similar al observado previamente para la proteína homóloga ING4, que muestra un conjunto de señales dispersas correspondientes al dominio PHD y otro conjunto de señales menos dispersos correspondientes a la región central desestructurada y al dominio N-terminal helicoidal. El menor número de señales en el espectro de ING5 en la zona central de la figura 28 se debe, muy probablemente, al mayor pH de la muestra (8.0 en comparación con pH 6.5), el cual favorece el intercambio de los

protones de los grupos amida expuestos con los protones del agua, disminuyendo su intensidad. Este es un claro indicio de que la región central es flexible y desordenada. El dominio N-terminal de ING4 es el dominio de dimerización (Palacios et al, 2010) y tiene una estructura elongada de *coiled-coil*, por ello sus señales están poco dispersas y son apenas visibles a 25 °C (debido a rápida relajación). El parecido entre los espectros de ING4 e ING5 sugiere que la estructura global de ING5 es muy similar.



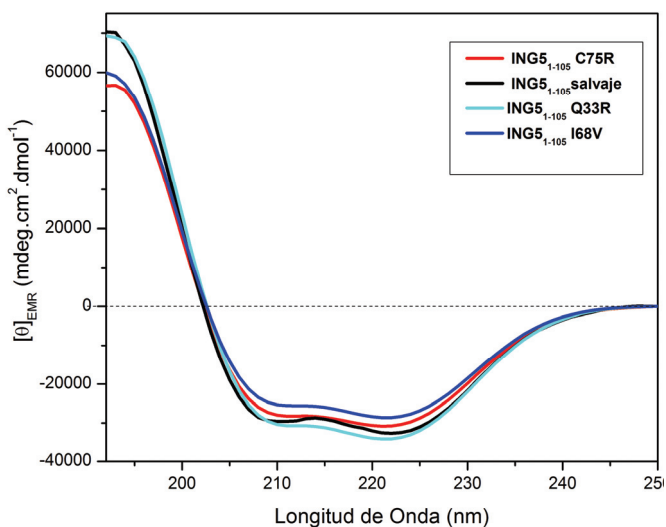
**Figura 28** Superposición de los espectros de RMN  $^1\text{H}$ - $^{15}\text{N}$ -HSQC de ING5 (color rojo) e ING4 (color negro). Los espectros fueron registrados a 25 °C y 800 MHz en las siguientes condiciones: ING4 130  $\mu\text{M}$  en 20 mM fosfato sódico pH 6.5, 50 mM NaCl, 1mM DTT, 5%  $\text{D}_2\text{O}$ , 0.01%  $\text{NaN}_3$ ; ING5 10  $\mu\text{M}$  en 20 mM TRIS pH 8.0, 300 mM NaCl, 1mM DTT, 0.01%.

### 1.28 Heterodimerización de ING4 e ING5

Se intentó detectar la posible formación de heterodimerización de las proteínas ING4 e ING5 mezclando los extractos en urea de los sedimentos obtenidos de los cultivos bacterianos y replegando la mezcla para la posterior separación de los homodímeros y el posible heterodímero mediante cromatografía, según se describe en la sección 3.4. Sin embargo el manejo de la mezcla de proteínas replegadas fue muy problemático, más aún que el de las proteínas aisladas, con grandes pérdidas de material en cada paso y no se pudo alcanzar una conclusión clara de estos experimentos.

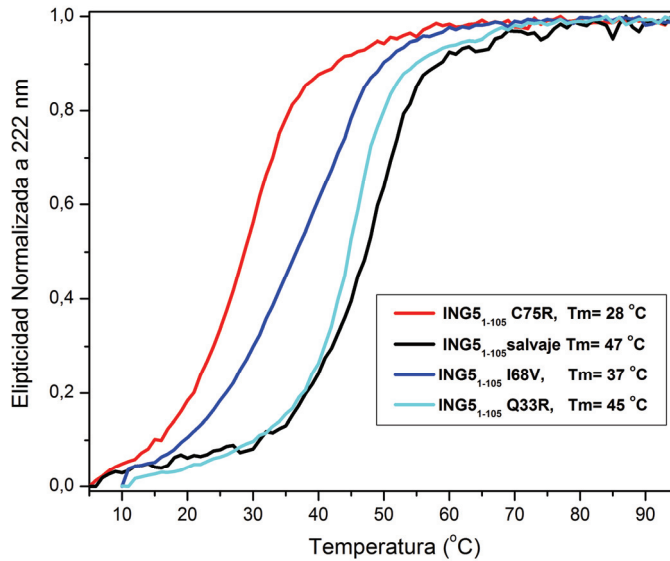
### 1.29 Análisis estructural del dominio N-terminal de ING5 (ING5<sub>1-105</sub>)

El espectro de dicroísmo circular de este dominio (Figura 29) muestra el típico patrón de mínimos y máximos de estructura en hélice alfa con una relación  $\theta_{222}/\theta_{208} = 1.2$ , que indica la formación de estructura de *coiled-coil*. En este caso el máximo a 191 nm es más intenso (con respecto a la intensidad de los mínimos a 208 y a 222 nm, en valor absoluto) que en el espectro de ING5 completa, lo que confirma que en la completa se suman contribuciones de regiones de la proteína no helicoidales, con elipticidad negativa en esta región.



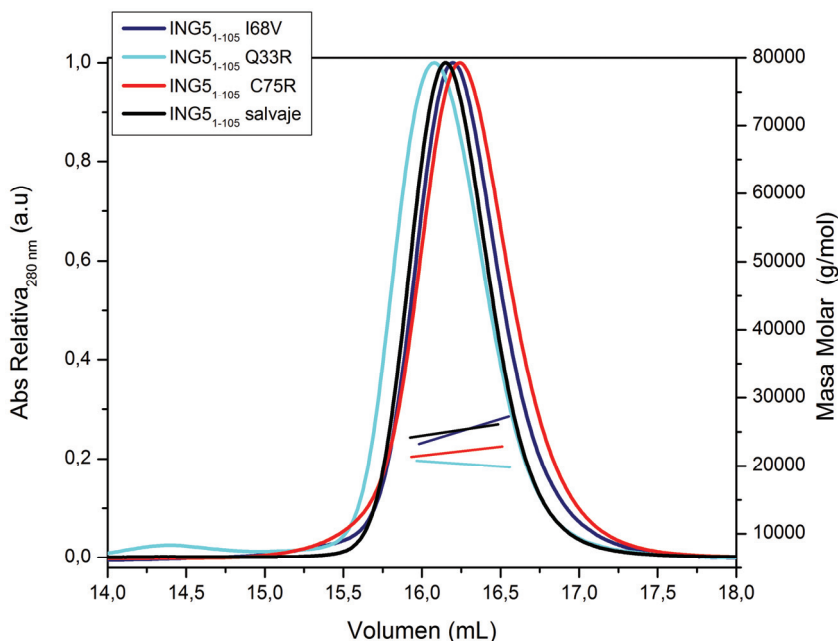
**Figura 29** Espectros de dicroísmo circular de ING5<sub>1-105</sub>. Se muestran superpuestos los espectros del dominio salvaje y de los mutantes puntuales Q33R, I68V y C75V, registrados a 25 °C con muestras a una concentración de 50  $\mu$ M en 20 mM fosfato sódico pH 6.5, 200 mM NaCl, 1 mM DTT.

La desnaturalización térmica (Figura 30) muestra una transición más cooperativa que la proteína completa, de nuevo consistente con la presencia en la proteína completa de regiones que no se desnaturalizan o lo hacen de modo poco cooperativo. El punto de inflexión de la curva, la temperatura media de la transición, es de 47 °C, un poco más alto que el medido para la proteína completa (42 °C). Esta temperatura es sensiblemente menor que la reportada para el dominio N-terminal de ING4, que es de 56 °C (Palacios et al, 2010).



**Figura 30** Desnaturalización térmica de ING5<sub>1-105</sub> y de los mutantes puntuales Q33R, I68V y C75R. Estas medidas se realizaron por dicroísmo circular a 222 nm a una concentración de 25  $\mu$ M en 20 mM fosfato sódico pH 6.5, 200 mM NaCl, 1 mM DTT.

En la purificación de este dominio la proteína eluye de la columna Superdex 75 16/60 a un volumen de 62 mL, que se corresponde con un peso molecular de una proteína globular de 70 kDa (cuando el peso molecular del monómero es de 12.5 kDa). Las medidas de SEC-MALLS (Figura 31) indican que el dominio ING5 posee una masa molar de 24.5 kDa, muy similar a la masa esperada para un dímero (25.1 kDa), confirmando que se trata de un homodímero. El peso molecular aparente de 70 kDa obtenido a partir del volumen de elución sugiere que el dímero tiene una estructura elongada, consistente con una estructura de hélices *coiled-coil*.



**Figura 31** Oligomerización de ING5<sub>1-105</sub> y de sus mutantes puntuales analizados mediante SEC-MALLS. Los experimentos se realizaron a temperatura ambiente en 20 mM Tris pH 8.0, 300 mM NaCl 1mM TECP 0.01 % NaN<sub>3</sub>. Para cada proteína se representa un par de curvas del mismo color. Una de ellas representa la elución de la columna de tamizado molecular (absorbancia a 280 nm, escala de la izquierda) y la otra la masa molar calculada a partir de los datos de dispersión de luz (escala de la derecha). Se tomó el parámetro *Mn* (promedio de la masa molar en el pico) calculado por el programa ASTRA como la masa molar experimental de la proteína.

El espectro de RMN <sup>1</sup>H-<sup>15</sup>N-HSQC de ING5<sub>1-105</sub> a 25 °C es similar al de la región central mostrada en el espectro de ING5 completa (Figura 28), con menos señales de las esperadas para una proteína de 105 aminoácidos. Esto es similar a lo que se había descrito antes para el dominio N-terminal de ING4, y es una consecuencia de la estructura dimérica, helicoidal y elongada (Palacios et al, 2010). Como se detallará más adelante (sección 4.10) se consiguieron espectros de muy buena calidad usando una muestra perdeuterada, lo que permitió la asignación de las resonancias del esqueleto y la identificación de los elementos de estructura secundaria en disolución.

### **1.30 Heterodimerización de strep-ING4<sub>2-107</sub> /his-ING5<sub>1-105</sub> análisis de mutaciones de ING5<sub>1-105</sub> detectadas en cáncer y su efecto en la dimerización.**

Con muestras de proteína pura correspondiente a los dominios N-terminales de ING4 e ING5 con colas de afinidad diferentes (Strep-ING4<sub>2-107</sub> e His-ING5<sub>1-105</sub>) se evaluó de nuevo la posible formación de heterodímeros mediante la estrategia de la purificación por cromatografía de afinidad en tándem usando las dos colas ortogonales. Se mezclaron cantidades equimolares de ambas proteínas puras, se desnaturalizaron en 8 M urea, se replegaron por dilución y se cromatografiaron sucesivamente en una columna de afinidad a Ni<sup>2+</sup> y en otra de afinidad a Strep-tactin y viceversa (como se detalla en la sección 3.4). El resultado final fue negativo, no detectándose la formación de heterodímeros. Sin embargo es interesante mencionar que en las primeras pruebas se obtuvieron resultados que aparentemente mostraban la copurificación de ambas proteínas. Esta se debía, sin embargo, a la unión de la proteína Strep-ING4<sub>2-107</sub> a la resina con iones de Ni<sup>2+</sup>. De forma inesperada esta proteína tiene una cierta afinidad por esta resina a pesar de tener solo 3 residuos y estar alejadas unas de otras en la secuencia y en la estructura tridimensional.

### **1.31 Análisis de mutaciones de ING5<sub>1-105</sub> detectadas en cáncer y su efecto en la dimerización.**

Puesto que el dominio N-terminal de ING5 está plegado de forma autónoma e independiente del resto de la proteína, se planteó estudiar el efecto de las tres mutaciones de ING5 detectadas en cáncer (las tres en el dominio N-terminal) en su estructura y estabilidad introduciéndolas en la construcción del dominio aislado. Esto facilita su manejo experimental y su análisis ya que, aunque el dominio también se produce de forma insoluble en los cultivos bacterianos, el rendimiento de este dominio en el proceso de purificación es mucho mayor que en el caso de la completa, y no tiende a precipitar tan fácilmente como esta.

Las tres mutaciones descritas dan lugar a los cambios de aminoácido Q33R, I68V, C75R (Cengiz et al, 2010). Los espectros de dicroísmo circular de los mutantes a 25 °C (Figura 29) son muy similares a los del dominio salvaje. La forma de los espectros es prácticamente la misma y difieren algo más en las elipticidades molares absolutas. Esto puede ser debido, al menos en parte, al error en la medida de la concentración de las diferentes muestras. En

## Resultados

---

todos los casos la relación  $\theta_{222}/\theta_{208} > 1.0$  (Tabla 8), que indica la formación de estructura de *coiled-coil* en todos ellos.

**Tabla 8** Ratios de la señal de dicroísmo circular  $\theta_{222}/\theta_{208}$  en los espectros de las distintas proteínas ING5, y temperaturas en el punto medio de la desnaturalización térmica medida a partir del cambio en la señal a 222 nm.

Proteína	$\theta_{222}/\theta_{208}$	$T_m$ (°C)
ING5 Completa	1.1	42
ING5 <sub>1-105</sub>	1.2	47
ING5 <sub>1-105</sub> Q33R	1.2	45
ING5 <sub>1-105</sub> I68V	1.1	37
ING5 <sub>1-105</sub> C75R	1.1	28

Las curvas de desnaturalización térmica observada mediante CD (Figura 30) revelan que los tres mutantes son menos estables que la proteína ING5<sub>1-105</sub> salvaje, siendo el orden de estabilidad  $ING5_{1-105} > Q33R > I68V > C75R$  (Tabla 8). Las desnaturalizaciones térmicas del dominio N-terminal y sus mutantes son irreversibles (al menos en las condiciones utilizadas) observándose un precipitado de proteína en la cubeta al final de proceso incluso al enfriar a 25 °C. Por tanto los puntos de inflexión de las curvas de la Tabla 8 no se corresponden necesariamente con las temperaturas medias de desnaturalización. Sin embargo son útiles para evaluar la estabilidad relativa de las distintas proteínas.

Las medidas de SEC-MALLS (Figura 31) confirman que todos los mutantes forman homodímeros en disolución (Tabla 9).

Para evaluar el impacto de las tres mutaciones puntuales C33R, I68V y C75R en la estabilidad y dimerización de ING5<sub>1-105</sub>. Se utilizaron los servidores SDM (Worth et al) y PISA (Krissinel & Henrick, 2007). El efecto de las mutaciones en la conformación oligomérica de mutantes puntuales de ING5<sub>1-105</sub> (C33R, I68V y C75R), fue evaluado mediante MALLS, como se observa en la (Figura 31). Tanto ING5<sub>1-105</sub> como sus mutantes se comportan como dímeros en solución Tabla 9.

**Tabla 9** Medidas de masa molar de las diferentes proteínas ING5 mediante SEC-MALLS.

Pesos moleculares de ING5 y mutantes puntuales calculados mediante SEC-MALLS			
Proteína	PM teórico monómero (kDa)	PM teórico dímero (kDa)	PM observado (kDa)
ING5 completa	30.1	60.2	62.0
ING5 <sub>1-105</sub>	12.6	25.1	24.5
ING5 <sub>1-105</sub> Q33R	12.6	25.2	23.9
ING5 <sub>1-105</sub> I68V	12.5	25.1	19.6
ING5 <sub>1-105</sub> C75R	12.6	25.2	21.4

### 1.32 Estructura cristalográfica de ING5<sub>1-105</sub>

De los diferentes cribados iniciales realizados, solo se obtuvieron formas cristalinas con diferentes optimizaciones de la condición 59 del cribado JCSG 1 (Tabla 4 y Figura 15), y solo difrataron los cristales provenientes de la condición 0.1M MES pH 6.5, 20% MPD, Isopropanol 4%, y de la condición MES pH 6.5, MPD 20-24%, PEG3350 2-4%, de los cuales se obtuvieron 3 conjuntos de datos de difracción (Tabla 5). Se obtuvieron cristales de ING5<sub>1-105</sub> con selenometionina pero no difrataron, y la estructura se resolvió mediante reemplazamiento molecular como se describe en la sección 3.15.

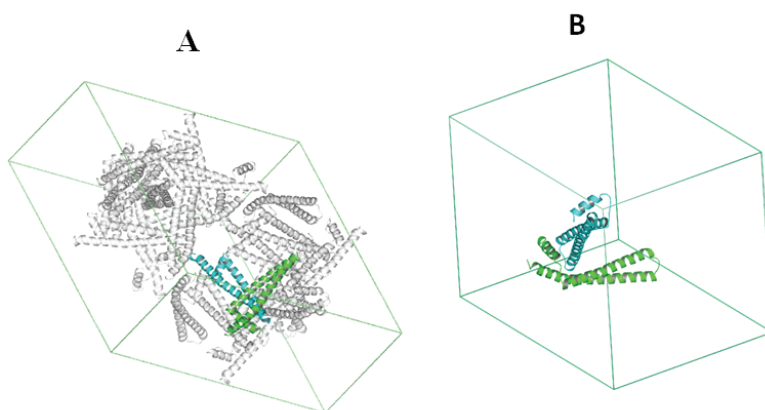
La unidad asimétrica contiene 2 cadenas polipeptídicas que forman un homodímero (Figura 32), y además contiene cinco moléculas de etilenglicol provenientes de la solución crioprotectora. Para determinar el número de moléculas en la unidad asimétrica se utilizó el coeficiente de Matthews, el cual genera una tabla con distintas combinaciones probables de números de moléculas y porcentajes de solvente en la unidad asimétrica. Inicialmente se selecciono la combinación más probable de 6 moléculas de ING5 y un contenido de solvente del 50%. Los reemplazamientos moleculares utilizando estas seis moléculas no fueron exitosos, por lo cual se decidió valorar el numero de moléculas mediante el parámetro LLG, el cual proporciona una medida del mejoramiento del modelo durante el proceso de refinamiento o cuando se comparan diferentes modelos contra los mismos datos



## Resultados

---

(Tabla 6). El parámetro indicaba que más de dos moléculas no mejoraban las estadísticas del refinamiento. La estructura de ING5<sub>1-105</sub> se resolvió con dos moléculas en la unidad simétrica y por ende según Matthews un contenido de solvente del 76%. En sistemas trigonales y hexagonales existe la posibilidad de macla. En este trabajo esta posibilidad se analizó mediante las pruebas de distribución de probabilidad acumulativa para reflexiones acéntricas y la prueba de Yeates-Padilla las cuales indicaron que no detectaban macla. El homodímero está conformado por la totalidad de la secuencia del clon ING5<sub>1-105</sub> más los residuos G-1, A0 en extremo N terminal, los cuales pertenecían al sitio de corte TEV y quedaron en la secuencia después de realizar el corte con la proteasa.

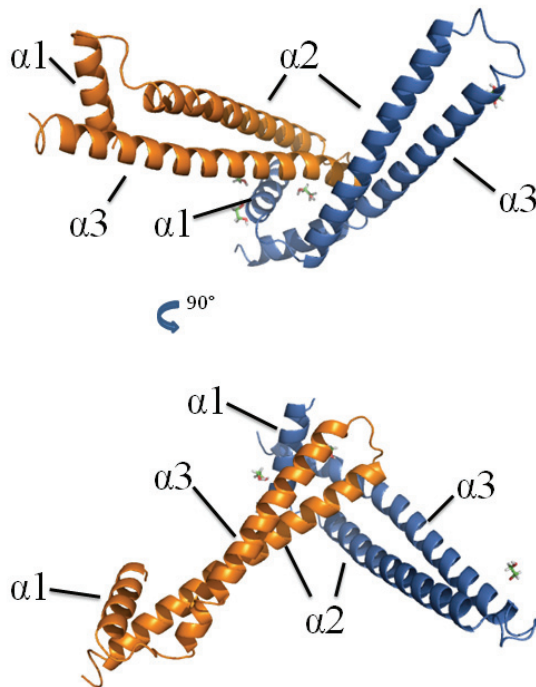


**Figura 32** Modelo estructural de ING5<sub>1-105</sub> en el cristal. (A) Celda unidad, exponiendo las dos moléculas de la unidad asimétrica en color verde y cian y las 11 x 2 moléculas relacionadas por simetría en color gris. (B) Unidad asimétrica del cristal exponiendo el homodímero de ING5<sub>1-105</sub>.

La estructura cristalográfica de ING5<sub>1-105</sub> después del proceso de refinamiento expone que el dímero está compuesto de dos cadenas A y B (Figura 33), cada una con por tres hélices- $\alpha$  prácticamente idénticas (Tabla 10). El programa DSSP identifica una hélice  $\alpha 3$  algo más corta en la cadena A debido a ligeras distorsiones en la última vuelta, que indica que aunque las hélices son muy similares, no son idénticas.

**Tabla 10** Elementos de estructura secundaria en las cadenas A y B de la estructura cristalográfica de ING5<sub>1-105</sub>. Los límites de las hélices fueron identificados utilizando el servidor DSSP ([www.cmbi.ru.nl/hsspsoap](http://www.cmbi.ru.nl/hsspsoap)).

Cadena	No. Residuos	Helice1 ( $\alpha 1$ )	Helice2 ( $\alpha 2$ )	Helice3 ( $\alpha 3$ )
A	1-105	1-11	19-52	58-98
B	1-105	1-11	19-52	58-103



**Figura 33** Dos vistas de la estructura cristalográfica del homodímero de ING5<sub>1-105</sub> resultantes de girar la molécula como se indica en la figura. Los monómeros correspondientes a las cadenas A y B se muestran en color naranja y azul, respectivamente. Se indican las tres hélices que conforman cada monómero ( $\alpha 1$ ,  $\alpha 2$  y  $\alpha 3$ ). Las moléculas de etilenglicol provenientes de la solución crioprotectora se muestran en formato stick coloreadas según el tipo de átomo.

