

# NANOMECÁNICA DEL CELULOSOMA

Implicaciones para la actividad del sistema

ALBERT GALERA PRAT

Departamento de Biología Molecular  
Facultad de Ciencias  
Universidad Autónoma de Madrid  
Octubre 2015

Universidad Autónoma de Madrid  
Facultad de Ciencias  
Departamento de Biología Molecular

# **Nanomecánica del celulosoma**

## Implicaciones para la actividad del sistema

Memoria presentada por **Albert Galera Prat**  
para optar al grado de Doctor en Biología Molecular.

El presente trabajo de tesis doctoral ha sido desarrollado en el  
Departamento de Neurobiología Molecular, Celular y del Desarrollo del  
Instituto Cajal (IC-CSIC), bajo la dirección del  
**Dr. Mariano Sixto Carrión Vázquez.**

*Madrid, octubre 2015*



Memoria presentada por **ALBERT GALERA PRAT**, licenciado en Bioquímica, para optar al grado de DOCTOR EN BIOLOGÍA MOLECULAR POR LA UNIVERSIDAD AUTÓNOMA DE MADRID.

**MARIANO-SIXTO CARRIÓN VÁZQUEZ, Doctor en Biología y Científico Titular del CSIC,**

**AUTORIZA** la presentación de la Tesis Doctoral titulada “**Nanomecánica del celulosoma: implicaciones para la actividad del sistema**”, de la que es autor el Lcdo. Albert Galera Prat y ha sido realizada bajo mi dirección en el Instituto Cajal del CSIC, Madrid.

Y para que así conste a los efectos oportunos, firmo el presente certificado en Madrid, el día 20 de octubre de 2015.

Fdo.: Mariano Sixto Carrión Vázquez

Instituto Cajal, CSIC  
Madrid

## AGRADECIMIENTOS

---

En primer lugar debo agradecer a mi director de tesis, Mariano Carrión Vázquez, por darme la oportunidad de trabajar en su laboratorio y participar en muchos de los proyectos en marcha. Gracias también por introducirme en la nanomecánica de proteínas, por la posibilidad de trabajar en un sistema de interés biotecnológico como es el celulosoma y por todas las discusiones que nos han llevado a mejorar el trabajo día a día. Finalmente quiero agradecer la confianza que ha depositado en mi y todas las oportunidades que me ha dado de participar en el diseño y solicitud de proyectos relacionados con la temática de la tesis.

Mi agradecimiento también a nuestros colaboradores del Weizmann Institute of Science, Edward A. Bayer y Yoav Barak, por su ayuda, consejo y buena disposición para la discusión y el diseño de algunos de los experimentos de esta tesis. Así como a todos los socios del consorcio CellulosomePlus por su entusiasmo que en todo momento me ha motivado para seguir adelante.

Debo agradecer igualmente a Jorgina Satrústegui por su ayuda y apoyo como tutora durante estos años. A la Fundación Ferrer y al MECD por las becas que me han permitido realizar esta tesis.

A Andrés, Javier, Ruben, Àngel, Pacho, Alex, Oscar, Laura, Mari Carmen, Isa, Adrián, Emilio, Ana, Anahí, Hector, Juan Carlos, Patricia, Elena (las tres), María, Marta, Aída y a todos los estudiantes que han ido pasando por el laboratorio durante todos estos años. Os quiero agradecer vuestra paciencia conmigo y todo lo que he aprendido de vosotros, por todos los buenos momentos que hemos pasado juntos y también por ayudarme a pasar los malos. Sin vuestra ayuda no habría llegado hasta aquí.

A mi familia, sobretudo a mis padres y a mi hermana que siempre me han apoyado en todo y me han animado a seguir adelante.

Finalmente quiero agradecer especialmente a Anna por cada día que hemos pasado juntos. Por su confianza y apoyo, por creer en mi, por su infinita paciencia conmigo y por ayudarme a crecer.



## SUMMARY

---

Cellulose is the major biopolymer and carbon source on the biosphere. It is composed of glucose, a sugar that can be processed by several microorganisms to produce added-value chemicals such as biofuels. The major bottleneck towards its utilization in industry is the saccharification step, i. e. the breakdown of the polymer to release the constituent soluble sugars. In nature, this process is carried out by several microorganisms with specialized enzymatic machinery. In particular, some anaerobic bacteria have developed a high molecular weight complex known as the cellulosome. This comprises a large usually non-catalytic protein called scaffoldin, capable of docking several enzymes and targeting them to the cellulosic substrate while binding it to the cell surface.