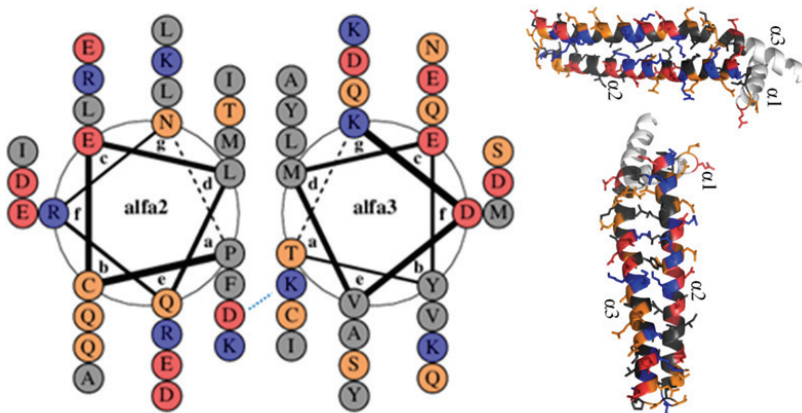
Utilizando la estructura cristalográfica de ING5<sub>1-105</sub> como archivo de entrada en el servidor Socket (<http://coiledcoils.chm.bris.ac.uk/socket/>), se determinó qué regiones de las hélices  $\alpha 2$  y  $\alpha 3$  cumplen los requerimientos estructurales típicos de una estructura *coiled-coil* y a

## Resultados

partir de ahí se identificó el patrón de héptadas  $(abcdefg)_n$  en la secuencia de ING5<sub>1-105</sub> (Tabla 11 y Figura 34).

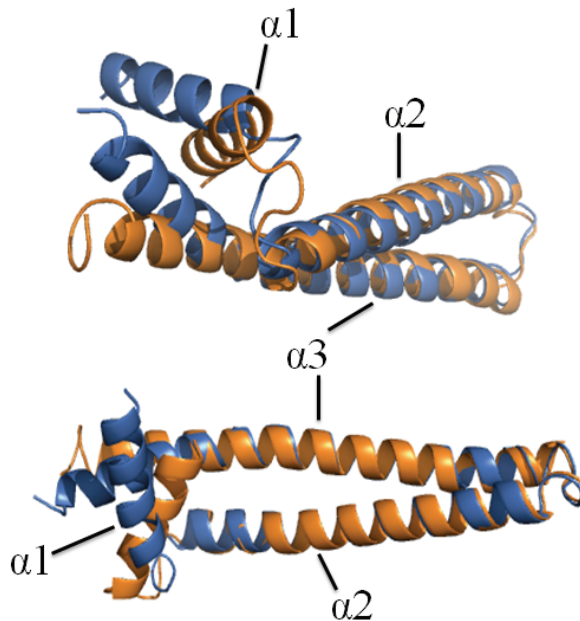
**Tabla 11** Análisis de la secuencia de héptadas presente en las secuencias de las hélices  $\alpha 2$  y  $\alpha 3$  de ING5<sub>1-105</sub>.

Héptadas en los residuos 21-45 de $\alpha 2$
defgabcdefgabcdefgabcdefg
LQRNFQLMRELDQRTEDEKKAIEDIL
Héptadas en los residuos 68-92 $\alpha 3$
MEYTMALQVKDDSYEKCKSYANQI
dcbagfedcbagfedcbagfedcba



**Figura 34** Estructura *coiled-coil* del monómero de ING5<sub>1-105</sub>. Secuencia de aminoácidos y patrón de héptadas en las regiones de las hélices  $\alpha 2$  y  $\alpha 3$  de ING5<sub>1-105</sub>. Distribución de residuos en la representación idealizada de las hélices (*helical wheels*) que están formando el *coiled-coil*. La figura se generó utilizando el programa Drawcoil 1.0 en el servidor <http://scripts.mit.edu/~gevorg/index.pl/drawcoil/>. Se indica con una línea discontinua de color azul el puente salino K82-D32 común para las dos cadenas (A y B de ING5<sub>1-105</sub>). A la derecha se muestran las interacciones hidrofóbicas intramoleculares de la cadena B de ING5<sub>1-105</sub>. Los residuos en colores de acuerdo a la predicción de estructura *coiled-coil* realizada en el programa Drawcoil. Hidrofóbicos en color gris, los residuos cargados negativamente (rojo), positivamente (azul), demás residuos en naranja. Los residuos que no forman parte de la estructura *coiled-coil* están en blanco.

La superposición de las hélices  $\alpha 2$  y  $\alpha 3$  de las dos cadenas de ING5 (Figura 35 expone que existen diferencias conformacionales notables entre los dos monómeros, principalmente en la región de la hélice  $\alpha 1$  y la región C-terminal de la hélice  $\alpha 3$ . La divergencia en la estructura del esqueleto de las dos cadenas evaluó mediante la medida de RMSD de los  $C^\alpha$  en la superposición global de la estructura, calculado con PROFIT (McLachlan, 1982). El valor es de 10.3 Å para la estructura global (residuos 1-105) y de 17.5 Å si se calcula sobre los residuos 1-20.



**Figura 35** Superposición de las hélices  $\alpha 2$  y  $\alpha 3$  de las cadenas A (naranja) y B (azul) de ING5<sub>1-105</sub>. Se muestran dos vistas de la misma superposición.

En el caso de la cadena B de ING5, las hélices  $\alpha 1$  y  $\alpha 3$  están más cerca espacialmente que las hélices  $\alpha 1$  y  $\alpha 3$  de la cadena A. Por esta razón nos referiremos más adelante a las cadenas como la conformación abierta (cadena A) y cerrada (cadena B).

#### 1.1.41 Descripción de las cadenas A y B de ING5<sub>1-105</sub>

Para analizar en detalle las diferencias existentes entre estas dos cadenas de ING5<sub>1-105</sub>, (Figuras 36 y 37) se identificaron las interacciones hidrofóbicas, enlaces de hidrogeno y

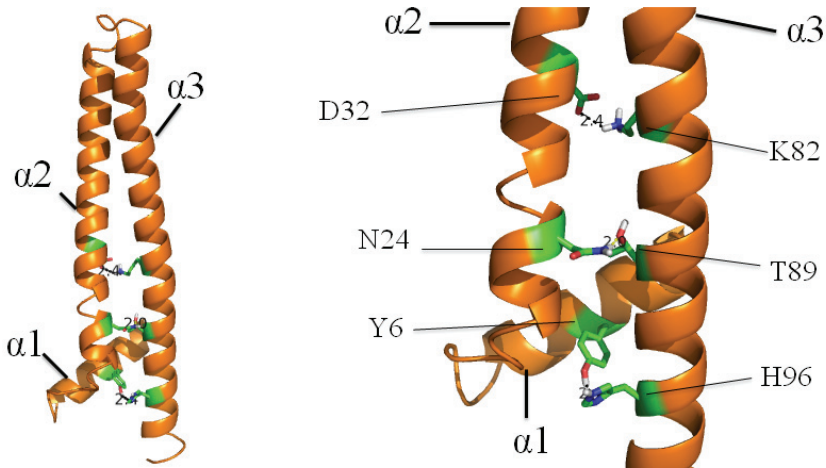
## Resultados

puentes salinos (listadas en las Tablas 13 y 14) mediante los programas PISA (Krissinel & Henrick, 2007) y Coilcheck (Alva et al, 2008) con el mismo resultado. Ambos programas se corrieron con cada cadena por separado (para identificar las interacciones intramoleculares) y con el dímero de la unidad asimétrica del cristal (interacciones intermoleculares). Los criterios utilizados por Coilcheck para determinar la existencia interacciones hidrofóbicas se basa en considerar residuos hidrofóbicos con una distancia igual o menor a 7 Å entre sus C<sup>β</sup>-C<sup>β</sup>, en los puentes salinos se consideran distancias iguales o menores a 4 Å entre los átomos con cargas opuestas de las cadenas laterales. Los enlaces de hidrogeno se identifican utilizando como criterio una distancia menor de 3.5 Å entre los átomos donador y aceptor (además de los ángulos que permitan la formación del enlace).

**Tabla 12** Interacciones intramoleculares de la cadena A. Los asteriscos indican aquellas interacciones exclusivas de la cadena A (no presentes en la cadena B) del homodímero.

<b>Interacciones hidrofóbicas intramoleculares de la cadena A</b>					
Nº	Residuos				Distancia (Å)
1 *	2	A	93	V	5.52
2 *	2	A	97	I	6.21
3 *	6	Y	93	V	5.00
4 *	6	Y	97	I	6.71
5	21	L	93	V	6.96
6	27	L	85	L	6.11
7	31	L	78	Y	5.85
8	31	L	85	L	6.58
9	42	I	68	I	6.02
10	42	I	71	A	4.80
11	42	I	72	Y	5.58
12	45	L	68	I	4.69
13	45	L	71	A	6.41
14	46	A	68	I	6.08
15 *	49	Y	68	I	6.83
<b>Puentes salinos intramoleculares de la cadena A</b>					

No.	Residuos				Distancia (Å)
1	32	D	82	K	2.40
<b>Puentes de hidrogeno intramoleculares cadena A</b>					
No.	Residuos				Distancia (Å)
1	89	T	24	N	2.00
2*	96	H	6	Y	2.40

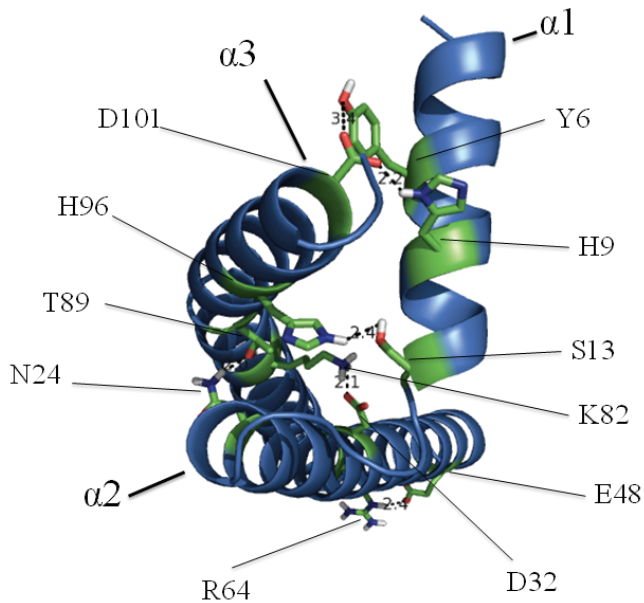


**Figura 36** Interacciones intramoleculares de la cadena A de ING5<sub>1-105</sub>. Izquierda vista general de la cadena A de la estructura cristalográfica donde se representa en verde la región del esqueleto y en *sticks* (con color según el átomo) las cadenas laterales de los residuos involucrados en el puente salino D32-K82 y los enlaces de hidrogeno N24-T89 y Y6-H96 que estabilizan la estructura *coiled-coil*. A la derecha se muestra un detalle ampliado de las interacciones indicando los residuos involucrados. Ver interacciones en Tabla 12).

## Resultados

**Tabla 13** Interacciones intramoleculares de la cadena B. Los asteriscos indican aquellas interacciones exclusivas de la cadena B (no presentes en la cadena A) del homodímero.

<b>Interacciones hidrofóbicas intramoleculares de la cadena B</b>					
N°	Residuos				Distancia (Å)
1	68	I	42	I	6.38
2	68	I	45	L	4.96
3	68	I	46	A	6.38
4	68	I	49	T	6.98
5	71	A	42	I	4.91
6	71	A	45	L	6.56
7	72	Y	42	I	5.72
8	78	Y	31	L	5.87
9	85	L	27	L	6.13
10	85	L	31	L	6.58
11*	93	V	21	L	6.98
12*	97	I	6	Y	5.49
13*	97	I	10	Y	5.37
<b>Puentes salinos intramoleculares de la cadena B</b>					
No.	Residuos				Distancia (Å)
1	82	K	32	D	2.10
2*	101	D	9	H	2.20
3*	64	R	48	E	2.99
<b>Puentes de hidrogeno intramoleculares cadena B</b>					
No.	Residuos				Distancia (Å)
1	89	T	24	N	2.20
2*	101	D	6	Y	3.40
3*	96	H	13	S	2.00

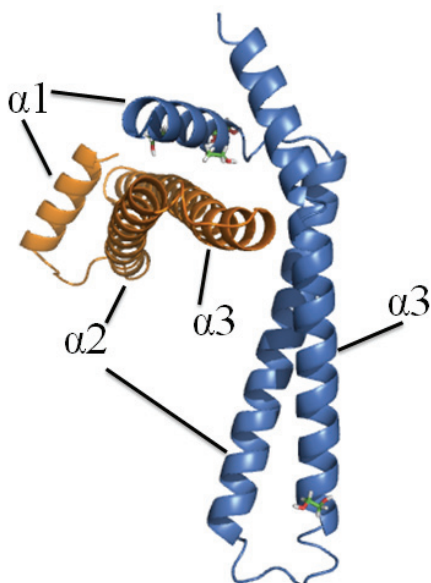


**Figura 37** Interacciones intramoleculares de la cadena B de ING5<sub>1-105</sub> donde se indican los residuos involucrados en los puentes salinos y enlaces de hidrógeno que estabilizan la estructura *coiled-coil* (la región del esqueleto está coloreada en verde y las cadenas laterales se muestran en *sticks* coloreados según el tipo de átomo). Ver interacciones en Tabla 13.

### 1.33 Descripción y análisis de la interfaz del homodímero de ING5<sub>1-105</sub>

El análisis de la interfaz del dímero de ING5<sub>1-105</sub> (Figuras 38 y 39) se realizó utilizando el servidor PISA ([http://www.ebi.ac.uk/msd-srv/prot\\_int/cgi-bin/piserver](http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver)) el cual genera un valor de relevancia de la interfaz (*formation Significance Score* CSS) que va desde 0 (no relevante) a 1 (máxima relevancia). Las interacciones de la interfaz están listadas en la Tabla 14. El dímero de ING5 posee un valor de CSS= 1. El valor de CSS, indica que la interfaz es termodinámicamente estable y juega un rol esencial en la formación del complejo, por tanto sugiere que el dímero puede existir en solución.





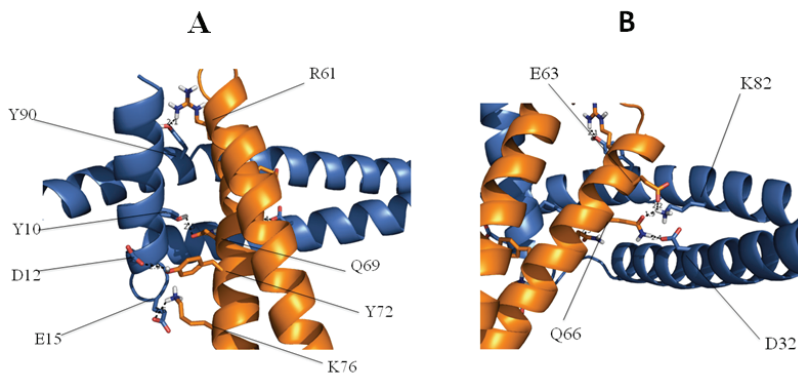
**Figura 38** Vista general de la interfaz de dimerización del dímero ING5<sub>1-105</sub>. Se observan la cadena A color naranja y la cadena B color azul. También se muestran las moléculas de etilenglicol en *sticks*.

La relevancia de la interfaz observada en el cristal se evaluó también mediante el programa EPPIC (*Evolutionary Protein-Protein Interfaz Classifier*) (Duarte et al, 2012). Este programa realiza un análisis teniendo en cuenta aspectos evolutivos (tasa mutacional entre residuos de la secuencia utilizando información de proteínas homólogas) generando un parámetro denominado *score-rim-score*, junto con análisis estructurales de los residuos involucrados en la interfaz desde el cual se genera el parámetro *core-surface-score*, que cuantifica la probabilidad de que la interfaz se produzca por azar, además de analizar la posible naturaleza artificial (contactos cristalinos debidos al empaquetamiento de la molécula en el cristal) o natural (relevancia biológica) de la interfaz estudiada. Los resultados del servidor EPPIC indican que los dos parámetros mencionados anteriormente convergen, por lo cual también apoyan la naturaleza biológica de la interfaz.

**Tabla 14** Interacciones intermoleculares del homodímero de ING5<sub>1-105</sub>. Las interacciones y fueron identificadas con el programa Coilcheck (Alva et al, 2008).

<b>Interacciones hidrofóbicas en la interfaz de dimerización</b>					
N°	Residuos Cadena B		Residuos Cadena A		Distancia (Å)
1	46	A	7	L	6.53
2	49	Y	7	L	6.58
3	50	I	0	A	5.20
4	50	I	4	A	5.63
5	62	V	86	A	4.15
6	62	V	90	Y	4.98
7	62	V	93	V	6.46
8	65	L	93	V	6.85
9	68	I	11	L	6.64
10	72	Y	11	L	5.16
11	72	Y	14	I	6.39
<b>Puentes salinos en la interfaz de homodimerización</b>					
N°	Residuos Cadena A		Residuos Cadena B		Distancia (Å)
1	63	E	82	K	2.2
2	76	K	15	E	2.6
<b>Enlaces de hidrógeno en la interfaz de dimerización</b>					
No.	Residuos Cadena A		Residuos Cadena B		Distancia (Å)
1	61	R	90	Y	2.10
2	66	Q	32	D	2.40
3	72	Y	12	D	2.90
4	66	Q	82	K	1.90
5	69	Q	10	Y	2.50

La interacción hidrófoba número 3 de la Tabla 14, es artificial ya que el residuo A0 no es natural, fue introducido en la secuencia para dejar más espacio entre el residuo G-1 (delante del cual corta la proteasa TEV) y la metionina inicial de ING5.

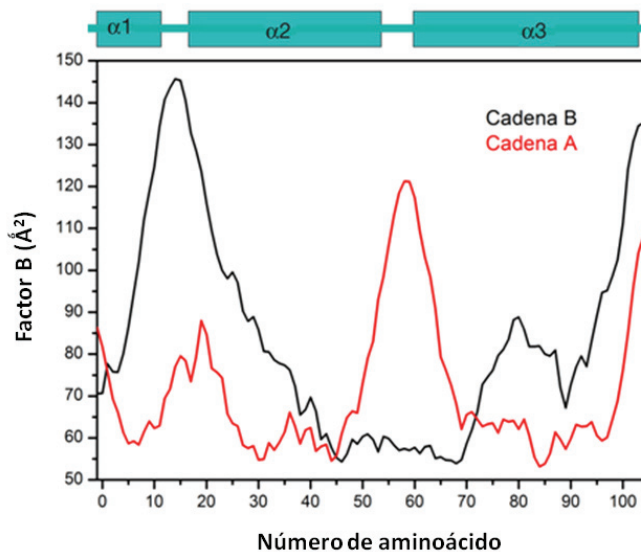


**Figura 39** Interacciones en la interfaz de dimerización de ING5<sub>1-105</sub> formada por las cadenas A (naranja) y B (azul). El detalle de las interacciones se encuentra en la Tabla 14.

### 1.34 Asimetría estructural del dímero de ING5<sub>1-105</sub>

La divergencia en la estructura del esqueleto de las dos cadenas, principalmente al nivel de la hélice  $\alpha 1$  y el extremo C-terminal de la hélice  $\alpha 3$ , así como la existencia de interacciones intramoleculares diferentes en cada uno de los dos monómeros de ING5<sub>1-105</sub> implica una asimetría en el homodímero. El grado de asimetría global y de la interfaz de dimerización se evaluó cuantitativamente según se describe en la sección 3.16. Según este análisis, el dímero de ING5<sub>1-105</sub> tiene un valor de asimetría global  $GloA\_Sc = 23.21 \text{ \AA}$  y un *score* de interfaz asimetría = 1, que indica completa asimetría en la interfaz del homodímero de ING5. En las tablas 14 y 15 de interacciones intramoleculares se puede observar que en la cadena B (conformación cerrada) se establece un puente salino entre los residuos H9 y D101 que no está en la cadena A (abierta). También en la cadena B, el residuo H96 forma un enlace de hidrógeno, con el residuo S13, mientras que en la cadena A, el enlace lo forma H96 con Y6.