As an adhesion system, we expect that the cellulosome is subjected to mechanical stress since the relative movement of the cell and its substrate would stretch the connecting region: the portion of the scaffoldin located between the two anchoring points. We hypothesized that the mechanical stability, i. e. the resistance to forced unfolding, of this region might be important to understand cellulosome activity since the unfolding of cohesins, which are the scaffoldin modules capable of binding enzymes, might result in enzyme release and activity loss.

We have used atomic force microscopy-based single molecule force spectroscopy in combination with molecular dynamics simulations to study the mechanical properties of several cohesins from different cellulosomes. We found that those cohesins located in the connecting region of scaffoldins that bind enzymes are more mechanically stable than those located outside the two anchoring points, in the external region. This principle applies for cohesins of different cellulosomes despite the sequence difference of these modules. We also study the effects of both the linker sequences located between cohesins and enzyme binding on the mechanical stability of cohesins which are elements that are found in natural cellulosomes. We did not observe any effect on cohesin mechanical stability nor on their mechanical clamp position.

## RESUMEN

---

La celulosa es el biopolímero y la fuente de carbono más abundante de la biosfera. Está compuesta por glucosa, un azúcar que puede ser procesado por microorganismos para producir productos químicos con valor añadido como los biocombustibles. El mayor cuello de botella para su utilización industrial se encuentra en el paso de la sacarificación, es decir, la degradación de este polímero para liberar los azúcares solubles constituyentes. En la naturaleza este proceso lo llevan a cabo algunos microorganismos que presentan una maquinaria enzimática especializada. En particular, algunas bacterias anaerobias han desarrollado un complejo conocido como celulosoma que comprende una proteína de elevado peso molecular y generalmente no catalítica llamada escafoldina, capaz de unir varias enzimas y dirigirlas al sustrato a la vez que las ancla a la superficie celular.

Por el hecho de tratarse de un sistema de adhesión, es de esperar que el celulosoma esté sometido a estrés mecánico dado que el movimiento relativo de la bacteria respecto al sustrato podría estirar la región conectora: la porción de la escafoldina que se encuentra entre los dos puntos de anclaje. Propusimos la hipótesis de que la estabilidad mecánica, es decir, la resistencia al desplegamiento por fuerza, de esta región puede ser importante para entender la actividad del celulosoma. Esto es debido a que el desplegamiento de las cohesinas, los módulos de la escafoldina capaces de unir enzimas, podría conducir a una pérdida de actividad.

Hemos empleado espectroscopia de fuerza de moléculas individuales mediante microscopía de fuerza atómica en combinación con simulaciones de dinámica molecular para estudiar las propiedades mecánicas de varios módulos cohesina de distintos celulosomas. Hemos observado que las cohesinas situadas en la región conectora presentan una mayor estabilidad mecánica comparada con la de aquellas que no se espera que estén sometidas a estrés mecánico. Este principio es aplicable a las cohesinas de distintos celulosomas a pesar de las diferencias de secuencia entre ellas. También hemos estudiado el efecto de las secuencias intermodulares y de la unión a enzimas sobre la estabilidad mecánica de las cohesinas, que son elementos que se encuentran en los celulosomas naturales. Hemos visto que ni la estabilidad mecánica de las cohesinas ni la posición de su parche mecánico se ven afectados por estos factores.

## ÍNDICE GENERAL

---

<b>I</b>	<b>INTRODUCCIÓN</b>	<b>1</b>
1	INTRODUCCIÓN	3
1.1	Lignocelulosa	3
1.2	Sistemas celulolíticos	5
1.2.1	Celulosoma	7
1.2.2	Sinergia	13
1.2.3	Mini-celulosomas	14
1.3	Bioprocesos consolidados	14
1.4	Mecánica en biología	15
1.5	Espectroscopia de fuerza de moléculas individuales	17
1.5.1	Modo de velocidad constante	19
1.5.2	Interpretación	19
1.5.3	Identificación de las proteínas	22
1.6	Simulaciones de dinámica molecular	24
1.7	Mecánica del celulosoma	26
<b>II</b>	<b>OBJETIVOS</b>	<b>29</b>
2	OBJETIVOS	31
<b>III</b>	<b>MATERIALES Y MÉTODOS</b>	<b>33</b>
3	MATERIALES Y MÉTODOS	35
3.1	Nomenclatura empleada para las cohesinas	35
3.2	Clonación molecular	35
3.2.1	Clonación de hetero-poliproteínas en pAFM	36
3.2.2	Homo-poliproteína de cohesina	38
3.3	Expresión	39
3.4	Purificación	39
3.4.1	Afinidad por Ni <sup>2+</sup>	39
3.4.2	Purificación por intercambio iónico	39
3.4.3	Purificación por exclusión molecular	40
3.5	Análisis de proteínas	40
3.5.1	Western-Blot	40
3.5.2	Electroforesis en condiciones nativas	40
3.6	Espectroscopia de fuerza de moléculas individuales	41
3.6.1	Cubreobjetos	41
3.6.2	Aparato y micropalancas	41
3.6.3	Análisis de SMFS	42
3.7	Modelado estructural	43
3.8	Simulaciones	43