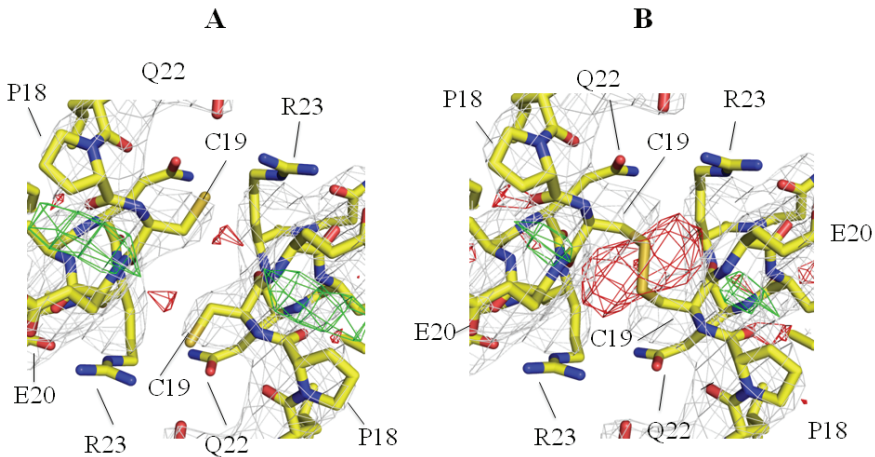
La relevancia de la flexibilidad en distintas regiones de dímeros asimétricos es controvertida debido a que no existen suficientes estudios que permitan concluir si la dimerización asimétrica requiere de niveles altos o bajos de flexibilidad (Swapna et al). La Figura 40 ilustra los valores de los factores B- de las dos cadenas de la estructura cristalográfica de ING5<sub>1-105</sub>. En la cadena B (conformación cerrada) la región correspondiente a la hélice  $\alpha 1$  expone un alto grado de flexibilidad que contrasta con el bajo nivel de flexibilidad para esta misma región (hélice  $\alpha 1$ ) de la cadena A. Lo cual es consistente con las diferencias conformacionales debido a las interacciones encontradas entre cadenas. En la cadena A existe otra región (residuos 50-70) con alta flexibilidad ausente en la cadena B. Estas diferencias podrían estar relacionadas con la estabilidad de las conformaciones abierta y cerrada encontrada en la estructura de ING5<sub>1-105</sub>.



**Figura 40** Factores B de la estructura cristalográfica de ING5<sub>1-105</sub>, donde se exponen las diferencias en flexibilidad entre las dos cadenas A y B.

Se analizaron las interacciones intermoleculares con otras cadenas de la red cristalina (contactos debidos al empaquetamiento del cristal o contactos cristalográficos) por si estos

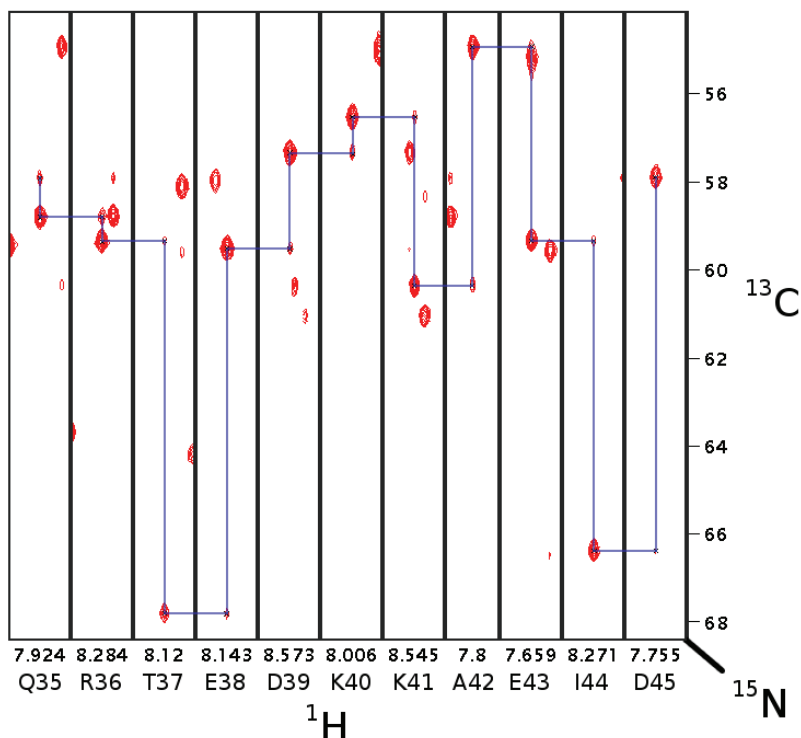
podieran explicar la presencia de un dímero asimétrico. Según este análisis los datos cristalográficos obtenidos de la difracción de cristales de ING5<sub>1-105</sub> se pueden agrupar en dos clases. La primera clase agrupa los cristales cuya condición es 0.1M MES pH 6.5, 20% MPD, Isopropanol 4%, y cuyos datos fueron colectados en el sincrotrón ESRF. La segunda clase corresponde a cristales provenientes de la condición 0.1M MES pH 6.5, MPD 20-24%, PEG3350 2-4% y cuyos datos fueron colectados en el sincrotrón ALBA. En general los contactos cristalográficos (detectados con el servidor del programa PISA), para la estructura obtenida de estos dos conjuntos de datos son similares, con una excepción importante. En la estructura resuelta con los datos del ALBA se deduce la existencia de un contacto cristalográfico intermolecular entre los residuos de C19 de dos cadenas A (conformación abierta) pertenecientes a dos dímeros de ING5<sub>1-105</sub> que no está presente en los datos del ESRF (Figura 41). Este contacto tiene la geometría de un puente disulfuro. El refinamiento de esta estructura indica que el puente disulfuro tiene una ocupancia baja, indicando que no es un contacto estable y que probablemente está presente solo en una cierta población de las moléculas del cristal. Por esta razón, y por su ausencia en el otro conjunto de datos, no parece probable que este puente disulfuro determine la conformación asimétrica del dímero de ING5<sub>1-105</sub>. El resto de contactos cristalográficos detectados por PISA, parecen que solo están involucrados en el empaquetamiento y ordenación de las moléculas en el cristal y carecen de relevancia para explicar la estructura del homodímero asimétrico de ING5<sub>1-105</sub>, por lo cual no se han tenido en cuenta para la discusión.



**Figura 41** Puente disulfuro en ING5<sub>1-105</sub>. Refinamiento del modelo de ING5<sub>1-105</sub> (datos obtenidos en el sincrotrón ESRF) sin y con puente disulfuro. Se exponen los mapas de densidad electrónica 2F<sub>o</sub>-F<sub>c</sub> y mapa de densidad diferencia F<sub>o</sub>-F<sub>c</sub>. (A) Estructura de ING5<sub>1-105</sub> refinada sin el puente disulfuro intermolecular entre las cisteínas C19 de dos cadenas A de dos dímeros. (B) Refinamiento de la estructura de ING5<sub>1-105</sub> donde se ha impuesto el puente disulfuro intermolecular entre las cisteínas C19. Se observa que al imponer el puente disulfuro el modelo no se ajusta a los datos experimentales. Por ende no existe puente disulfuro intermolecular en la estructura de ING5<sub>1-105</sub>. Clave de mapas de densidad: F<sub>o</sub>-F<sub>c</sub> (verde positivo 3.0 sigma, rojo negativo -3 sigma), 2F<sub>o</sub>-F<sub>c</sub> (color gris) a un valor de sigma =1.

### Asignación del espectro <sup>1</sup>H-<sup>15</sup>N- HSQC asignado de ING5<sub>1-105</sub>

Con muestras perdeuteradas (pero protonadas en los grupos amida), se pudieron registrar espectros de triple resonancia de ING5<sub>1-105</sub> con calidad suficiente para la asignación de las resonancias del esqueleto y obtener información estructural detallada de esta proteína en disolución. En la figura 42 se ilustra el proceso de asignación del segmento de secuencia 35-45 en el experimento HNCA, el cual, aunque contiene en sí mismo la conectividad intraresiduo y secuencial, se complementa y se analiza conjuntamente con el experimento HN(CO)CA.

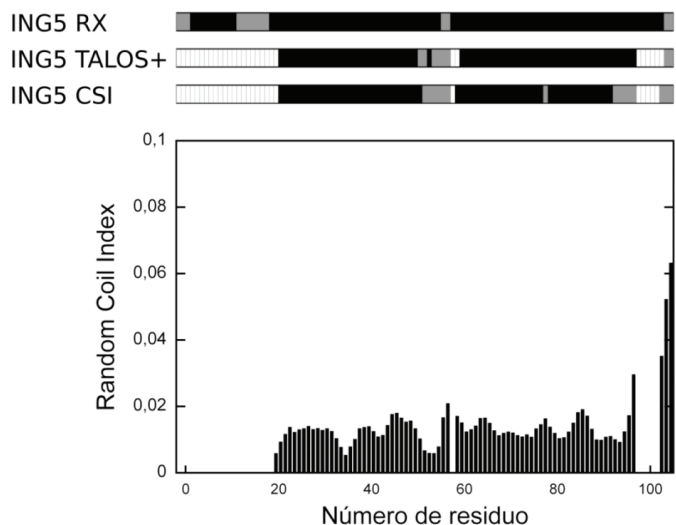


**Figura 42** Asignación de las resonancias de la cadena principal de ING<sub>5</sub><sub>1-105</sub>. Se muestra un *strip plot* del experimento HNCA para los residuos 35-45. En el eje vertical se representa la dimensión de  $^{13}\text{C}$ , mientras que en el horizontal se representa la dimensión de  $^1\text{H}$ . Cada una de las tiras es una región del plano  $^1\text{H}$  y  $^{13}\text{C}$  correspondiente a un valor de  $^{15}\text{N}$  (eje perpendicular al plano del papel). De esta manera cada tira está centrada en los desplazamientos químicos de  $^1\text{H}$  y del residuo correspondiente y los correlaciona con los carbonos  $^{13}\text{C}\alpha$  del propio residuo (señal roja intensa) y del residuo precedente (señal roja débil),

La asignación de los núcleos del esqueleto observados en los espectros de ING<sub>5</sub><sub>1-105</sub> se lista en el Anexo 2. La asignación de los grupos NH de los residuos observados se ilustra en la Figura 43.





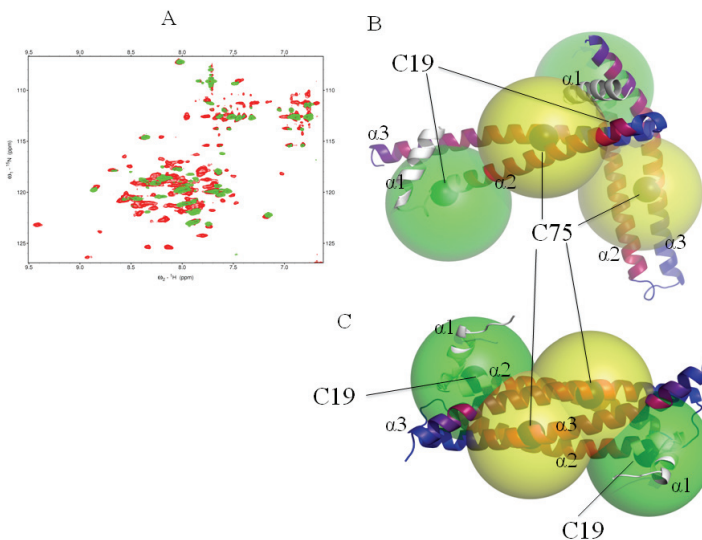


**Figura 44** Análisis de los desplazamientos químicos de las resonancias del esqueleto y C $\beta$  de ING5<sub>1-105</sub>. En la parte superior se muestran los elementos de estructura secundaria (residuos en hélice en color negro, en *coil* en gris, y no observados en blanco) identificados en la estructura cristalográfica y las determinadas por RMN usando los programas TALOS+, y CSI. El gráfico de la parte inferior muestra el valor del *Random Coil Index* (RCI). El anexo 3 contiene la lista de ángulos diedros del esqueleto derivados de TALOS+.

El análisis de los desplazamientos químicos mediante TALOS+ y CSI permite observar dos hélices con límites y longitudes similares a las hélices  $\alpha 2$  y  $\alpha 3$  observadas en el cristal. Las principales diferencias se localizan en las regiones no observadas por RMN (extremo N-terminal de  $\alpha 2$  y la región próxima al extremo C-terminal de  $\alpha 3$ ) y en la conexión  $\alpha 2$ -  $\alpha 3$  donde está el residuo P58. Teniendo esto en cuenta, y que la correlación entre los desplazamientos químicos y la estructura secundaria que se usa en estos programas es empírica (no es directa como la observada por cristalografía), no cabe concluir que haya diferencias esenciales en los elementos de estructura secundaria que se pueden observar en disolución. Los valores de RCI son un poco más altos en el extremo C-terminal de  $\alpha 3$  que en el resto de la cadena, sugiriendo un cierto grado de desorden o flexibilidad. Sin embargo los valores absolutos son muy inferiores a los típicamente obtenidos en proteínas desordenadas o en regiones flexibles de proteínas (con valores cercanos a 1), indicando que el extremo C-terminal de  $\alpha 3$  puede ser algo más flexible que el resto de la hélice.

### 1.35 Efecto de la relajación paramagnética en el espectro $^1\text{H}$ - $^{15}\text{N}$ - HSQC de ING5<sub>1-105</sub> marcado con MTSL en C19 y C75.

Aprovechando la existencia de dos residuos de cisteína en la cadena de ING5<sub>1-105</sub> (C19 y C75) se planteó realizar estas medidas para obtener información adicional sobre la estructura del dímero en disolución, ya que el análisis de los desplazamientos químicos mostrado en la figura 43 solo proporciona información de la estructura local. La sonda paramagnética MTSL unida covalentemente al grupo tiol de estos dos residuos de cisteína podría ocasionar una reducción de intensidades en las señales del espectro de ING5<sub>1-105</sub> que permitiera observar si el modo de dimerización en disolución es similar al que muestra en el cristal o similar al del dominio homólogo de ING4<sub>1-105</sub> (dímero simétrico).



**Figura 45** Efecto de la relajación paramagnética en el espectro de ING5<sub>1-105</sub> marcado con la sonda paramagnética MTSL simultáneamente en los residuos C19 y C75. (A) Espectros  $^1\text{H}$ - $^{15}\text{N}$ -HSQC de ING5<sub>1-105</sub> (muestra diamagnética en color rojo, y muestra paramagnética en color verde). (B y C) Modelos estructurales de ING5<sub>1-105</sub> exponiendo las cisteínas 19 y 75 marcadas con MTSL como esferas negras y un radio de acción de la sonda paramagnética simulado con esferas de diámetro de 15 Å de color verde (C19) y amarillo (C75). La proteína esta coloreada con un gradiente de rojo a azul, en el que el rojo representa una desaparición de la señal y el azul la ausencia de cambio de intensidad. El modelo B es el cristalográfico y el C es teórico y asume que ING5<sub>1-105</sub> dimeriza de igual forma a ING4<sub>2-105</sub>.

Comparando las intensidades de las señales correspondientes en los espectros  $^{15}\text{N}$ -HSQC de ING5<sub>1-105</sub> en forma diamagnética (sin MTSL) y paramagnética (con MTSL; Figura 45A), se observa que hay muchas señales que desaparecen, otras con intensidad disminuida y algunas que apenas se ven afectados por la acción de la sonda paramagnética. Los residuos A47, I50, V53, K54, L56, S57, Q60, V62, E63, R64, Q66, R68, I98 A102 y L104, representados en azul en las figura 45 B y C. Estos últimos se encuentran fuera de las esferas de acción de la sonda MTSL en ambas conformaciones. El resultado es que estos datos no permiten discriminar entre los dos modelos.

### **1.36 Efecto de de las mutaciones puntuales en la estabilidad de la estructura de ING5<sub>1-105</sub>**

Las tres mutaciones puntuales detectadas en cáncer en el dominio N-terminal de ING5 producen una desestabilización de la proteína aunque forman dímeros con estructura de *coiled-coil* como la proteína nativa. Con el fin de examinar la correlación entre la desestabilización observada y la estructura del dominio determinada por cristalografía se calculó el efecto desestabilizante de las tres mutaciones analizadas. Este cálculo se realizó con el programa *Site Directed Mutator* (Worth et al, 2011). Los resultados se resumen en la tabla 15, donde también se dan los resultados utilizando un modelo de ING5<sub>1-105</sub> construido sobre la estructura cristalográfica de su homólogo ING4<sub>1-105</sub> (que se muestra en la Figura 45C).

**Tabla 15** Efectos de las mutaciones puntuales en la estabilidad de la estructura de ING5<sub>1-105</sub>. Cálculos realizados en el servidor utilizando el servidor <http://mordred.bioc.cam.ac.uk/~sdm/sdm.php>.

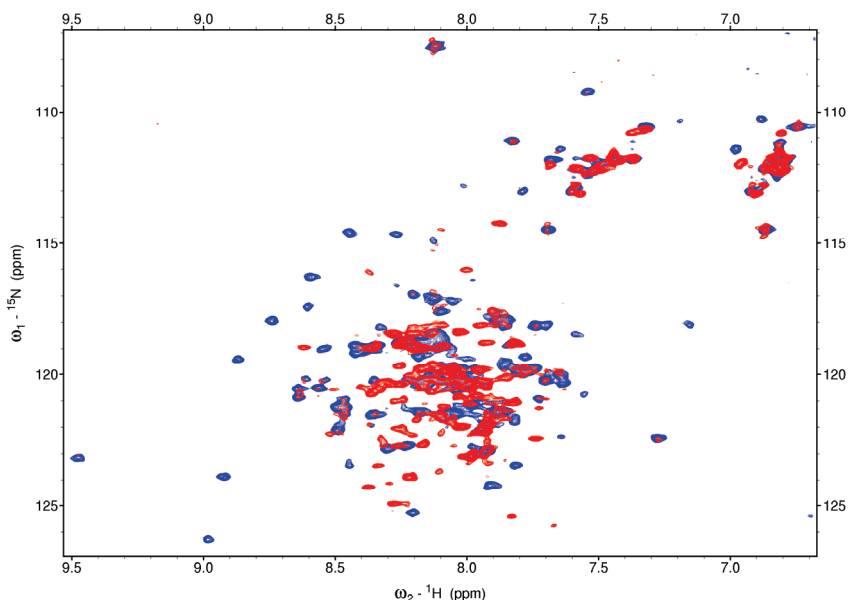
Modelos mutantes ING5 <sub>1-105</sub> basados en la estructura RX de ING5 <sub>1-105</sub>	% de accesibilidad al solvente de la proteína salvaje	% de Accesibilidad al solvente del mutante	Efecto en la estabilidad Pseudo $\Delta\Delta G$ (kcal/mol)
Q33R	75.9 (accesible)	75.7 (accesible)	-0.07 (neutral)
I68V	16.0 (enterrado)	16.9 (enterrado)	0.04 (neutral)
C75R	6.6 (enterrado)	25.1 (parcialmente accesible)	3.49 (altamente desestabilizante)
Modelos mutantes ING5 <sub>1-105</sub> basados en la estructura RX de ING4 <sub>2-105</sub>	% de accesibilidad al solvente de la proteína salvaje	% de Accesibilidad al solvente del mutante	Efecto en la estabilidad Pseudo $\Delta\Delta G$ (kcal/mol)
Q33R	72.4 (accesible)	75.5 (accesible)	-0.27 (neutral)
I68V	18.1 (parcialmente accesible)	3.5 (enterrado)	0.53 (ligeramente desestabilizante)
C75R	18.1 (parcialmente accesible)	39.6 (parcialmente accesible)	0.66 (ligeramente desestabilizante)

Como se puede observar, la mutación C75R se predice como muy desestabilizante en ambos modelos, de ING5<sub>1-105</sub> aunque más en el cristalográfico. La mutación Q33R se predice neutral en ambos y la I68V ligeramente desestabilizante en el modelo basado en el dímero simétrico de ING4. Los resultados no concuerdan del todo con los experimentales, y no discriminan entre los dos modelos.

### 1.37 Interacción ING5<sub>1-105</sub> y GST-JADE1L\_dominio IIB experimento <sup>1</sup>H-<sup>15</sup>N-HSQC

Como se describe en la sección 3.3.5, los bajos rendimientos de la purificación de JADE1L\_dominio IIB, y su estado aparentemente agregado nos obligó a realizar los experimentos de interacción con ING5<sub>1-105</sub> directamente con la proteína obtenida después de la cromatografía de afinidad a GST, la cual es una mezcla de GST sola y proteína de fusión. En el espectro de ING5<sub>1-105</sub> en presencia de GST+GST-JADE1L\_dominio IIB se observa la desaparición de un gran número de señales y/o grandes perturbaciones en los desplazamientos químicos de otras. Esto demuestra que existe una interacción. Sin embargo

un análisis detallado y fiable de los residuos de ING5<sub>1-105</sub> implicados no fue posible debido a los drásticos cambios observados y a las limitaciones en la calidad de los datos experimentales.

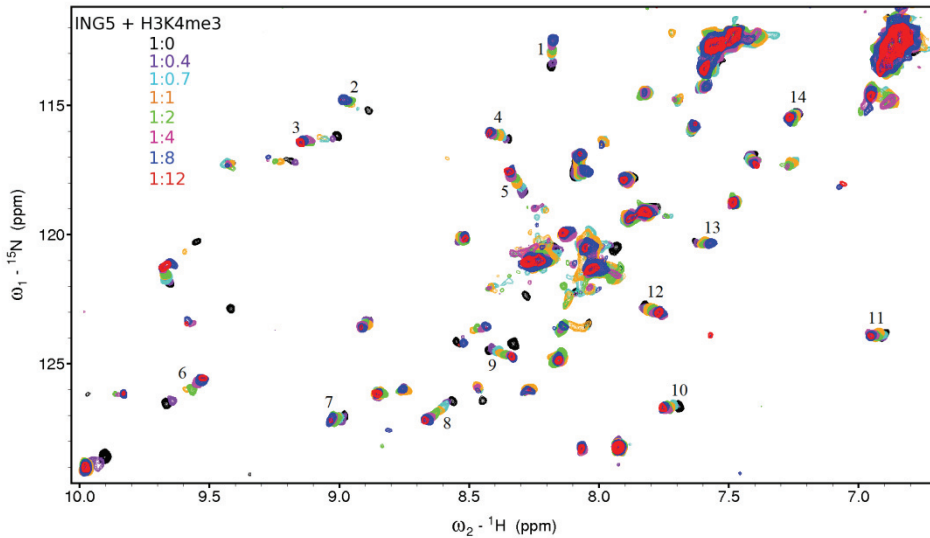


**Figura 46** Interacción ING5-JADE1L observada por RMN. Superposiciones de los espectros  $^1\text{H}$ - $^{15}\text{N}$ -TROSY de ING5<sub>1-105</sub> sin (azul) y con (rojo) GST-JADE1L\_dominioIIb en un exceso de 5 veces (molar). Condiciones de las muestras: ING5<sub>1-105</sub> 17  $\mu\text{M}$  en tampón 20 mM fosfato sódico pH 7.0. 50 mM NaCl, 5% D<sub>2</sub>O, 0.01% NaN<sub>3</sub>, 1mM DTT en ausencia o en presencia de 85  $\mu\text{M}$  de mezcla GST+GST-JADE1L\_dominio IIb. La muestra contenía también una pequeña cantidad de una mezcla de inhibidores de proteasas.

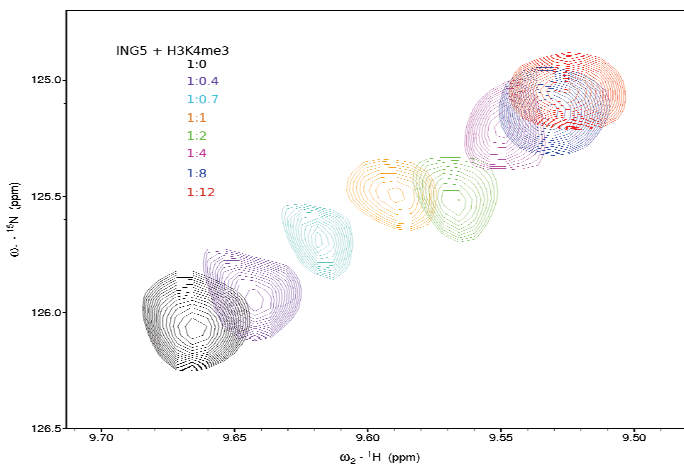
### 1.38 Interacción de ING5 con el péptido de histona H3K4me3

La interacción de ING5 con el fragmento del extremo N-terminal de histona H3 trimetilado en el residuo K4 se evaluó por RMN de la misma que se ha reportado para la proteína ING4 (Palacios et al, 2010). Las únicas diferencias están en la composición del tampón (pH 8.0, 300 mM NaCl en la muestra de ING5), debido a la baja solubilidad de ING5 en las condiciones utilizadas previamente con ING4 (pH 8.0, 300 mM NaCl). En la titulación de ING5 con el péptido se observa cómo muchas de las señales del conjunto de señales dispersas de ING5 (las correspondientes al dominio PHD) experimentan cambios en sus frecuencias de resonancia hasta alcanzarse un punto muy cercano a la saturación (Figura

47). La mayoría de las señales se observan en todos los puntos de la titulación, pero algunas desaparecen o ven disminuida su intensidad en los puntos intermedios (Figura 48).



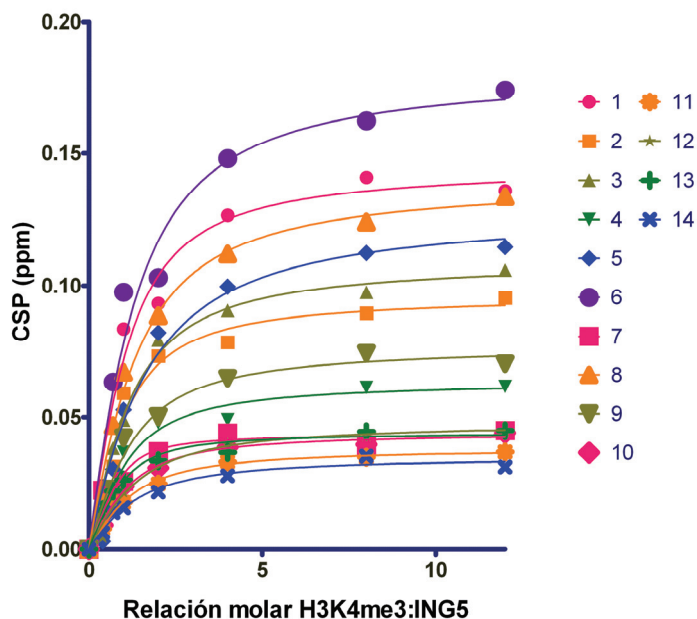
**Figura 47** Superposición de los espectros  $^1\text{H}$ - $^{15}\text{N}$  HSQC de una muestra  $11\ \mu\text{M}$  de ING5 en presencia de cantidades crecientes del péptido H3K4me3, indicadas en la figura como relaciones molares (en base al monómero de ING5). Los números adyacentes a algunas señales (14 en total) indican las señales de ING5 utilizadas para calcular la constante de disociación. Condiciones experimentales: 20 mM Tris pH 8.0, 300 mM NaCl, 1mM DTT, 0.01% de  $\text{NaN}_3$ , 800 MHz,  $25\ ^\circ\text{C}$ .



**Figura 48** Ampliación de la región de la figura, correspondiente al pico número 6.

## Resultados

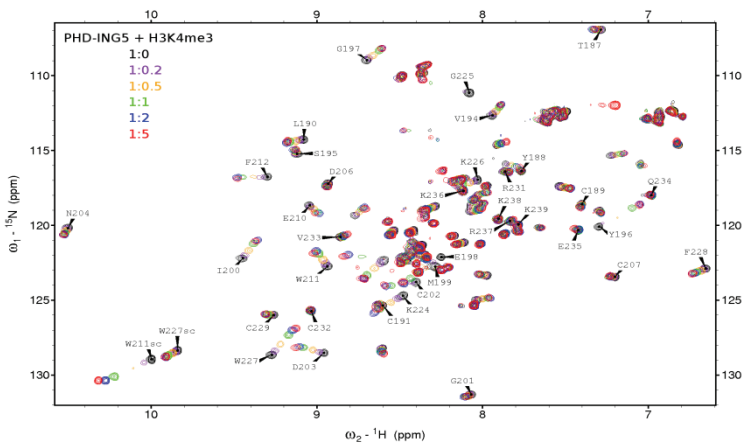
El análisis de los datos se realizó ajustando las perturbaciones de desplazamiento químico de cada señal a un equilibrio de formación de complejo con un único conjunto de sitios de unión equivalentes (ver sección 3.21). Las distintas curvas se muestran en la Figura 49, y los valores de las constantes de disociación individuales en el Anexo 4.



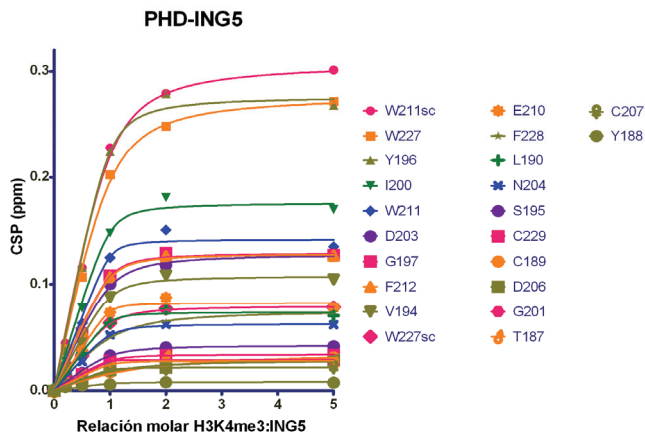
**Figura 49** Ajustes de las medidas para 14 señales del espectro de RMN de ING5 en la titulación con H3K4me3. Los datos experimentales se muestran con círculos y la simulación obtenida del ajuste con una línea del mismo color.

La constante de disociación calculada a partir de las medidas de las 14 señales mediante RMN para ING5 completa fue  $K_d = 6.7 \pm 2.3 \mu\text{M}$  (valor medio y desviación estándar). Este valor es alrededor de tres veces mayor que la calculada para la interacción entre el dominio PHD de ING5 aislado libre y un péptido H3K4me3 muy similar a partir de medidas de fluorescencia ( $K_d = 2.4 \pm 1.0 \mu\text{M}$ ). Además las constantes de afinidad calculadas para las 14 señales de ING5 completa, presentan un rango de valores que va desde 2.1 a 11.5  $\mu\text{M}$ . Para valorar mejor la posible diferencia en la afinidad de ING5 y su dominio PHD se analizaron los experimentos de RMN correspondientes a una titulación similar

realizada en el laboratorio de la Dra. Kutateladze en Colorado pero aún sin analizar (ver secciones 3.1.2 y 3.21). La superposición de los espectros se muestra en la Figura 50, el análisis en la Figura 51, y el listado de los valores de las constantes de disociación individuales en el Anexo 5.



**Figura 50** Superposición de los espectros  $^1\text{H}$ - $^{15}\text{N}$  HSQC de una muestra  $20\ \mu\text{M}$  del dominio PHD de ING5 en presencia de cantidades crecientes del péptido H3K4me3, indicadas en la figura como relaciones molares. Los números adyacentes a algunas señales indican los residuos que se pudieron asignar en base a dos espectros NOESY TOCSY editados en  $^{15}\text{N}$  de baja calidad. Condiciones experimentales: 10 mM Tris pH 6.5, 150 mM NaCl, 1mM DTT, 10 mM  $\text{NaN}_3$ , 500 MHz a  $25\ ^\circ\text{C}$ .

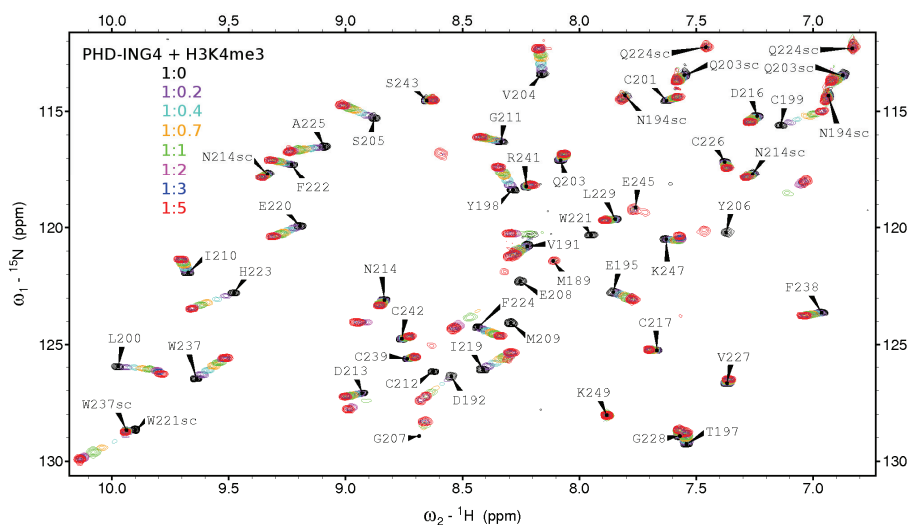


**Figura 51** Ajustes de las medidas para 22 señales del espectro de RMN del dominio PHD de ING5 en la titulación con H3K4me3 (correspondientes a los espectros de la figura 50). Los datos experimentales se muestran con círculos y la simulación obtenida del ajuste con una línea del mismo color.

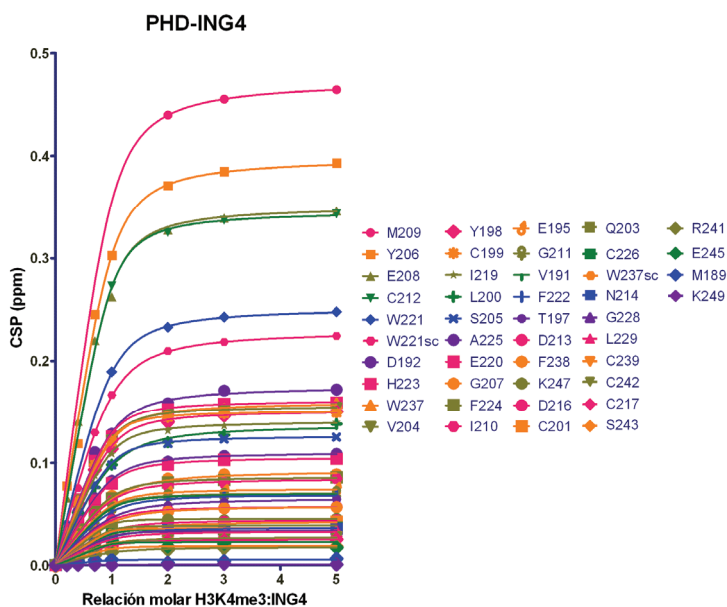


## Resultados

El resultado de este análisis da un valor de  $K_d = 12.0 \pm 10.0 \mu\text{M}$  (promedio y desviación estándar) para el dominio PHD de ING5, este valor es unas 5 veces mayor que el valor obtenido por fluorescencia para la misma interacción, y dos veces mayor que el valor medido para ING5 completa aunque en condiciones de pH y concentración de sal diferentes. Para evaluar mejor la incidencia se realizaron medidas simultáneas en varios residuos y se calculó el valor promedio así como el efecto del pH y la sal. Se analizaron los espectros correspondientes a una titulación con el dominio PHD de ING4 del que solo se había analizado el cambio en un residuo (Palacios et al, 2006). La superposición de los espectros se muestra en la Figura 52, el análisis en la Figura 53, y el listado de los valores de las constantes de disociación individuales en el Anexo 6.



**Figura 52** Superposición de los espectros  $^1\text{H}$ - $^{15}\text{N}$  HSQC de una muestra  $50 \mu\text{M}$  del dominio PHD de ING4 en presencia de cantidades crecientes del péptido H3K4me3, indicadas en la figura como relaciones molares. Los números adyacentes a las señales indican el número de aminoácido. Condiciones experimentales: 20 mM fosfato sódico pH 6.5, 50 mM NaCl, 1mM DTT, 9-10%  $\text{D}_2\text{O}$ , 0.03% de  $\text{NaN}_3$ , 600 MHz,  $25^\circ\text{C}$ .



**Figura 53** Ajustes de las medidas para un conjunto de señales del espectro de RMN del dominio PHD de ING4 en la titulación con H3K4me3 (correspondientes a los espectros de la figura 52). Los datos experimentales se muestran con círculos y la simulación obtenida del ajuste con una línea del mismo color.

Los resultados dan un valor de  $K_d = 4.0 \pm 1.2 \mu\text{M}$  (promedio más desviación estándar) para la interacción del dominio PHD de ING4 con H3K4me3. Este valor es muy similar al obtenido a partir de la señal del W237 ( $K_d = 3.9 \pm 0.7 \mu\text{M}$ ) y es aproximadamente la mitad del valor  $K_d = 7.9 \pm 2 \mu\text{M}$  calculado por fluorescencia para la misma interacción (Peña et al, 2006).

El conjunto de este análisis sugiere que las diferencias de afinidad por H3K4me3 observadas entre las proteínas ING y sus dominios no son significativas ya que se pueden explicar por las diferencias en las condiciones experimentales y el método de medida.





## **5 Discusión**



Antes de empezarse este trabajo la única información estructural disponible para la proteína ING5 era la estructura cristalográfica del dominio PHD unido a un fragmento de la histona H3. En este trabajo se ha conseguido preparar muestras de la proteína completa con la pureza y en la cantidad suficiente para su análisis estructural por dicroísmo circular, SEC-MALLS y RMN, incluida la unión de la proteína completa al fragmento de histona H3. Realizar un análisis estructural más detallado del llevado a cabo en esta tesis se ve por el bajo rendimiento en la purificación y sobre todo a la baja solubilidad de la proteína.

Los resultados obtenidos indican que la proteína ING5 humana es una proteína dimérica con tres dominios estructuralmente independientes. El dominio N-terminal está constituido por los residuos 1-105, le sigue una región central desestructurada de 69 aminoácidos y un dominio C-terminal (residuos 174-240) con estructura de PHD. La comparación de estos resultados con los obtenidos previamente para su homólogo ING4 indican que la estructura global es muy similar.

El análisis estructural del dominio N-terminal de ING5 aislado muestra que es dimérico, y que por tanto contiene el sitio de dimerización de la proteína, como se había observado para el homólogo de ING4). Los altos rendimientos en la purificación del dominio y su estabilidad permitieron intentar su cristalización, y se pudo resolver su estructura cristalográfica con una resolución de 3.2 Å<sup>2</sup>. En la estructura se observan en cada protómero tres hélices ( $\alpha 1$  de corta longitud,  $\alpha 2$  y  $\alpha 3$ , más largas, que forman un *coiled-coil*), similares a las observadas en el homólogo ING4. Sin embargo, o la estructura del dímero muestra una interfaz de dimerización diferente. Además, mientras que el dímero de ING4 es simétrico y la conformación de los dos protómeros es la misma, el dímero de ING5<sub>1-105</sub> es asimétrico y la conformación de los protómeros es distinta.

La superposición de los dos protómeros de ING5<sub>1-105</sub> (cadenas A y B en la estructura cristalográfica) muestra una diferencia global en la estructura caracterizada por un valor de RMSD = 10.3 Å para los C $\alpha$ . A nivel local las diferencias principales están en la disposición espacial de las hélices  $\alpha 1$  y de los extremos C-terminales de la hélices  $\alpha 1$ . Estas diferencias dan lugar a lo que en este trabajo se han denominado como la conformación abierta (cadena A) y cerrada (cadena B). La conformación cerrada está estabilizada por varios tipos de interacciones hidrofóbicas (Y6-I97, Y10-I97, L21-V93),

puentes salinos (H9-D101) y enlaces de hidrogeno (Y6-D101, S13-H96) que solo están presentes en la cadena B. De forma similar la conformación abierta de la cadena A se encuentra estabilizada por interacciones hidrofóbicas (A2-V93, A2 -I97, Y6-V93, Y6-I97), un puente salino (D32-K82) y un enlace de hidrogeno (Y6-H96) presentes solo en la cadena B. Las diferencias entre los dos protómeros se ajustan a la definición de cuasisimetría en la cual idénticas subunidades de un complejo adoptan una conformación similar pero no idéntica. Pero la asimetría del dímero se puede caracterizar cuantitativamente (Swapna et al, 2012). El dímero de ING5 que posee un  $GloA\_Sc = 23.21$ , un valor que es similar al del homodímero asimétrico de las subunidades  $\alpha$  y  $\beta$  de la tubulina (PDB:1JFF) y al de *adenovirus DNA-binding protein* (PDB:1ADV) con valores de  $GloA\_Sc = 23.09$  y  $GloA\_Sc = 23.42$  respectivamente. Los homodímeros globalmente asimétricos tienden a tener interfaces de interacción menores a  $1800 \text{ \AA}^2$ , en contraste con los homodímeros simétricos que pueden tener interfaces muy grandes (llegando a  $25000 \text{ \AA}^2$ , aunque la mayoría posee interfaces menores de  $5000 \text{ \AA}^2$ ). En el caso de la interfaz de dimerización de  $ING5_{1-105}$  los valores son de  $777 \text{ \AA}^2$  para la cadena A y de  $829 \text{ \AA}^2$  para la cadena B. El homodímero de  $ING5_{1-105}$  posee además valores de superficie de interfaz acordes con la teoría de homodimerización asimétrica.

Mayoritariamente los homodímeros exhiben organización simétrica (Levy et al, 2006), ya que la simetría confiere más estabilidad al complejo frente a la desnaturalización cuando se comparan con proteínas monoméricas; la simetría en oligómeros también confiere cooperatividad y unión multivalente a ligandos. Además existen otras ventajas de la oligomerización a nivel celular, como el control de la síntesis, eficiencia de la codificación y regulación del ensamblaje del oligómero. Los oligómeros simétricos son favorecidos entonces al ser más estables y tener un control en el ensamblaje para evitar la formación de matrices infinitas entre sus subunidades, las cuales se conocen son causantes de estados patológicos, como las enfermedades causadas por priones y en Alzheimer (Dobson, 2003; Wright et al, 2005). La simetría es interrumpida o modificada de muchas formas, por ejemplo la cuasi simetría, en la cual idénticas subunidades adoptan una conformación similar pero no igual; Pleomorfismo, en el cual subunidades idénticas forman diferentes

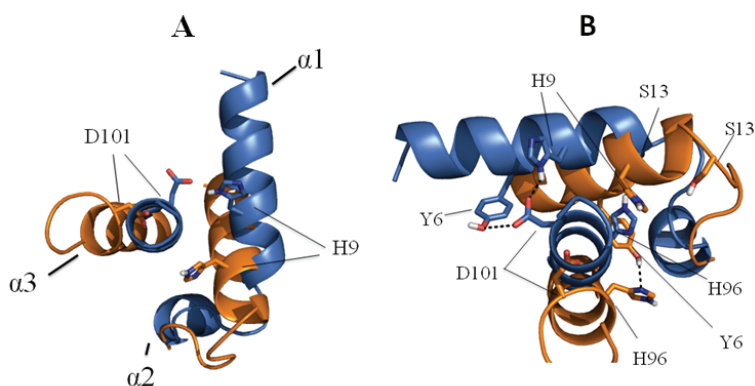
complejos; Pseudosimetría, en la cual diferentes moléculas forman complejos aproximadamente simétricos; y simetría *Mismatch* en la cual oligómeros con diferente simetría interactúan a lo largo de sus respectivos ejes de simetría. La mayoría de homodímeros exponen asimetría local a nivel de la conformación de las cadenas laterales. Se ha reportado que algunos complejos exponen alternancia entre las conformaciones simétrica y asimetría, en un ciclo donde las dos conformaciones se intercambian en el tiempo, en donde la asimetría global es rara vez observada, pero existen casos donde la asimetría permite al oligómero cumplir funciones especializadas (Goodsell & Olson, 2000). Los homodímeros biológicamente relevantes presentes en las bases de datos son predominantemente simétricos y exponen baja asimetría global. Alrededor del 76% de los homodímeros tienen un  $GloA\_Sc \leq 0.4$  y casi el 90% tiene un  $GloA\_Sc \leq 1$ . En contraste con el 3% de complejos asimétricos biológicamente relevantes encontrados en las bases de datos estructurales (Levy et al, 2008).