3.8.1	Simulación del desplegamiento mecánico mediante simulaciones de Monte Carlo	44
<b>IV</b>	<b>RESULTADOS</b>	47
4	RESULTADOS	49
4.1	Nanomecánica de cohesinas individuales	49
4.1.1	Nanomecánica de cohesinas de <i>C. thermocellum</i>	49
4.1.2	Nanomecánica de cohesinas de <i>C. cellulolyticum</i>	56
4.1.3	Nanomecánica de cohesinas de <i>A. cellulolyticus</i>	58
4.2	Nanomecánica de fragmentos del celulosoma	59
4.2.1	Nanomecánica de cohesinas con secuencias intermodulares	60
4.2.2	Nanomecánica de cohesinas unidas a doquerina	64
<b>V</b>	<b>DISCUSIÓN</b>	69
5	DISCUSIÓN	71
5.1	Nanomecánica de cohesinas de <i>C. thermocellum</i>	74
5.2	Nanomecánica de cohesinas de <i>A. cellulolyticus</i>	76
5.3	Nanomecánica de cohesinas de <i>C. cellulolyticum</i> y secuencias intermodulares	77
5.4	Nanomecánica de cohesinas interaccionando con doquerina	79
5.5	Generalidad de la hipótesis	80
<b>VI</b>	<b>CONCLUSIONES</b>	83
6	CONCLUSIONES	85
	BIBLIOGRAFÍA	87
<b>VII</b>	<b>APÉNDICE</b>	103
A	RESUMEN DE LAS FUERZAS DE DESPLEGAMIENTO	105

## ÍNDICE DE FIGURAS

---

Figura 1	Composición de la pared celular de plantas.	4
Figura 2	Modo de acción de celulasas.	6
Figura 3	El celulosoma de <i>C. thermocellum</i> .	8
Figura 4	Estructuras de componentes del celulosoma.	9
Figura 5	Diversidad de celulosomas.	12
Figura 6	Sistema Lpp-OmpA para presentar proteínas en la superficie de <i>E. coli</i> .	16
Figura 7	Microscopio de fuerza atómica.	18
Figura 8	Experimento de espectroscopia de fuerza de moléculas individuales por AFM en el modo de velocidad constante.	20
Figura 9	Efecto de la aplicación de fuerza sobre la barrera energética.	21
Figura 10	Incremento de longitud de contorno de una proteína.	24
Figura 11	Hipótesis mecánica del celulosoma.	27
Figura 12	Alineamiento de las secuencias de cohesinas de CipA de <i>C. thermocellum</i> .	50
Figura 13	Resultados SMFS módulos cohesina de <i>C. thermocellum</i> .	51
Figura 14	Dependencia de la velocidad.	52
Figura 15	Validación del modelado estructural.	54
Figura 16	Simulaciones del estirado de CtA1 y CtA9.	56
Figura 17	Alineamiento de cohesinas de CipC de <i>C. cellulolyticum</i> .	57
Figura 18	Resultados CcC8.	58
Figura 19	Resultados de nanomecánica de cohesinas de <i>A. cellulolyticus</i> .	60
Figura 20	Resultados CtA7 con secuencias intermodulares.	61
Figura 21	Resultados SMFS y SMD de secuencias intermodulares en cohesinas de <i>C. cellulolyticum</i> .	62
Figura 22	Resultados de simulaciones del estirado de CcC3.	64
Figura 23	Control interacción pAFM-CtA2:doquerina.	65
Figura 24	Resultados de nanomecánica de cohesinas en presencia de doquerina.	66
Figura 25	Resultados simulaciones estirado de complejos cohesina-doquerina.	67
Figura 26	Correlación de las fuerzas medidas con otros parámetros.	72
Figura 27	Ausencia de intermedios en el desplegamiento de cohesinas mediante espectroscopia de fuerza de moléculas individuales.	73
Figura 28	Comparación entre CtA1 y CtA2 o CtA7 y CtA9.	74
Figura 29	Diferencias entre CcC1 y CcC3.	79
Figura 30	Región de unión a doquerina y parche mecánico de una cohesina.	80