Entre las funciones que se le otorgan a la presencia de simetría en homodímeros, esta ayudar a prevenir la formación de matrices infinitas de subunidades, prevención de la agregación, proveer al complejo de cooperatividad y dar la oportunidad a nivel conformacional de que en un mismo complejo existan varios sitios de unión a ligandos (Dobson, 2003; Wright et al, 2005). En los homodímeros asimétricos se sugiere que existen tres mecanismos por los cuales estos evitan la formación de matrices infinitas. El primero es el uso de interfaces de dimerización parcialmente superpuestas, lo cual permite un grado de impedimento estérico para que se formen cadenas infinitas de protómeros; el segundo mecanismo es en el cual existen cambios estructurales que regulan la asociación–disociación del homodímero; y el tercero es la formación de interfaces de homodimerización débiles por las cuales el complejo es estable, pero la dimerización mediante grandes superficies de interfaz solo ocurre en presencia del ligando (Swapna et al, 2012).

En un dímero asimétrico como el de ING5 hay una interfaz de dimerización que está libre, por lo cual debe existir un mecanismo que evite la formación de una cadena ilimitada de



dímeros de ING5. En la estructura del dímero de ING5 se observa que en la interfaz de dimerización está implicada la región con la hélice  $\alpha 1$  de la conformación cerrada, mientras que esta región en la otra cadena está en la conformación abierta. Esas observaciones se podrían interpretar como que en el proceso de formación del dímero se establece la diferencia entre las conformaciones de cada protómero. En este proceso jugarían un papel especialmente relevante la flexibilidad de las hélices  $\alpha 1$  y el equilibrio de protonación-desprotonación de la histidina H9, la cual si está protonada es capaz de formar el puente salino con D101 facilitando el resto de interacciones de la conformación cerrada que están ausentes en la conformación abierta (Figura 54).

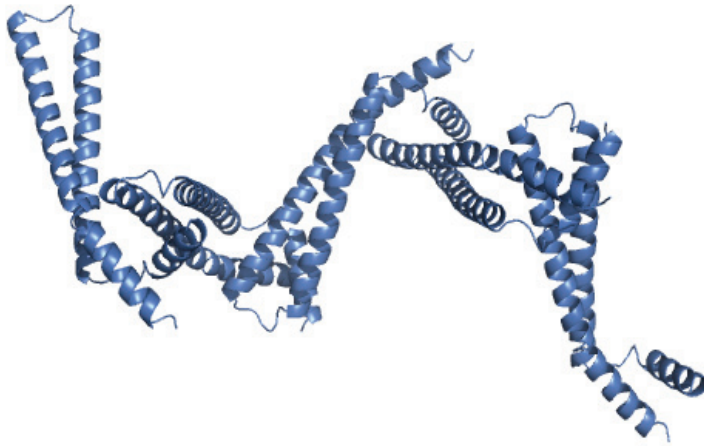


**Figura 54** Detalle de la superposición de las hélices de las cadenas A (conformación abierta) y B (conformación cerrada) del dímero de ING5. (A) En la Cadena B (azul) se observa el par iónico H9 y D101, el cual está ausente en la cadena A (naranja). La interacción electrostática permite además que se generen las otras interacciones estabilizantes de la conformación cerrada de la cadena B. (B) Detalle de la superposición de las hélices  $\alpha 1$  de las cadenas A y B del dímero de ING5. En la Cadena B (azul) se observan las interacciones estabilizantes de la conformación cerrada (Y6-D101 y S13-H96), posiblemente facilitadas por la formación del par iónico H9-D101. La cadena A en contraste no forma el par iónico H9-D101, pero mantiene el enlace de hidrogeno Y6-H96. Estos dos residuos forman cada uno enlaces de hidrogeno con diferentes residuos en la cadena B.

Si todas las moléculas se encontrasen en la conformación cerrada, sería posible la formación de una cadena infinita (Figura 56). A pHs bajos donde la cadena lateral de H9 estuviera protonada en una población cercana al 100 % el dímero formaría a agregados. En el caso particular de las condiciones de cristalización de ING5<sub>1-105</sub> el pH de la condición es pH=6.5 coincide con el valor de  $pK_a$  de la cadena lateral de una histidina, lo cual permitiría

que en la solución de cristalización la mitad de las histidinas presentes estuvieran protonadas (conformación cerrada) y la otra mitad desprotonadas (conformación abierta). La estructura cristalográfica podría representar esta situación en el contexto de un solo protómero.

Un estudio complementario a realizar en el futuro de un mutante H9A podría proporcionar información relevante sobre el papel de este residuo y su equilibrio de ionización en la formación de dímeros asimétricos.



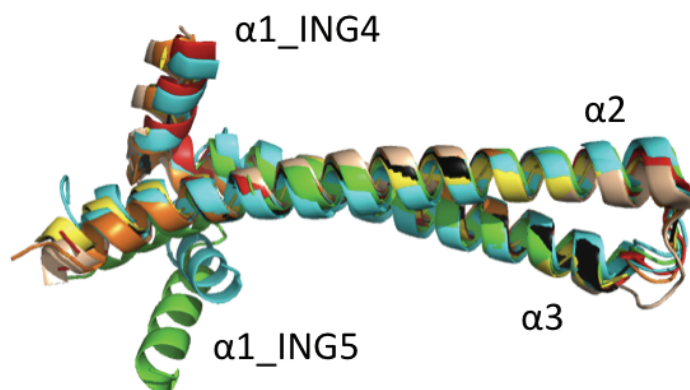
**Figura 55** Formación de cadenas infinitas de  $ING5_{1-105}$ . Cuando  $ING5$  se encuentra en una solución de  $pH < 6.5$ , la conformación estructural predominante podría ser la cerrada, la cual permite que cada una de las cadenas presentes en la solución posean una interfaz de dimerización, permitiendo la formación de una cadena infinita de  $ING5_{1-105}$  representada en la figura por la formación de un pentámero de cadenas cerradas.

Las diferencias estructurales entre el dímero cristalográfico de  $ING5_{1-105}$  y el  $ING4_{2-105}$  son muy pronunciadas, y sorprendentes considerando la alta identidad de secuencia y que el modelo construido a partir del protómero de  $ING4$  sirvió para resolver la estructura de  $ING5$  mediante reemplazamiento molecular. Aunque el reemplazamiento molecular solo tuvo éxito cuando se eliminaron de las cadenas la hélice corta  $\alpha 1$  y la conexión  $\alpha 2$  y  $\alpha 3$ . Además los residuos involucrados en la estabilización del homodímero de  $ING4_{2-105}$  están conservados en  $ING5_{1-105}$ , por lo tanto era razonable esperar que las dos proteínas

## Discusión

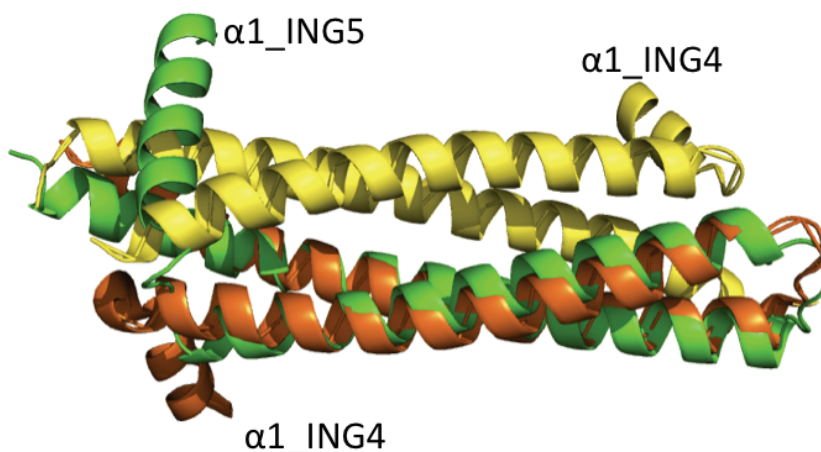
---

dimerizasen de la misma forma. ING4<sub>2-105</sub> es un homodímero con una gran superficie de interfaz de dimerización (de 1889 Å<sup>2</sup> para la cadena A y 1884 Å<sup>2</sup> para la cadena C). Con un valor de *score* de asimetría global GloA\_Sc= 0.0638 típico de un homodímero simétrico (Swapna et al, 2012).. En contraste el homodímero cristalográfico de ING5<sub>1-105</sub> es un dímero asimétrico con un valor de GloA\_Sc= 23.21. La diferencia estructural más notable entre los dominios N-terminales de ING4 e ING5 resueltos mediante cristalografía, es la ubicación espacial de las hélices  $\alpha 1$  en los protómeros. En una superposición de los protómeros de ING4 e ING5 se observa que en ING5 la hélice  $\alpha 1$  está girada unos 270 grados respecto a la hélice  $\alpha 1$  de ING4, lo cual causa que la zona de dimerización simétrica en ING4 esté “bloqueada” en ING5 (Figura 56).



**Figura 56** Superposición de las cadenas A-F de ING4 (en diferentes colores) y las cadenas A y B de ING5 (verde y cian). Se puede observar que en ING5 la hélice  $\alpha 1$  está ubicada unos 270 grados respecto a la hélice  $\alpha 1$  de ING4.

La ubicación de las cadenas  $\alpha 1$  de ING5 impide que ING5 forme dímeros simétricos del tipo ING4, ya que bloquean el acceso a la región de interacción entre los protómeros (Figura 57).



**Figura 57** Superposición ING4-ING5. Superposición del homodímero de ING5 cadenas (A: color verde y C: color cian) con la cadena A y C (colores café y amarillo) de ING4. Se observa la homología estructural entre las hélices  $\alpha 2$  y  $\alpha 3$  de ING4 y las hélices  $\alpha 2$  y  $\alpha 3$  de ING5. La conformación que adopta la hélice  $\alpha 1$  de la cadena B en ING5 permite la homodimerización asimétrica entre los dos protómeros al adoptar la conformación cerrada.

En la estructura cristalográfica de ING4<sub>2-105</sub>, los residuos de la hélice  $\alpha 1$  no establecen contactos con las hélices  $\alpha 2$  y  $\alpha 3$ , sino que parecen estabilizadas por contactos cristalográficos con otras moléculas de la red cristalina. En el cristal de ING5<sub>1-105</sub> los contactos intermoleculares de la hélice  $\alpha 1$  tienen lugar con la otra cadena del dímero exclusivamente.

Sería interesante realizar en el futuro un estudio de los dominios terminales de ING4 e ING5 en los que se haya eliminado la hélice  $\alpha 1$ , e intentar cristalizarlos. Quizás esto permitiría entender mejor su papel en la dimerización simétrica o asimétrica.

El alto grado de identidad de secuencia y la similitud estructural entre los protómeros de las estructuras cristalográficas de ING4<sub>2-105</sub> e ING5<sub>1-105</sub>, sugieren que ING5<sub>1-105</sub> podría dimerizar de forma similar a ING4<sub>2-105</sub> (dímero simétrico antiparalelo). Los residuos involucrados en las interacciones electrostáticas que estabilizan el dímero de ING4<sub>2-105</sub> están conservados en ING5<sub>1-105</sub>. La principal evidencia experimental de que ING5<sub>1-105</sub>

podría dimerizar en disolución como  $ING4_{2-105}$  proviene de los espectros de RMN de  $ING5_{1-105}$ , en los cuales solo se observa una señal por residuo (salvo en el caso de los residuos 1-20, 58 y 98-102 para los que no se detecta ninguna señal). En un dímero asimétrico se observarían dos señales de RMN en el caso de los residuos con entornos químicos diferentes en cada protómero (la mayoría si no todo en el caso del dímero cristalográfico de  $ING5_{1-105}$ ). Estos resultados indican que  $ING5_{1-105}$  puede dimerizar al menos en dos conformaciones posibles: una como la encontrada en la estructura cristalina (dímero asimétrico), y otra como el dímero observado en solución (dímero simétrico). La estructura del dímero simétrico posiblemente sea como la cristalográfica de  $ING4_{2-105}$ . Aunque se ha intentado obtener información sobre la estructura del dímero en disolución mediante medidas de relajación paramagnética, los resultados han sido ambiguos. Posiblemente esto se pueda mejorar en el futuro utilizando mutantes puntuales de cisteína de  $ING5$  para obtener moléculas con un solo grupo paramagnético MTSL unido en lugar de dos simultáneamente, que producen la desaparición de un gran número de señales.

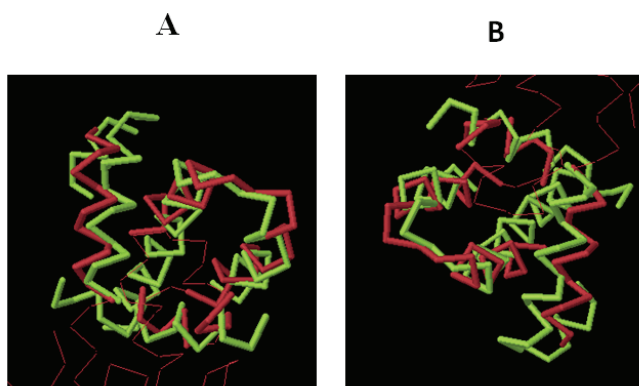
La ausencia de señales de RMN para los residuos en la región 1-19 se debe probablemente a un equilibrio conformacional en una escala de tiempo desfavorable para su observación por RMN. Aunque se ha intentado confirmar esto registrando espectros a distintas temperaturas y campos magnéticos no ha sido posible observar nuevas señales. Este equilibrio conformacional sin caracterizar probablemente esté relacionado con los modos de dimerización, ya que la región invisible está implicada en la interfaz de dimerización observada en la estructura cristalográfica de  $ING5_{1-105}$ . Los cristales de  $ING5$  se obtuvieron a pH 6.5, y los espectros de RMN a pH 7.0. Pudiera ser que la estructura cristalina corresponda a una especie que existe en equilibrio en disolución con dímeros simétricos y que se concentra en el cristal mientras que la especie simétrica no cristaliza. Las poblaciones de las distintas especies podrían estar reguladas por el equilibrio de ionización del residuo H9, como se ha discutido más arriba. La relevancia de las condiciones experimentales para la observación de un dímero u otro quizás se podría evaluar realizando experimentos de RMN a distintos pHs y también con mutantes puntuales de H9 (y quizás también de H96).

¿Puede ser el dímero asimétrico un artefacto de la cristalización? El análisis de contactos cristalográficos no apoya esta posibilidad. En la estructura resuelta con los datos obtenidos en el ALBA hay evidencia (derivada de los valores de ocupancia atómica en el refinamiento) de que existe en el cristal una mezcla de moléculas, en algunas la cisteína 19 está oxidada y en otras moléculas la cisteína 19 está reducida. Como consecuencia existen moléculas en el cristal con las formas oxidadas de la cisteína 19 con un puente disulfuro con el mismo residuo cisteína 19 perteneciente a una copia simétrica obtenida al aplicar el eje de simetría doble al cristal. En contraste este puente disulfuro no se encuentra en la estructura resuelta con los datos colectados en el ESRF, ya que al imponer el puente disulfuro en el refinamiento el resultado es un mapa de densidad residual  $F_o - F_c$  que indica que el enlace no es compatible con los datos de difracción (Figura 41). Una posible explicación para la presencia de la mezcla de moléculas con y sin puente disulfuro en los datos obtenidos del ALBA podría ser que la concentración de TCEP en el tampón de la muestra de ING5<sub>1-105</sub> con que se hicieron las gotas que generaron los cristales, no fuese suficiente para mantener reducidas todas las cisteínas de los puentes disulfuro de las moléculas presentes, por lo tanto el cristal estaría formado por moléculas con y sin enlace disulfuro. Como el empaquetamiento de la molécula de ING5<sub>1-105</sub> no cambia con la presencia del puente disulfuro, la estructura puede ser determinada y refinada con un modelo que contiene 50% de las cisteínas 19 en estado reducido y el 50% en estado oxidado. Pero la ausencia del puente disulfuro en la estructura obtenida en de los datos ESRF demuestra que este puente disulfuro no es relevante para la conformación estructural y empaquetamiento de las moléculas de ING5<sub>1-105</sub>.

Si la estructura cristalina de ING5<sub>1-105</sub> tiene relevancia fisiología tendrá que ser evaluado en futuros experimentos, pero hay argumentos que apoyan que lo es. Las interfaces de dimerización de homodímeros biológicamente relevantes suelen estar conservadas a nivel de secuencia tanto a nivel de familia de proteínas como en dominios homólogos. Estudios recientes concluyen que en la interfaz de homodímeros simétricos y asimétricos existen residuos con un alto grado de recurrencia (F, Y y L), por el contrario los residuos (I y M) son más comunes en interfaces simétricas y el residuo (Q) es común en interfaces asimétricas (Swapna et al, 2012). Otros estudios indican que los residuos (Q, D y A)

tienden a romper la simetría y que los residuos W y H son marcadores de simetría conformacional. En la interfaz de ING5 no existen marcadores de simetría (W y H) y hay tres residuos del tipo de los que rompen la simetría (Q, D y A).

La búsqueda de interfaces similares a la existente en ING5 mediante el servidor DALI, indica que la conformación estructural de esta interfaz no es singular y que muchas proteínas poseen disposiciones de las hélices alfa similar a la que presenta ING5 (Figura 58). Los resultados obtenidos por el servidor DALI (Holm & Rosenstrom, 2010), detectan 819 proteínas con un ensamblaje similar al encontrado en la interfaz del dímero de ING5. Las proteínas detectadas poseen un rango de diferencia en RMSD entre 1.2 Å y 10 Å, y un valor de Z (similitud en secuencia) entre 4.4 y 2. El resultado con la puntuación más alta (Z=4.4 y RMSD= 2.7 Å) es la estructura de la glucosiltransferasa de *Legionella pneumophila*.



**Figura 58** Interfaz de interacción similar a la de ING5. Dos vistas (relacionadas por un giro de 180 °) de la superposición de la interfaz de dimerización de ING5 con parte de la cadena A de la proteína glucosiltransferasa de *Legionella pneumophila* ( PDB: 2WZG). En color verde se muestra la interfaz de dimerización de ING5 y en color rojo los segmentos de la cadena A de la glucosiltransferasa con homología estructural con ING5.

En el análisis de la interfaz de la estructura de ING5, realizada mediante los servidores PISA y EPPIC, (Duarte et al; Krissinel & Henrick, 2007), los residuos que forman parte de la interfaz se encuentran conservados en eucariotas, tanto en homólogos de secuencia como en homólogos estructurales, sugiriendo que la interfaz tiene relevancia biológica. La conservación de residuos en la interfaz es además coherente con la estructura filogenética

de la familia de proteínas ING. La secuencia de interfaz de dimerización de ING5 tiene un 80% de identidad con ING4, pero algunas interacciones que estabilizan la interfaz de ING5 (como el puente salino E63-K82 y los enlaces de hidrogeno R61-90Y y Q66-D32) no se podrían replicar en ING4 (en el caso hipotético de que ING4 dimerizara asimétricamente). Es interesante notar que los residuos que están presentes en la interfaz de ING5 y que no están en ING4 se encuentran presentes en ING3, del cual se ha propuesto que es el miembro ancestral de la familia ING (He et al, 2005).

La conformación asimétrica expuesta en el dímero de ING5<sub>1-105</sub> resuelta en esta tesis, podría ser el resultado de un proceso evolutivo de conservación de residuos ligados a alguna de las funciones de la proteína que requiere una conformación estructural específica o la capacidad que podría tener ING5 de sufrir cambios conformacionales ya sea para formar parte de los diferentes complejos donde interactúa con otras proteínas.

La asimetría puede cumplir una función en la interacción de proteínas con sus ligandos. Uno de los casos estudiados es el homodímero (NSP3) del rotavirus. Esta proteína interactúa con el RNA y se ha reportado que la estructura asimétrica del homodímero es responsable de que la estequiometría de NP3:RNA sea 2:1 (Deo et al, 2002); (Piron et al, 1999). En otro ejemplo de la funcionalidad de la asimetría en proteínas se observa en la proteína Hsp70, la cual posee asimetría en su interfaz de dimerización, provocando variación en los sitios de unión con sus ligandos lo cual juega un papel fundamental en la regulación de estas interacciones (Nagradova, 2001). De esta forma la asimetría permite un mecanismo de unión de una subunidad proteica a un homodímero (Zhang et al, 2005). El mutante de delección ING4Δ1-56 (que no puede dimerizar) mantiene su capacidad de unirse al complejo HBO1 (Culurgioni et al, 2012). ING5 también se une al complejo HBO1 pero no existen datos experimentales que demuestren en qué estado oligomérico de ING5 se lleva a cabo esta unión, pero tomando en cuenta la similitud en secuencia y estructura entre ING4<sub>57-105</sub> y ING5<sub>57-105</sub> podemos inferir que ING5<sub>57-105</sub> también podría unirse al complejo de forma monomérica y esta unión estaría favorecida por la conformación asimétrica de la estructura del dímero de ING5, en la cual la región 57-105 está expuesta al solvente. Existen varios ejemplos donde la asimetría encontrada en homodímeros, está directamente relacionada con la forma en que estos se unen a sus ligandos. El represor transcripcional



*RevErb* el cual está presente en varias especies (Zhao et al, 1998), pertenece a la superfamilia de receptores nucleares, los cuales consisten en un complejo de diferentes factores de transcripción, que se unen a regiones específicas del ADN mediante dominios de unión al ADN los cuales forman dímeros solo en presencia de la región específica de ADN al cual se unen. La asimetría encontrada en estos homodímeros funcionalmente está implicada en la unión y la asimetría conformacional de los homodímeros se encuentra direccionada por la conformación cabeza cola del sitio de unión al ADN (Zhao et al, 1998).

En el análisis de homodímero de ING4 se presentaron tres posibles formas de unión al nucleosoma, en las cuales el homodímero antiparalelo podría unirse mediante los dos dominios PHD presentes a un solo nucleosoma, a dos nucleosomas simultáneamente y por último simultáneamente a un nucleosoma y a otro ligando posiblemente otra proteína. La estructura del homodímero de ING5 resuelta en esta tesis además de ser compatible con las anteriores formas de interacción, es factible que pudiera presentar estados de unión a complejos mediante conformación monoméricas e incluso dimerizar de forma similar a ING4 en presencia de los ligandos específicos dentro de los complejos HAT de los cuales forma parte.

El análisis de los mutantes puntuales del ING5<sub>1-105</sub> detectados en cáncer (I68V, Q33R y C75R) s indica que todos ellos mantienen una estructura dimérica helicoidal de tipo *coiled-coil* como la proteína nativa. Sin embargo su estabilidad térmica muestra que todos son menos estables.

A la luz de la estructura cristalográfica de ING5<sub>1-105</sub> se puede racionalizar esta desestabilización. La mutación I68V afecta a la interfaz de dimerización pero la nueva cadena puede establecer similares interacciones hidrofóbicas. En los otros dos casos Q33R y C75R, estas mutaciones no forman parte activa del área de dimerización asimétrica, pero están cerca de residuos que estabilizan la interfaz. La mutación Q33R queda muy cerca del enlace de hidrogeno intermolecular D32-Q66 y podría perturbarlo por interacción electrostática repulsiva con el residuo D32. Una situación similar puede presentar la mutación C75R cercana al puente salino intermolecular K66-E15.

De los tres mutantes, C75R es el menos estable, con una temperatura media aparente (debido a que el proceso es irreversible) de 28 °C, en comparación con los 47 °C del

domino nativo). Extrapolando a condiciones fisiológicas, el mutante C75R estaría parcialmente desestructurado en la célula. Al menos en este caso, se puede proponer una relación entre esta mutación y el fenotipo de cáncer de las células en las que se detectó, es decir, podría ser una mutación *driver* en contraste con las *passengers*. En el caso de las otras dos mutaciones, la desestabilización es menos drástica. La relevancia de estas mutaciones para la actividad de ING5 como supresor tumoral deberá ser confirmada mediante ensayos de proliferación celular e inducción de apoptosis en ensayos con células humanas.

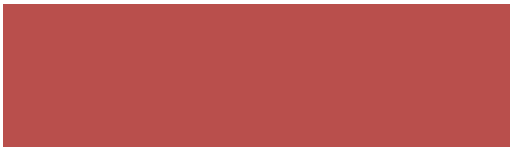
Previamente se había reportado que el dominio IIB de JADE1L, componente de complejo HAT HBO1, interacciona con ING5 de forma independiente de su dominio PHD. Esta interacción se había detectado en ensayos de *pull-down* con detección por *western blotting* utilizando proteínas fusionadas a GST. Los resultados obtenidos en esta tesis permiten confirmar que existe la interacción en disolución y que la interacción implica al dominio N-terminal de ING. Sin embargo los bajos rendimientos en la purificación de GST-JADE1L\_dominoIIB y la imposibilidad de obtener el dominioIIB aislado sin que formase agregados han impedido realizar un análisis más profundo e identificar el sitio de interacción. Quizás con nuevas construcciones del dominioIIB se puedan realizar experimentos que proporcionen esta información.

Dado que ING5 dimeriza por su dominio N-terminal, el cual está separado del dominio PHD por una región desestructurada y flexible, ING5 es un receptor bivalente de H3K4me3 con dos dominios PHD químicamente equivalentes. Como consecuencia, la proteína completa se une al péptido de histona H3K4me3 con una afinidad ( $K_d = 6.7 \pm 2.3 \mu\text{M}$ ) muy similar (esencialmente la misma teniendo en cuentas las condiciones experimentales) al su dominio PHD aislado. Así mismo, la afinidad es del mismo orden a la medida para ING4 y para su dominio PHD, lo cual es consistente con el alto grado de identidad de secuencia entre los dos dominios PHD. También es consistente con la gran similitud en el modo de unión observado en las estructuras cristalográficas de ambos complejos. Sin embargo, al igual que se ha propuesto para ING4, en el contexto de la cromatina, con dos colas flexibles de histona H3 en cada nucleosoma, la afinidad podría ser sensiblemente mayor, ya que el

## Discusión

---

segundo evento de unión tendría que pagar un coste entrópico menor que el primero. Esto se podría investigar en el futuro midiendo la unión de ING5 a nucleosomas reconstituidos (con el residuo K4 de histona H3 trimetilado) o con moléculas quiméricas que incorporen dos secuencias H3K4me3 en una cadena lineal que mimetice la disposición de las dos colas en el nucleosoma.



## **6 Conclusiones**



**CONCLUSIONES**

1. El supresor tumoral ING5 humano es una proteína dimérica constituida por tres dominios estructuralmente independientes: un dominio N-terminal plegado, un dominio central desestructurado y flexible, y un dominio C-terminal plegado de tipo PHD a través del cual interacciona con la histona H3 del nucleosoma.
2. El dominio N-terminal contiene el sitio de dimerización de ING5. Este dominio tiene una estructura predominantemente helicoidal de tipo *coiled-coil*.
3. La estructura cristalina del dominio N-terminal de ING5 consiste en un dímero asimétrico. Cada protómero está constituido por tres hélices: una hélice corta  $\alpha 1$  y dos hélices más largas  $\alpha 2$  y  $\alpha 3$  que forman una estructura de *coiled-coil*. Sin embargo la estructura global de cada protómero es diferente, existiendo una conformación abierta y otra cerrada, según la posición relativa de las hélices  $\alpha 1$  y  $\alpha 3$ .
4. La estructura cristalina del dominio N-terminal de ING5 difiere de la de su homólogo ING4, en el cual también se observan las hélices  $\alpha 1$ ,  $\alpha 2$  y  $\alpha 3$  pero con un ensamblaje simétrico del dímero. Esta diferencia podría deberse a la distinta orientación de las hélices cortas  $\alpha 1$  con respecto a las otras dos en ambas proteínas. En ING4  $\alpha 1$  no interacciona con las otras dos hélices y se aleja de ellas favoreciendo el ensamblaje de las cuatro hélices del dímero en un *coiled-coil*. En los protómeros de ING5  $\alpha 1$  interacciona con  $\alpha 3$  y se ubica de forma incompatible con el tipo de dimerización observado en ING4.
5. La estructura del dominio N-terminal de ING5 observada en disolución por RMN es simétrica, con hélices  $\alpha 2$  y  $\alpha 3$  similares a las observadas en el cristal y con los

## Conclusiones

---

primeros 19 residuos invisibles, implicados en un equilibrio conformacional que está aún por caracterizar.

6. Las interacciones observadas en la estructura cristalina y la conservación de los residuos implicados sugieren que las estructuras simétrica y asimétrica podrían estar reguladas por el equilibrio de protonación de residuos de histidina.
7. Los mutantes puntuales en el dominio N-terminal de ING5 detectados en cáncer Q33R, I68V y C75R son menos estables que la proteína nativa. Especialmente inestable es el mutante C75R, con una temperatura media aparente de desnaturalización por debajo de la fisiológica.
8. Los dominios N-terminales de ING4 e ING5 no forman heterodímeros estables detectables en experimentos de co-purificación.
9. El dominio N-terminal de ING5 contiene el sitio de interacción con el dominio IIb de la proteína JADE1L, un componente del complejo de acetilación de histonas HBO1 del que ING5 forma parte.
10. Los dos dominios PHD de la proteína ING5 dimérica son químicamente equivalentes y, por tanto, ING5 es un receptor bivalente de la cola N-terminal de histona H3 trimetilada en K4.
11. La proteína ING5 se une a un péptido correspondiente al extremo N-terminal de la histona H3 trimetilado en el residuo de lisina 4 con una constante de disociación  $K_d = 6.7 \pm 2.3 \mu\text{M}$  (valor medio y desviación estándar de medidas para varios residuos). La afinidad de la interacción es similar a las medidas para su dominio PHD aislado y también a las medidas para ING4 y su dominio PHD, teniendo en cuenta las diferencias en las condiciones experimentales utilizadas en cada caso.



## **7 Referencias**





- Aasland R, Gibson TJ, Stewart AF (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci* **20**(2): 56-59.
- Abad M, Menendez C, Fuchtbauer A, Serrano M, Fuchtbauer EM, Palmero I (2007) Ing1 mediates p53 accumulation and chromatin modification in response to oncogenic stress. *J Biol Chem* **282**(42): 31060-31067
- Aguissa-Toure AH, Wong RP, Li G (2011) The ING family tumor suppressors: from structure to function. *Cell Mol Life Sci* **68**(1): 45-54
- Akhtar A, Becker PB (2000) Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in *Drosophila*. *Mol Cell* **5**(2): 367-375
- Alva V, Syamala Devi DP, Sowdhamini R (2008) COILCHECK: an interactive server for the analysis of interface regions in coiled coils. *Protein Pept Lett* **15**(1): 33-38
- Anandakrishnan R, Aguilar B, Onufriev AV (2004) ) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* **40**(Web Server issue): W537-541
- Avvakumov N, Cote J (2007) The MYST family of histone acetyltransferases and their intimate links to cancer. *Oncogene* **26**(37): 5395-5407
- Avvakumov N, Lalonde ME, Saksouk N, Paquet E, Glass KC, Landry AJ, Doyon Y, Cayrou C, Robitaille GA, Richard DE, Yang XJ, Kutateladze TG, Cote J (2011) Conserved molecular interactions within the HBO1 acetyltransferase complexes regulate cell proliferation. *Mol Cell Biol* **32**(3): 689-703
- Bell SP, Stillman B (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**(6374): 128-134
- Berardi P, Russell M, El-Osta A, Riabowol K (2004) Functional links between transcription, DNA repair and apoptosis. *Cell Mol Life Sci* **61**(17): 2173-2180
- Berjanskii MV, Wishart DS (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc* **127**(43): 14970-14971

## Referencias

---

Berjanskii MV, Wishart DS (2008) Application of the random coil index to studying protein flexibility. *J Biomol NMR* **40**(1): 31-48

Bienz M (2006) The PHD finger, a nuclear protein-interaction domain. *Trends Biochem Sci* **31**(1): 35-40

Blanc E, Roversi P, Vonrhein C, Flensburg C, Lea SM, Bricogne G (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2210-2221

Bolin KA, Hanson P, Wright SJ, Millhauser GL (1998) An NMR investigation of the conformational effect of nitroxide spin labels on Ala-rich helical peptides. *J Magn Reson* **131**(2): 248-253

Bomar MG, Pai MT, Tzeng SR, Li SS, Zhou P (2007) Structure of the ubiquitin-binding zinc finger domain of human DNA Y-polymerase eta. *EMBO Rep* **8**(3): 247-251

Brunger AT (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**(6359): 472-475

Campos EI, Chin MY, Kuo WH, Li G (2004a) Biological functions of the ING family tumor suppressors. *Cell Mol Life Sci* **61**(19-20): 2597-2613.

Campos EI, Martinka M, Mitchell DL, Dai DL, Li G (2004b) Mutations of the ING1 tumor suppressor gene detected in human melanoma abrogate nucleotide excision repair. *Int J Oncol* **25**(1): 73-80.

Cantley LC (2002) The phosphoinositide 3-kinase pathway. *Science* **296**(5573): 1655-1657

Carvin D. I, S.A., Sternberg, M.J., Blundell, T.L., Stubbs, M.T. y Huber, R. (2006) Isomorphus replacement. En *International Tables of Crystallography. F: Crystallography of biological macromolecules* **12**: 8

Cengiz B, Gunduz E, Gunduz M, Beder LB, Tamamura R, Bagci C, Yamanaka N, Shimizu K, Nagatsuka H (2010) Tumor-specific mutation and downregulation of ING5 detected in oral squamous cell carcinoma. *Int J Cancer* **127**(9): 2088-2094

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**(3): 289-302

Cote J, Richard S (2005) Tudor domains bind symmetrical dimethylated arginines. *J Biol Chem* **280**(31): 28476-28483

Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**(Pt 9): 1002-1011

Culurgioni S, Munoz IG, Moreno A, Palacios A, Villate M, Palmero I, Montoya G, Blanco FJ (2012) The crystal structure of the inhibitor of growth 4 (ING4) dimerization domain reveals the functional organization of the ING family of chromatin binding proteins. *J Biol Chem*

Champagne KS, Saksouk N, Pena PV, Johnson K, Ullah M, Yang XJ, Cote J, Kutateladze TG (2008) The crystal structure of the ING5 PHD finger in complex with an H3K4me3 histone peptide. *Proteins* **72**(4): 1371-1376

Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 1): 12-21

Cheung KJ, Jr., Li G (2002) p33(ING1) enhances UVB-induced apoptosis in melanoma cells. *Exp Cell Res* **279**(2): 291-298

Cheung P, Allis CD, Sassone-Corsi P (2000) Signaling to chromatin through histone modifications. *Cell* **103**(2): 263-271

De Biasio A, Campos-Olivas R, Sanchez R, Lopez-Alonso JP, Pantoja-Uceda D, Merino N, Villate M, Martin-Garcia JM, Castillo F, Luque I, Blanco FJ (2012) Proliferating cell nuclear antigen (PCNA) interactions in solution studied by NMR. *PLoS One* **7**(11): e48390

Deo RC, Graft CM, Rajashankar KR, Burley SK (2002) Recognition of the rotavirus mRNA 3' consensus by an asymmetric NSP3 homodimer. *Cell* **108**(1): 71-81

Dobson CM (2003) Protein folding and disease: a view from the first Horizon Symposium. *Nat Rev Drug Discov* **2**(2): 154-160

- Doyon Y, Cayrou C, Ullah M, Landry AJ, Cote V, Selleck W, Lane WS, Tan S, Yang XJ, Cote J (2006) ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol Cell* **21**(1): 51-64
- Duarte JM, Srebniak A, Scharer MA, Capitani G (2012) Protein interface classification by evolutionary analysis. *BMC Bioinformatics* **13**: 334
- Ducruix AG, R. (1992) *Crystallization of Nucleic Acids and Proteins: A Practical Approach*.
- Dutta A, Bell SP (1997) Initiation of DNA replication in eukaryotic cells. *Annu Rev Cell Dev Biol* **13**: 293-332
- Dyson HJ, Wright PE (2005) Elucidation of the protein folding landscape by NMR. *Methods Enzymol* **394**: 299-321
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-2132
- Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**: 10
- Evans PR (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* **67**: 10
- Feng X, Hara Y, Riabowol K (2002) Different HATS of the ING1 gene family. *Trends Cell Biol* **12**(11): 532-538
- Flanagan JF, Mi LZ, Chruszcz M, Cymborowski M, Clines KL, Kim Y, Minor W, Rastinejad F, Khorasanizadeh S (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438**(7071): 1181-1185
- Folta-Stogniew E, Williams KR (1999) Determination of molecular masses of proteins in solution: Implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. *J Biomol Tech* **10**(2): 51-63

Foy RL, Song IY, Chitalia VC, Cohen HT, Saksouk N, Cayrou C, Vaziri C, Cote J, Panchenko MV (2008) Role of Jade-1 in the histone acetyltransferase (HAT) HBO1 complex. *J Biol Chem* **283**(43): 28817-28826

French GS, Wilson, K.S. (19978) TRUNCATE(part of the Groningen BIOMOL package). Incorporated into TRUNCATE by Martyn Winn.

Garkavtsev I, Grigorian IA, Ossovskaya VS, Chernov MV, Chumakov PM, Gudkov AV (1998) The candidate tumour suppressor p33ING1 cooperates with p53 in cell growth control. *Nature* **391**(6664): 295-298.

Garkavtsev I, Kazarov A, Gudkov A, Riabowol K (1996) Suppression of the novel growth inhibitor p33ING1 promotes neoplastic transformation. *Nat Genet* **14**(4): 415-420.

Garkavtsev I, Kozin SV, Chernova O, Xu L, Winkler F, Brown E, Barnett GH, Jain RK (2004) The candidate tumour suppressor protein ING4 regulates brain tumour growth and angiogenesis. *Nature* **428**(6980): 328-332.

Garman E (2003) 'Cool' crystals: macromolecular cryocrystallography and radiation damage. *Curr Opin Struct Biol* **13**(5): 545-551

Goddard TK, DG. (2008) Sparky-NMR assignment and integration software.

Gong W, Suzuki K, Russell M, Riabowol K (2005) Function of the ING family of PHD proteins in cancer. *Int J Biochem Cell Biol* **37**(5): 1054-1065

Goodsell DS, Olson AJ (2000) Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct* **29**: 105-153

Gordon PM, Soliman MA, Bose P, Trinh Q, Sensen CW, Riabowol K (2008) Interspecies data mining to predict novel ING-protein interactions in human. *BMC Genomics* **9**: 426

Gorrini C, Squatrito M, Luise C, Syed N, Perna D, Wark L, Martinato F, Sardella D, Verrecchia A, Bennett S, Confalonieri S, Cesaroni M, Marchesi F, Gasco M, Scanziani E, Capra M, Mai S, Nuciforo P, Crook T, Lough J, Amati B (2007) Tip60 is a haplo-insufficient tumour suppressor required for an oncogene-induced DNA damage response. *Nature* **448**(7157): 1063-1067

## Referencias

---

Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**(4): 305-308

Gozani O, Karuman P, Jones DR, Ivanov D, Cha J, Lugovskoy AA, Baird CL, Zhu H, Field SJ, Lessnick SL, Villasenor J, Mehrotra B, Chen J, Rao VR, Brugge JS, Ferguson CG, Payrastra B, Myszka DG, Cantley LC, Wagner G, Divecha N, Prestwich GD, Yuan J (2003) The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. *Cell* **114**(1): 99-111.

Green D (1954) The Structure of Haemoglobin. IV. Sign Determination by the Isomorphous Replacement Method. *Proc R Soc Lond* **225**(1162): 21

Grigoryan G, Keating AE (2008) Structural specificity in coiled-coil interactions. *Curr Opin Struct Biol* **18**(4): 477-483

Ha S, Park S, Yun CH, Choi Y (2002) Characterization of nuclear localization signal in mouse ING1 homolog protein. *Biochem Biophys Res Commun* **293**(1): 163-166

Han X, Feng X, Rattner JB, Smith H, Bose P, Suzuki K, Soliman MA, Scott MS, Burke BE, Riabowol K (2008) Tethering by lamin A stabilizes and targets the ING1 tumour suppressor. *Nat Cell Biol* **10**(11): 1333-1340

He GH, Helbing CC, Wagner MJ, Sensen CW, Riabowol K (2005) Phylogenetic analysis of the ING family of PHD finger proteins. *Mol Biol Evol* **22**(1): 104-116

Helbing CC, Veillette C, Riabowol K, Johnston RN, Garkavtsev I (1997) A novel candidate tumor suppressor, ING1, is involved in the regulation of apoptosis. *Cancer Res* **57**(7): 1255-1258

Hendrickson WA, Smith, J.L. y Sheriff, S. (1985) Direct Phase determination based on anomalous scattering. *Methods Enzymol* **115**: 5

Hendrickson W, Wilson, M.M. (1981) Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature* **290**: 8

Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**(Web Server issue): W545-549

Jafarnejad SM, Li G (2012) Regulation of p53 by ING family members in suppression of tumor initiation and progression. *Cancer Metastasis Rev* **31**(1-2): 55-73

Joerger AC, Rajagopalan S, Natan E, Veprintsev DB, Robinson CV, Fersht AR (2009) Structural evolution of p53, p63, and p73: implication for heterotetramer formation. *Proc Natl Acad Sci U S A* **106**(42): 17705-17710

Jung YS ZM (2004) Mars-robust automatic backbone assignment of proteins. *J Biomol NMR* (30): 12

Kabsch W (2006) Integration, scaling, space-group assignment and post refinement. *International Tables of Crystallography Volumen F: Crystallography of biological macromolecules* **11**: 8

Kahn R, Fourme, R., Bosshard, R., Chiadmi, M., Risler, J.L., Dideberg, O. y Wery JP (1985) Crystal structure study of Opsanus tau parvalbumin by multiwavelength anomalous diffraction. *FEBS Lett* **179**(1): 5

Khorasanizadeh S (2004) The nucleosome: from genomic organization to genomic regulation. *Cell* **116**(2): 259-272

Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128**(4): 693-705

Krebs JE, Kuo MH, Allis CD, Peterson CL (1999) Cell cycle-regulated histone acetylation required for expression of the yeast HO gene. *Genes Dev* **13**(11): 1412-1421

Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**(3): 774-797

Kuzmichev A, Zhang Y, Erdjument-Bromage H, Tempst P, Reinberg D (2002) Role of the Sin3-histone deacetylase complex in growth regulation by the candidate tumor suppressor p33(ING1). *Mol Cell Biol* **22**(3): 835-848

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**(5259): 680-685



## Referencias

---

Larrieu D, Ythier D, Binet R, Brambilla C, Brambilla E, Sengupta S, Pedoux R (2009) ING2 controls the progression of DNA replication forks to maintain genome stability. *EMBO Rep* **10**(10): 1168-1174

Latham JA, Dent SY (2007) Cross-regulation of histone modifications. *Nat Struct Mol Biol* **14**(11): 1017-1024

Lee J, Thompson JR, Botuyan MV, Mer G (2008) Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nat Struct Mol Biol* **15**(1): 109-111

Leslie AG (2006) The integration of macromolecular diffraction data. *Crystallogr D Biol Crystallogr* **62**: 5

Levy ED, Boeri Erba E, Robinson CV, Teichmann SA (2008) Assembly reflects evolution of protein complexes. *Nature* **453**(7199): 1262-1265

Levy ED, Pereira-Leal JB, Chothia C, Teichmann SA (2006) 3D complex: a structural classification of protein complexes. *PLoS Comput Biol* **2**(11): 155

Luger K, Rechsteiner TJ, Flaus AJ, Wayne MM, Richmond TJ (1997) Characterization of nucleosome core particles containing histone proteins made in bacteria. *J Mol Biol* **272**(3): 301-311

Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33**(2): 491-497

McFerrin MB, Snell, E. H. (2002) The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. *J Appl Cryst*(35): 7

McLachlan AD (1982) Rapid Comparison of Protein Structures. *Acta Crystallogr D Biol Crystallogr* **A38**: 3

McNamara C, Zinkernagel AS, Macheboeuf P, Cunningham MW, Nizet V, Ghosh P (2008) Coiled-coil irregularities and instabilities in group A Streptococcus M1 are required for virulence. *Science* **319**(5868): 1405-1408

Nagashima M, Shiseki M, Pedoux RM, Okamura S, Kitahama-Shiseki M, Miura K, Yokota J, Harris CC (2003) A novel PHD-finger motif protein, p47ING3,  
160

modulates p53-mediated transcription, cell cycle control, and apoptosis. *Oncogene* **22**(3): 343-350

Nagradova NK (2001) Interdomain interactions in oligomeric enzymes: creation of asymmetry in homo-oligomers and role in metabolite channeling between active centers of hetero-oligomers. *FEBS Lett* **487**(3): 327-332

Norma JG (1999) Applications of circular dichroism in protein and peptide analysis. *Trends Biochem Sci* **18**: 9

Otwinowski Z, Minor W, Charles W. Carter, Jr. (1997) Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* Vol. Volume 276, pp 307-326. Academic Press

Palacios A, Garcia P, Padro D, Lopez-Hernandez E, Martin I, Blanco FJ (2006) Solution structure and NMR characterization of the binding to methylated histone tails of the plant homeodomain finger of the tumour suppressor ING4. *FEBS Lett* **580**(30): 6903-6908

Palacios A, Moreno A, Oliveira BL, Rivera T, Prieto J, Garcia P, Fernandez-Fernandez MR, Bernado P, Palmero I, Blanco FJ (2010) The dimeric structure and the bivalent recognition of H3K4me3 by the tumor suppressor ING4 suggests a mechanism for enhanced targeting of the HBO1 complex to chromatin. *J Mol Biol* **396**(4): 1117-1127

Palacios A, Munoz IG, Pantoja-Uceda D, Marcaida MJ, Torres D, Martin-Garcia JM, Luque I, Montoya G, Blanco FJ (2008) Molecular basis of histone H3K4me3 recognition by ING4. *J Biol Chem* **283**(23): 15956-15964

Pascual J, Martinez-Yamout M, Dyson HJ, Wright PE (2000) Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. *J Mol Biol* **304**(5): 723-729.

Pedersen A, K H, Enberg J, Karlsson G (2011) Rational improvement of cell-free protein synthesis. *New Biotechnology*: in press

Peña PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, Zhao R, Kutateladze TG (2006) Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**(7098): 100-103

## Referencias

---

Peña PV, Hom RA, Hung T, Lin H, Kuo AJ, Wong RP, Subach OM, Champagne KS, Zhao R, Verkhusha VV, Li G, Gozani O, Kutateladze TG (2008) Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor. *J Mol Biol* **380**(2): 303-312

Piche B, Li G (2010) Inhibitor of growth tumor suppressors in cancer progression. *Cell Mol Life Sci* **67**(12): 1987-1999

Piron M, Delaunay T, Grosclaude J, Poncet D (1999) Identification of the RNA-binding, dimerization, and eIF4GI-binding domains of rotavirus nonstructural protein NSP3. *J Virol* **73**(7): 5411-5421

Powell HR (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr D Biol Crystallogr* **55**(Pt 10): 1690-1695

Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7**: 95-99

Redondo P, Prieto J, Munoz IG, Alibes A, Stricher F, Serrano L, Cabaniols JP, Daboussi F, Arnould S, Perez C, Duchateau P, Paques F, Blanco FJ, Montoya G (2008) Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature* **456**(7218): 107-111

Rizzo PJ (2003) Those amazing dinoflagellate chromosomes. *Cell Res* **13**(4): 215-217

Rossmann MG (2001) Molecular replacement--historical background. *Acta Crystallogr D Biol Crystallogr* **57**(Pt 10): 1360-1366

Rossmann MG (2006) Automatic indexing of data processing. *International Tables of Crystallography Volumen F: Crystallography of biological macromolecules* **11**: 3

Rupp B (2009) *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*.

Russell M, Berardi P, Gong W, Riabowol K (2006) Grow-ING, Age-ING and Die-ING: ING proteins link cancer, senescence and apoptosis. *Exp Cell Res* **312**(7): 951-961

- Saksouk N, Avvakumov N, Champagne KS, Hung T, Doyon Y, Cayrou C, Paquet E, Ullah M, Landry AJ, Cote V, Yang XJ, Gozani O, Kutateladze TG, Cote J (2009) HBO1 HAT complexes target chromatin throughout gene coding regions via multiple PHD finger interactions with histone H3 tail. *Mol Cell* **33**(2): 257-265
- Santos-Rosa H, Caldas C (2005) Chromatin modifier enzymes, the histone code and cancer. *Eur J Cancer* **41**(16): 2381-2402
- Scott M, Boisvert FM, Vieyra D, Johnston RN, Bazett-Jones DP, Riabowol K (2001a) UV induces nucleolar translocation of ING1 through two distinct nucleolar targeting sequences. *Nucleic Acids Res* **29**(10): 2052-2058
- Scott M, Bonnefin P, Vieyra D, Boisvert FM, Young D, Bazett-Jones DP, Riabowol K (2001b) UV-induced binding of ING1 to PCNA regulates the induction of apoptosis. *J Cell Sci* **114**(Pt 19): 3455-3462
- Schindler U, Beckmann H, Cashmore AR (1993) HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. *Plant J* **4**(1): 137-150
- Schrödinger. (2010) The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC. In 1.5.0.4 (ed.), 1.5.0.4
- Shen JC, Unoki M, Ythier D, Duperray A, Varticovski L, Kumamoto K, Pedoux R, Harris CC (2007) Inhibitor of growth 4 suppresses cell spreading and cell migration by interacting with a novel binding partner, liprin alpha1. *Cancer Res* **67**(6): 2552-2558
- Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**(4): 213-223
- Shiseki M, Nagashima M, Pedoux RM, Kitahama-Shiseki M, Miura K, Okamura S, Onogi H, Higashimoto Y, Appella E, Yokota J, Harris CC (2003) p29ING4 and p28ING5 bind to p53 and p300, and enhance p53 activity. *Cancer Res* **63**(10): 2373-2378.
- Sims RJ, 3rd, Chen CF, Santos-Rosa H, Kouzarides T, Patel SS, Reinberg D (2005) Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J Biol Chem* **280**(51): 41789-41792

- Soliman MA, Riabowol K (2007) After a decade of study-ING, a PHD for a versatile family of proteins. *Trends Biochem Sci* **32**(11): 509-519
- Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J Appl Cryst* **41**: 3
- Swapna LS, Srikeerthana K, Srinivasan N (2012) Extent of structural asymmetry in homodimeric proteins: prevalence and relevance. *PLoS One* **7**(5): 16
- Taipale M, Rea S, Richter K, Vilar A, Lichter P, Imhof A, Akhtar A (2005) hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol Cell Biol* **25**(15): 6798-6810
- Talbert PB, Henikoff S (2010) Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* **11**(4): 264-275
- Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14**(11): 1025-1040
- Taylor G (2003) The phase problem. *Acta Crystallogr D Biol Crystallogr* **59**(Pt 11): 1881-1890
- Theobald DL, Wuttke DS (2006) THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics* **22**(17): 2171-2172
- Thomas T, Dixon MP, Kueh AJ, Voss AK (2008) Mof (MYST1 or KAT8) is essential for progression of embryonic development past the blastocyst stage and required for normal chromatin architecture. *Mol Cell Biol* **28**(16): 5093-5105
- Tsurimoto T (1999) PCNA binding proteins. *Front Biosci* **4**: D849-858
- Ullah M, Pelletier N, Xiao L, Zhao SP, Wang K, Degerny C, Tahmasebi S, Cayrou C, Doyon Y, Goh SL, Champagne N, Cote J, Yang XJ (2008) Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes. *Mol Cell Biol* **28**(22): 6828-6843

---

Van Durme J, Delgado J, Stricher F, Serrano L, Schymkowitz J, Rousseau F (2011) A graphical interface for the FoldX forcefield. *Bioinformatics*

Vieyra D, Loewith R, Scott M, Bonnefin P, Boisvert FM, Cheema P, Pastyrzyeva S, Meijer M, Johnston RN, Bazett-Jones DP, McMahon S, Cole MD, Young D, Riabowol K (2002) Human ING1 proteins differentially regulate histone acetylation. *J Biol Chem* **277**(33): 29832-29839

Villate M, Merino N, Blanco FJ (2012) Production of meganucleases by cell-free protein synthesis for functional and structural studies. *Protein Expr Purif* **85**(2): 246-249

Voss AK, Collin C, Dixon MP, Thomas T (2009) Moz and retinoic acid coordinately regulate H3K9 acetylation, Hox gene expression, and segment identity. *Dev Cell* **17**(5): 674-686

Voss AK, Thomas T (2009) MYST family histone acetyltransferases take center stage in stem cells and development. *Bioessays* **31**(10): 1050-1061

Walshaw J, Woolfson DN (2001) Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol* **307**(5): 1427-1450

Warbrick E (1998) PCNA binding through a conserved motif. *Bioessays* **20**(3): 195-199

Warbrick E, Heatherington W, Lane DP, Glover DM (1998) PCNA binding proteins in *Drosophila melanogaster* : the analysis of a conserved PCNA binding domain. *Nucleic Acids Res* **26**(17): 3925-3932

Weiss MS (2000) Global indicators of X-ray data quality. *J Appl Crystallogr* **34**: 6

Wider G, Wuthrich K (1999) NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Curr Opin Struct Biol* **9**(5): 594-601

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ,

Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4): 235-242

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shift referencing in biomolecular NMR. *J Biomol NMR* **6**(2): 135-140

Wishart DS, Sykes BD (1994) The <sup>13</sup>C chemical-shift index: a simple method for the identification of protein secondary structure using <sup>13</sup>C chemical-shift data. *J Biomol NMR* **4**(2): 171-180

Worth CL, Preissner R, Blundell TL (2011) SDM--a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res* **39**(Web Server issue): W215-222

Wright CF, Teichmann SA, Clarke J, Dobson CM (2005) The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* **438**(7069): 878-881

Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P, Wu C, Allis CD (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**(7098): 86-90

Yang XJ, Ullah M (2007) MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells. *Oncogene* **26**(37): 5408-5419

Yee AA, Savchenko A, Ignachenko A, Lukin J, Xu X, Skarina T, Evdokimova E, Liu CS, Semesi A, Guido V, Edwards AM, Arrowsmith CH (2005) NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. *J Am Chem Soc* **127**(47): 16512-16517

Ythier D, Larrieu D, Brambilla C, Brambilla E, Pedoux R (2008) The new tumor suppressor genes ING: genomic structure and status in cancer. *Int J Cancer* **123**(7): 1483-1490

Zhang M, Windheim M, Roe SM, Peggie M, Cohen P, Prodromou C, Pearl LH (2005) Chaperoned ubiquitylation--crystal structures of the CHIP U box E3 ubiquitin ligase and a CHIP-Ubc13-Uev1a complex. *Mol Cell* **20**(4): 525-538

Zhao Q, Khorasanizadeh S, Miyoshi Y, Lazar MA, Rastinejad F (1998) Structural elements of an orphan nuclear receptor-DNA complex. *Mol Cell* **1**(6): 849-861

Zwart PH, Afonine PV, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, McKee E, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Storoni LC, Terwilliger TC, Adams PD (2008) Automated structure solution with the PHENIX suite. *Methods Mol Biol* **426**: 419-435

Zwart PH, Grosse-Kunstleve, R.W., Adams, P.D. (2005) Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation In Newsletter C (ed.), *Contribution 7*





## **8 ANEXOS**



### 1.39 Anexo 1 Secuencia del gen sintético de ING5 con codones optimizados para su expresión en E. coli.

En rojo y verde se indican los sitios de restricción NcoI y BanHI, respectivamente, y en azul el codón de terminación de la traducción.

ACCATGGGCTCTTCCCACCATCACCATCATCACTCTAGCGGTGAAAACCTGTAC  
TTCCAGGGCGCTATGGCAACCGCAATGTACCTGGAACACTATCTGGACTCTATC  
GAGAACCTGCCGTGCGAACTGCAGCGTAACTTCCAGCTGATGCGTGAACCTGGA  
TCAGCGTACCGAAGACAAAAAGCGGAGATTGACATCCTGGCTGCGGAATACA  
TCTCCACGGTGAAAACCCTGTCTCCAGACCAGCGCGTAGAACGCCTGCAGAAA  
ATCCAGAACGCCTATAGCAAATGCAAAGAATACTCTGACGACAAAGTGCAGCT  
GGCAATGCAGACCTACGAAATGGTTGACAAACACATTCGCCGTCTGGATGCTG  
ATCTCGCGGTTTTGAAGCTGACCTCAAAGACAAAATGGAGGGCTCCGATTTTCG  
AATCTTCCGGCGGTTCGTGGTCTGAAAAAAGGTCGTGGTCAGAAAGAGAAACGT  
GGTAGCCGTGGTTCGTGGTTCGTCGCACTTCTGAGGAAGACACGCCGAAAAAAAA  
AAAACACAAAGGCGGCAGCGAATTCACTGATACCATCCTGAGCGTTCATCCGT  
CTGATGTACTGGATATGCCAGTTGACCCGAACGAACCGACCTACTGCCTGTGCC  
ACCAGGTTTCCTATGGCGAGATGATTGGTTGCGATAACCCAGACTGTCCGATTG  
AATGGTTCCACTTTGCCTGCGTGGATCTGACCACTAAACCGAAAGGCAAATGGT  
TCTGTCCACGCTGTGTCCAGGAGAAACGAAAAAAAAAATAAGGATCC

**1.40 Anexo 2 Desplazamientos químicos de los núcleos asignados de la cadena principal y los  $^{13}\text{C}_\beta$  de ING5 <sup>1-105</sup>**

Residuo	AA	$^{13}\text{C}_\alpha$	$^{13}\text{C}_\beta$	$^{13}\text{CO}$	$^1\text{HN}$	$^{15}\text{N}$
-1	G	-	-	-	-	-
0	A	-	-	-	-	-
1	M	-	-	-	-	-
2	A	-	-	-	-	-
3	T	-	-	-	-	-
4	A	-	-	-	-	-
5	M	-	-	-	-	-
6	Y	-	-	-	-	-
7	L	-	-	-	-	-
8	E	-	-	-	-	-
9	H	-	-	-	-	-
10	Y	-	-	-	-	-
11	L	-	-	-	-	-
12	D	-	-	-	-	-
13	S	-	-	-	-	-
14	I	-	-	-	-	-
15	E	-	-	-	-	-
16	N	-	-	-	-	-
17	L	-	-	-	-	-
18	P	-	-	-	-	-
19	C	-	-	-	-	-
20	E	59.025	46.431	178.277	-	-
21	L	56.9	41.051	177.936	7.079	118.081
22	Q	59.434	27.547	178.991	7.816	117.985
23	R	58.374	28.507	178.397	8.103	118.108
24	N	55.757	36.386	177.869	7.533	120.158
25	F	60.978	-	178.854	8.058	118.338
26	Q	58.531	26.777	178.727	7.896	119.981
27	L	57.234	40.736	180.003	8.382	121.035
28	M	60.277	32.994	177.426	8.333	118.982
29	R	58.971	28.962	179.742	7.483	118.142
30	E	58.948	28.406	179.366	8.097	120.974
31	L	57.666	40.793	180.381	8.32	120.776
32	D	57.455	39.213	178.549	8.208	122.405

33	Q	58.304	27.228	178.105	7.925	121.088
34	R	58.849	28.55	180.02	8.284	118.528
35	T	67.302	55.492	175.864	8.119	116.827
36	E	59.02	28.081	179.928	8.143	122.279
37	D	56.851	39.398	179.197	8.573	120.481
38	K	56.038	29.291	178.192	8.007	121.578
39	K	59.87	31.074	178.753	8.544	120.201
40	A	54.432	16.836	180.394	7.801	120.954
41	E	58.836	29.106	179.021	7.66	119.529
42	I	65.878	36.838	176.735	8.272	120.004
43	D	57.38	39.567	179.552	7.755	118.324
44	I	64.185	37.054	178.762	7.451	120.674
45	L	57.234	41.818	180.846	8.368	120.76
46	A	55.066	18.19	179.387	9.398	122.909
47	A	54.717	16.909	180.794	7.659	119.566
48	E	59.055	28.555	178.86	8.07	119.02
49	Y	61.221	37.891	176.417	8.265	121.311
50	I	62.957	-	177.392	8.506	117.175
51	S	60.395	60.165	175.989	7.743	110.714
52	T	61.602	52.495	176.945	7.445	108.871
53	V	65.245	29.248	175.736	7.769	123.231
54	K	58.659	-	177.471	8.347	118.927
55	T	61.058	54.131	174.753	8.018	107.171
56	L	53.712	42.365	177.247	7.177	122.133
57	S	56.351	61.68	-	8.821	118.966
58	P	-	-	-	-	-
59	D	56.52	39.404	179.198	-	-
60	Q	57.844	28.965	179.184	7.634	119.985
61	R	60.546	29.051	178.302	8.487	120.144
62	V	65.978	30.908	178.352	7.626	117.887
63	E	59.071	28.728	179.484	7.553	119.632
64	R	57.889	28.808	179.371	8.06	117.258
65	L	57.43	40.005	179.026	8.09	119.956
66	Q	58.277	27.212	178.662	8.221	118.563
67	K	59.275	31.338	180.252	7.665	118.827
68	I	65.798	37.486	177.194	7.671	121.452
69	Q	59.429	27.473	179.151	8.522	119.809
70	N	55.391	37.454	177.616	8.659	117.487
71	A	54.91	17.303	180.295	7.797	123.917

72	Y	63.191	37.493	179.643	8.457	118.7
73	S	61.701	61.68	177.106	8.523	116.039
74	K	56.559	29.639	176.885	7.893	122.605
75	C	63.682	25.894	177.121	7.957	116.849
76	K	58.882	31.205	177.499	7.928	120.27
77	E	59.089	28.594	179.187	7.601	120.537
78	Y	57.614	36.634	178.999	8.021	116.804
79	S	61.27	61.219	177.082	8.209	114.394
80	D	57.029	39.202	179.506	8.8	123.577
81	D	56.882	39.393	179.097	8.42	121.777
82	K	59.44	30.322	177.845	8.399	121.652
83	V	66.397	30.504	177.347	7.975	118.61
84	Q	58.409	27.527	178.837	7.568	117.556
85	L	56.842	41.397	180.158	7.945	119.93
86	A	55.327	16.988	178.773	8.879	125.786
87	M	57.327	-	179.868	8.027	114.477
88	Q	58.181	28.043	178.879	7.75	117.603
89	T	67.287	-	176.345	8.247	118.061
90	Y	62.453	-	178.731	8.468	119.964
91	E	59.328	28.491	178.721	7.738	119.285
92	M	58.625	32.96	179.459	8.01	118.751
93	V	66.77	30.754	176.704	8.443	121.17
94	D	57.576	41.53	177.568	7.769	119.171
95	K	58.866	31.503	178.716	7.887	116.162
96	H	58.992	-	177.175	8.06	118.752
97	I	62.297	-	-	8.198	118.599
98	R	-	-	-	-	-
99	R	-	-	-	-	-
100	L	-	-	-	-	-
101	D	-	-	-	-	-
102	A	-	-	-	-	-
103	D	54.276	40.25	176.186	-	-
104	L	54.56	41.23	176.185	7.855	121.64
105	A	53.387	19.253	175.54	7.703	108.747

**1.41 Anexo 3. Ángulos diedros del esqueleto de ING5<sub>1-105</sub> derivados de TALOS+**

Número de Residuo	Residuo	Ángulo $\phi$ (°)	Ángulo $\varphi$ (°)
-1	G		
0	A		
1	M		
2	A		
3	T		
4	A		
5	M		
6	Y		
7	L		
8	E		
9	H		
10	Y		
11	L		
12	D		
13	S		
14	I		
15	E		
16	N		
17	L		
18	P		
19	C		
20	E		
21	L	-64.56	-38.95
22	Q	-62.1	-41.71
23	R	-65.24	-37.61
24	N	-63.52	-41.3
25	F	-66	-40.14



26	Q	-66.4	-39.07
27	L	-63.58	-40.39
28	M	-62.06	-42.57
29	R	-62.34	-41.09
30	E	-64.98	-42.46
31	L	-62.19	-39.41
32	D	-66.22	-42.75
33	Q	-67.13	-38.55
34	R	-64.29	-39.92
35	T	-65.64	-37.7
36	E	-60.23	-42.34
37	D	-69.27	-40.21
38	K	-67.11	-37.55
39	K	-60.93	-45.01
40	A	-64.34	-38.39
41	E	-64.92	-43.32
42	I	-63.95	-42.25
43	D	-63.2	-38.33
44	I	-64.18	-42.27
45	L	-62.79	-39.04
46	A	-61.67	-42.46
47	A	-63.28	-39.8
48	E	-63.56	-39.22
49	Y	-61.36	-46.98
50	I	-71.31	-36.57
51	S	-69.58	-19.38
52	T	-71.22	-25
53	V	-66.27	-41.69
54	K	-87.67	-2.738
55	T	63.782	28.369
56	L	-95.13	143.31
57	S	-74.6	130.78
58	P		
59	D		
60	Q	-58.83	-40.3

61	R	-59.07	-44.91
62	V	-65.31	-41.61
63	E	-61.11	-43.65
64	R	-65.98	-38.68
65	L	-64.08	-39.8
66	Q	-63.76	-35
67	K	-64.16	-42.83
68	I	-65.17	-42.09
69	Q	-61.39	-41.67
70	N	-61.8	-43.28
71	A	-62.43	-42.39
72	Y	-63.63	-43.15
73	S	-67	-38.42
74	K	-65.2	-38.9
75	C	-61.28	-42.49
76	K	-64.65	-37.93
77	E	-63.11	-30.79
78	Y	-70.63	-27.3
79	S	-64.23	-37.81
80	D	-65.54	-42.27
81	D	-62.21	-44.72
82	K	-65.83	-39.3
83	V	-63.24	-44.42
84	Q	-60.68	-41.45
85	L	-64.59	-39.74
86	A	-61.52	-38.79
87	M	-67.16	-34.27
88	Q	-66.64	-40.95
89	T	-64.78	-42.77
90	Y	-59.84	-44.91
91	E	-64.45	-39.19
92	M	-66.07	-41.37
93	V	-64.59	-43.1
94	D	-62.68	-43.53
95	K	-66.78	-32.98

96	H	-64.87	-40.35
97	I	-81.06	-28.22
98	R		
99	R		
100	L		
101	D		
102	A		
103	D		
104	L	-90.95	-11.62
105	A		

**1.42 Anexo 4 Constantes de disociación ( $K_d$ ) de ING5 por el péptido de histona H3K4me3 calculadas a partir de la titulación por RMN. Se indica el valor de la perturbación máxima del desplazamiento químico combinado de  $^1\text{H}$  y  $^{15}\text{N}$  (CSPmax, valor ajustado junto con el de  $K_d$ ) y el error de ajuste de la  $K_d$ ).**

ING5 FL			
Señal	$K_d$ ( $\mu\text{M}$ )	CSPmax (ppm)	Error de ajuste en $K_d$ ( $\mu\text{M}$ )
1	5.746	0.1458	2.062
2	5.304	0.09656	2.017
3	7.034	0.1099	1.119
4	5.87	0.06374	2.219
5	11.51	0.1288	4.061
6	8.314	0.1819	2.585
7	2.077	0.04386	1.357
8	8.416	0.1401	1.639
9	7.22	0.07764	1.985
10	5.614	0.04442	0.9344
11	6.659	0.0384	1.062
12	8.214	0.04818	2.056
13	3.292	0.04445	1.045
14	7.963	0.03531	2.144

**1.43 Anexo 5 Constantes de disociación (Kd) del dominio PHD de ING5 por el péptido de histona H3K4me3 calculadas a partir de la titulación por RMN. Se indica el valor de la perturbación máxima del desplazamiento químico combinado de  $^1\text{H}$  y  $^{15}\text{N}$  (CSPmax, valor ajustado junto con el de Kd) y el error de ajuste de la Kd).**

ING5			
Residuo	Kd ( $\mu\text{M}$ )	CSPmax (ppm)	Error de ajuste en Kd ( $\mu\text{M}$ )
T187	4.332	0.02819	7.91
Y188	9.858	0.008273	25.63
C189	143.4	0.03625	31.89
L190	3.722	0.07363	3.06
V194	9.036	0.1081	5.744
S195	14.33	0.04259	6.796
Y196	9.327	0.2771	6.648
G197	7.382	0.1297	4.445
I200	5.87	0.1763	4.469
G201	0.6456	0.02888	2.689
D203	15.37	0.1289	4.103
N204	7.46	0.06278	3.678
D206	48.67	0.02896	24.7
C207	2.858	0.02182	1.601
E210	2.431	0.08247	3.178
W211	3.021	0.1422	3.774
W211sc	23.19	0.3087	7.965
F212	7.768	0.1289	2.947
W227	24.82	0.2784	7.047
W227sc	11.4	0.07965	3.75
F228	37.44	0.07593	8.71
C229	7.528	0.03403	4.11

Los residuos marcados en color gris no se usaron para el cálculo de la  $K_d$  promedio debido a su bajo valor de CSP o a comportamientos anómalos de las señales a lo largo de la titulación (como desaparición de señales en puntos intermedios).

**1.44 Anexo 6 Constantes de disociación (Kd) del dominio PHD de ING4 por el péptido de histona H3K4me3 calculadas a partir de la titulación por RMN. Se indica el valor de la perturbación máxima del desplazamiento químico combinado de  $^1\text{H}$  y  $^{15}\text{N}$  (CSPmax, valor ajustado junto con el de Kd) y el error de ajuste de la Kd).**

ING4			
Residuo	Kd ( $\mu\text{M}$ )	CSPmax (ppm)	Error de ajuste en Kd ( $\mu\text{M}$ )
M189	~0.5000	0.005345	
V191	2.837	0.07032	0.642
D192	4.641	0.1757	2.133
E195	3.23	0.07502	1.004
T197	6.202	0.06576	0.7024
Y198	3.592	0.1521	0.8393
C199	2.089	0.1516	0.6082
L200	6.293	0.1381	1.475
C201	6.668	0.0432	1.857
Q203	0.8283	0.03885	0.8328
V204	2.761	0.1566	0.5872
S205	2.916	0.1269	0.6243
Y206	3.84	0.399	1.188
G207	6.733	0.09285	4.334
E208	3.345	0.3515	0.8868
M209	4.2	0.4743	0.01967
I210	4.509	0.085	0.6939
G211	2.555	0.07081	0.7403
C212	2.972	0.3467	0.254
D213	3.926	0.05801	0.3634
N214	4.809	0.03713	0.805
D216	4.062	0.04433	1.25
C217	4.408	0.02555	0.9399
I219	2.993	0.141	0.6003

E220	3.605	0.106	0.7078
W221	4.293	0.2524	0.6159
W221sc	5.471	0.2302	0.5392
F222	4.263	0.0696	0.5004
H223	2.656	0.1618	0.5283
F224	3.517	0.08676	0.8854
A225	3.956	0.1106	1.014
C226	1.744	0.03651	0.7654
G228	1.838	0.03399	0.9365
L229	3.829	0.0331	0.5215
W237	3.779	0.1602	0.6803
W237sc	~0.5000	0.03624	
F238	5.705	0.05821	0.7038
C239	2.968	0.02694	1.148
R241	6.69	0.01788	1.603
C242	4.257	0.02754	0.9027
S243	1.473	0.01881	0.7883
E245	~0.5000	0.02254	
K247	0.8603	0.04544	0.6335
K249	35.27	0.001191	65.91

Los residuos marcados en color gris no se usaron para el cálculo de la Kd promedio debido a su bajo valor de CSP o a comportamientos anómalos de las señales a lo largo de la titulación (como desaparición de señales en puntos intermedios)





## **9 Abreviaturas**



ADN	Ácido desoxirribonucleico
ARN	Ácido ribonucleico
ASA	Área de superficie accesible
ATP	Adenosina-5-trifosfato
BPTF	Bromodomain PHD finger transcription factor
BSA	Bovine Serum Albumin
CD	Dicroísmo Circular
CSP	Perturbación del desplazamiento químico
CV	Volumenes de Columna
DLS	Dynamic Light Scattering
DMSO	Dimetil sulfoxido
DSS	Sodium 2,2-dimethylsilapentane-5-sulphonate
DTT	Ditiotreitol
EDTA	EthyleneDiamineTetraacetic Acid
EPR	Electron Paramagnetic Resonance
ESRF	European Synchrotron Radiation Facility
FID	Free Induction Decay
FPLC	Fast Protein Liquid Chromatography
FRET	Transferencia de Energía por Resonancia de Fluorescencia
GdnHCl	Cloruro de Guanidinio
GST	Glutation S-transferasa
JADE	Gene (J) for Apoptosis and Differentiation in Epithelial
GST-JADE-Dominio-lib	Proteína de fusión GST con el dominio IIb de JADE
H++	Programa de predicción del pK a de las proteínas
H2A/2B/3/4	Histona 2 A. 2B. 3 y 4

H3 <sub>15</sub> K4me3	Fragmento de histona 3 (1-15) trimetilada en K4
H3K4me3	Histona 3 trimetilada en lisina 4
HAT	Acetiltransferasa de histonas
HBO1	Acetiltransferasa de histonas de unión al complejo del origen de replicación
HDAC	Desacetilasa de histonas
hEaf6	Proteína ortóloga humana del factor 6 asociado a Esa-1
HIF-1 $\alpha$	Factor de transcripción inducible por hipoxia
HMT	Metiltransferasa de histona
HNSCC	Carcinoma celular escamoso de cabeza y cuello
HPLC	High Pressure/Performance Liquid Chromatography
HSQC	Heteronuclear single quantum coherence spectroscopy
ING	Inhibitor of Growth
IPTG	Isopropil $\beta$ -D-1-tiogalactopiranosido
ITC	Calorimetría isoterma de titulación
LB	Medio de crecimiento de células Luria Bertani
LZL	Leucine-Zipper-Like. motivo tipo cremallera de leucinas
MALDI-TOF	Matrix Assited Laser Ionisation
MALLS	Multi-Angle Laser Light Scattering
MPD	2-metil-2.4-pentanodiol
MW	Peso molecular
NCS	Simetría no cristalográfica
NF- $\kappa$ B	Factor nuclear $\kappa$ B
NLS	Secuencia de localización nuclear
NOE	Nuclear overhauser effect
NOESY	Nuclear overhauser effect spectroscopy
OD	Densidad óptica
ORF	Open Reading Frame

PAGE	Electroforesis en gel de poliacrilamida
PBR	Región polibásica
PBS	Tampón fosfato salino
PCNA	Antígeno nuclear de células proliferantes
PCR	Dominio potencial de regulación de cromatina
PDB	Protein Data Bank
PDIM	Motivo de interacción dependiente de fosforilación
PEG	Polietilén glicol
PHD	Homeodominio de plantas. Dominio de dedos de zinc
pRb	Proteína de retinoblastoma
PtdInsPs	Fosfatos de fosfatidilinositol
RMN	Resonancia magnética nuclear
SAP30	Proteína asociada al complejo Sin3A
SAXS	Dispersión de rayos X a bajo ángulo
SDS	Dodecil sulfato sódico
TMS	TetraMetil Silano
TOCSY	Total Correlation Spectroscopy
TSG	Gen supresor tumoral
UV	Ultravioleta
WT	Wild type



## **10 Índices de Figuras y Tablas**





<b>Figura 1</b> Alineamiento de secuencias de las proteínas ING humanas.....	5
<b>Figura 2</b> Alineamiento de las secuencias N-terminales de las proteínas de la familia ING y un esquema de los elementos de estructura secundaria .....	7
<b>Figura 3</b> Alineamiento de secuencias de la región central de la familia de proteínas ING...	8
<b>Figura 4</b> Alineamiento de las secuencias de la región C-terminal de las proteínas ING. ...	10
<b>Figura 5</b> Modificaciones postraduccionales. ....	16
<b>Figura 6</b> Estructura de diferentes dominios de reconocimiento de modificaciones postraduccionales de histonas.....	18
<b>Figura 7</b> Modelo de dímero de ING4 unido a nucleosoma. ....	20
<b>Figura 8</b> Árbol filogenético de la familia de proteínas ING humanas. ....	21
<b>Figura 9</b> Alineamiento de secuencias de dominios de interacción tipo JADE.....	23
<b>Figura 10</b> Distribución de residuos en las estructuras coiled-coil idealizadas de ING4 e ING5 .....	25
<b>Figura 11</b> Resultados representativos de la expresión, purificación y grado de pureza de las proteínas ING5 e ING5 <sub>1-105</sub> .....	46
<b>Figura 12</b> Análisis por SDS-PAGE y separación por cromatografía de los productos de la reacción de corte de la proteína de fusión GST-JADE1L_dominio IIb.....	50
<b>Figura 13</b> Espectros de DC de proteínas.....	53
<b>Figura 14</b> Cristalización de proteínas mediante difusión de vapor. ....	57
<b>Figura 15</b> Formas cristalinas de ING51-105.....	60
<b>Figura 16</b> Representación de las interferencias constructiva (izquierda) y destructiva (derecha) de acuerdo a la ley de Bragg.....	62
<b>Figura 17</b> Imagen de difracción del cristal de ING5 <sub>1-105</sub> .....	67
<b>Figura 18</b> Representación de $I/\sigma(I)$ versus Index.....	69
<b>Figura 19</b> Representación del valor de $R_{\text{medido}}$ ( $R_{\text{meas}}$ ) versus Resolución .....	70
<b>Figura 20</b> Función de autorotación del cristal de ING5 <sub>1-105</sub> .....	70
<b>Figura 21</b> Modelo de búsqueda en MR.....	72
<b>Figura 22</b> Esquema de la espectroscopia de RMN multidimensional .....	79
<b>Figura 23</b> Experimento HSQC-TROSY .....	81
<b>Figura 24</b> Diagramas de transferencia de la magnetización. ....	83
<b>Figura 25</b> Estructura secundaria de ING5.....	94
<b>Figura 26</b> Estabilidad de ING5. Desnaturalización térmica de ING5.....	95
<b>Figura 27</b> Oligomerización de ING5. ....	96
<b>Figura 28</b> Superposición de los espectros de RMN <sup>1</sup> H- <sup>15</sup> N-HSQC de ING5 .....	97
<b>Figura 29</b> Espectros de dicroísmo circular de ING5 <sub>1-105</sub> .....	98
<b>Figura 30</b> Desnaturalización térmica de ING5 <sub>1-105</sub> y de los mutantes puntuales . ....	99
<b>Figura 31</b> Oligomerización de ING5 <sub>1-105</sub> . ....	100
<b>Figura 32</b> Modelo estructural de ING5 <sub>1-105</sub> en el cristal. ....	104
<b>Figura 33</b> Dos vistas de la estructura cristalográfica del homodímero de ING5 <sub>1-105</sub> .....	105
<b>Figura 34</b> Estructura <i>coiled-coil</i> del monómero de ING5 <sub>1-105</sub> . ....	106
<b>Figura 35</b> Superposición de las hélices $\alpha 2$ y $\alpha 3$ de ING5 <sub>1-105</sub> . ....	107
<b>Figura 36</b> Interacciones intramoleculares de la cadena A de ING5 <sub>1-105</sub> .....	109

<b>Figura 37</b> Interacciones intramoleculares de la cadena B de ING5 <sub>1-105</sub> .....	111
<b>Figura 38</b> Vista general de la interfaz de dimerización del dímero ING5 <sub>1-105</sub> .....	112
<b>Figura 39</b> Interacciones en la interfaz de dimerización de ING5 <sub>1-105</sub> .....	114
<b>Figura 40</b> Factores B de la estructura cristalográfica de ING5 <sub>1-105</sub> ,.....	115
<b>Figura 41</b> Puente disulfuro en ING5 <sub>1-105</sub> .....	117
<b>Figura 42</b> Asignación de las resonancias de la cadena principal de ING5 <sub>1-105</sub> .....	118
<b>Figura 43</b> Espectro <sup>1</sup> H- <sup>15</sup> N- TROSY de ING5 <sub>1-105</sub> .....	119
<b>Figura 44</b> Análisis de los desplazamientos químicos de las resonancias del esqueleto y Cβ de ING5 <sub>1-105</sub> .....	120
<b>Figura 45</b> Efecto de la relajación paramagnética en el espectro de ING5 <sub>1-105</sub> marcado con la sonda paramagnética MTLs.....	121
<b>Figura 46</b> Interacción ING5-JADE1L observada por RMN.....	124
<b>Figura 47</b> Superposición de los espectros <sup>1</sup> H- <sup>15</sup> N HSQC de una muestra 11 μM de ING5 en presencia de cantidades crecientes del péptido H3K4me3.....	125
<b>Figura 48</b> Ampliación de la región de la figura, correspondiente al pico número 6.....	125
<b>Figura 49</b> Ajustes de las medidas para 14 señales del espectro de RMN de ING5 en la titulación con H3K4me3.....	126
<b>Figura 50</b> Superposición de los espectros <sup>1</sup> H- <sup>15</sup> N HSQC de una muestra 20 μM del dominio PHD de ING5 en presencia de cantidades crecientes del péptido H3K4me3.....	127
<b>Figura 51</b> Ajustes de las medidas para 22 señales del espectro de RMN del dominio PHD de ING5 en la titulación con H3K4me3.....	127
<b>Figura 52</b> Superposición de los espectros <sup>1</sup> H- <sup>15</sup> N HSQC de una muestra 50 μM del dominio PHD de ING4 en presencia de cantidades crecientes del péptido H3K4me3.....	128
<b>Figura 53</b> Ajustes de las medidas para un conjunto de señales del espectro de RMN del dominio PHD de ING4 en la titulación con H3K4me3 (correspondientes a los espectros de la figura 52).....	129
<b>Figura 54</b> Detalle de la superposición de las hélices de las cadenas A (conformación abierta) y B (conformación cerrada) del dímero de ING5.....	136
<b>Figura 55</b> Formación de cadenas infinitas de ING5 <sub>1-105</sub> .....	137
<b>Figura 56</b> Superposición de las cadenas A-F de ING4 (en diferentes colores) y las cadenas A y B de ING5 (verde y cian).....	138
<b>Figura 57</b> Superposición ING4-ING5.....	139
<b>Figura 58</b> Interfaz de interacción similar a la de ING5.....	142

<b>Tabla 1</b> Interacciones proteína-proteínas descritas para las proteínas de la familia ING....	13
<b>Tabla 2</b> Resumen de las constantes de disociación (Kd) medidas para las reacciones de unión de los distintos dominios PHD de las proteínas ING y péptidos correspondientes a la región N-terminal de histona H3 con distintos grados de metilación en el residuo K4. ....	19
<b>Tabla 3</b> Listado de los cebadores empleados para la obtención de los diferentes clones....	34
<b>Tabla 4</b> Resumen de las condiciones de cristalización de ING5 <sub>1-105</sub> .....	58
<b>Tabla 5</b> Estadística de la colección de datos de difracción de los cristales de ING5 <sub>1-105</sub> refinamiento de los diferentes modelos. ....	68
<b>Tabla 6</b> Los valores de LLG ( <i>Logarithm- Likelihood Gain</i> ) y los parámetros RFZ ( <i>Rotation Function Z-score</i> ), TFZ ( <i>Translation Function Z-score</i> ) indican que el modelo no mejora al incluir una tercera copia, por ende dos copias son suficientes para continuar con el proceso de refinamiento. ....	73
<b>Tabla 7</b> Parámetros de los experimentos de RMN utilizados en la asignación de ING5 <sub>1-105</sub> . ....	84
<b>Tabla 8</b> Ratios de la señal de dicroísmo circular $\theta_{222}/\theta_{208}$ en los espectros de las distintas proteínas ING5, y temperaturas en el punto medio de la desnaturalización térmica medida a partir del cambio en la señal a 222 nm. ....	102
<b>Tabla 9</b> Medidas de masa molar de las diferentes proteínas ING5 mediante SEC-MALLS. ....	103
<b>Tabla 10</b> Elementos de estructura secundaria en las cadenas A y B de la estructura cristalográfica de ING5 <sub>1-105</sub> . Los límites de las hélices fueron identificados utilizando el servidor DSSP ( <a href="http://www.cmbi.ru.nl/hsspsoap">www.cmbi.ru.nl/hsspsoap</a> ).....	105
<b>Tabla 11</b> Análisis de la secuencia de héptadas presente en las secuencias de las hélices $\alpha 2$ y $\alpha 3$ de ING5 <sub>1-105</sub> . ....	106
<b>Tabla 12</b> Interacciones intramoleculares de la cadena A. Los asteriscos indican aquellas interacciones exclusivas de la cadena A (no presentes en la cadena B) del homodímero. ....	108
<b>Tabla 13</b> Interacciones intramoleculares de la cadena B. Los asteriscos indican aquellas interacciones exclusivas de la cadena B (no presentes en la cadena A) del homodímero. ....	110
<b>Tabla 14</b> Interacciones intermoleculares del homodímero de ING5 <sub>1-105</sub> . Las interacciones y fueron identificadas con el programa Coilcheck (Alva et al, 2008).....	113
<b>Tabla 15</b> Efectos de las mutaciones puntuales en la estabilidad de la estructura de ING5 <sub>1-105</sub> . Calculos realizados en el servidor utilizando el servidor <a href="http://mordred.bioc.cam.ac.uk/~sdm/sdm.php">http:// mordred. bioc. cam. ac. uk/~ sdm /sdm.php</a> .....	123