## ACRÓNIMOS

---

AFM	microscopio de fuerza atómica (en inglés Atomic Force Microscope)
CBM	módulo de unión a carbohidratos (en inglés Carbohydrate Binding Module)
CBP	bioproceso consolidado (en inglés Consolidated BioProcessing)
CEM	complejo celulosa-enzimas-microorganismo
CG	modelo de grano grueso (en inglés Coarse-Grain)
$\Delta L_C$	incremento de longitud de contorno
$\Delta x_D$	distancia entre el estado nativo y el estado de transición en la coordenada de reacción.
$F_D$	fuerza de desplegamiento
I27	módulo inmunoglobulina I27 de titina cardíaca humana
IPTG	Isopropil- $\beta$ -D-tiogalactopiranosido
$L_C$	longitud de contorno
MD	dinámica molecular (en inglés Molecular Dynamics)
NTA	ácido nitrilotriacético
$OD_{595}$	densidad óptica medida a 595 nm
$p$	longitud de persistencia
PDB	Protein Data Bank
RMSD	raíz de la desviación media cuadrática
RMSF	raíz de la fluctuación media cuadrática
SLH	dominios de homología de capa de superficie (en inglés Surface Layer Homology)
SMD	dinámica molecular dirigida (en inglés Steered Molecular Dynamics)
SMFS	espectroscopía de fuerza de moléculas individuales (en inglés Single Molecule Force Spectroscopy)
WLC	cadena gusaniforme (en inglés Worm Like Chain)

Parte I

INTRODUCCIÓN

## INTRODUCCIÓN

---

La vida en la tierra depende mayoritariamente de la fotosíntesis, a partir de la cual se produce la biomasa siendo la celulosa el componente principal [Lynd et al., 2002]. La energía proveniente del sol y fijada en el carbono mediante la fotosíntesis es diez veces superior a la consumida en el mundo [Demain et al., 2005], su utilización por parte de microorganismos es responsable de uno de los mayores flujos de materia de la biosfera y además es el sustento de los rumiantes, que constituyen la mayor fuente de proteína alimentaria [Lynd et al., 2002]. Por todo ello, se puede concluir que la fotosíntesis aporta de forma directa o indirecta la mayor parte de la energía usada por la vida en la tierra [Wei et al., 2009].

Con el fin de utilizar la energía almacenada en la biomasa nace el concepto de biorefinería, en contraposición a una refinería de petróleo, en la que se integra tanto el tratamiento como la conversión de la biomasa para producir biocombustibles, energía y productos químicos de interés [Hasunuma et al., 2013].

El interés por el uso de biomasa lignocelulósica como fuente de energía se fundamenta en la amplia disponibilidad de este recurso, que se encuentra a gran escala y en todo el mundo y es relativamente barata [Lynd et al., 2002, 2005, Hasunuma et al., 2013, Huang et al., 2014]. Por otra parte, los ciclos de uso de biomasa lignocelulósica como fuente de energía tienen una producción neta de gases de efecto invernadero mucho menor [Borrion et al., 2012] y la conversión de estos materiales a etanol presenta un balance energético muy positivo [Demain et al., 2005, Lynd et al., 2005].

La biomasa lignocelulósica se obtiene como subproducto en agricultura o como residuo industrial y urbano [Bayer et al., 2007]. Además muchas de estas fuentes de biomasa se están estudiando, e incluso en algún caso usando, en la actualidad para la producción de etanol, una de las alternativas más prometedoras a los combustibles basados en el petróleo [Hasunuma et al., 2013].

### 1.1 LIGNOCELULOSA

El uso de lignocelulosa como fuente de energía supone un reto científico y técnico considerable ya que se trata de un compuesto sólido y muy heterogéneo formado por carbohidratos fuertemente interconectados con una fracción de compuestos aromáticos complejos [Chundawat et al., 2011]. En peso seco, la celulosa (20 a 50%), los polisacáridos no celulósicos (15 a 35%) y lignina (10 a 30%) son los componentes mayoritarios y sirven como la base estructural que proporciona a las plantas su rigidez y les permiten resistir la descomposición física y química [Lynd et al., 2002, Bayer et al., 2013]. Además, también se encuentran otros componentes minoritarios como proteínas (3 a 10%), lípidos (1 a 5%), azúcares solubles (1 a 10%) y minerales (5 a 10%).



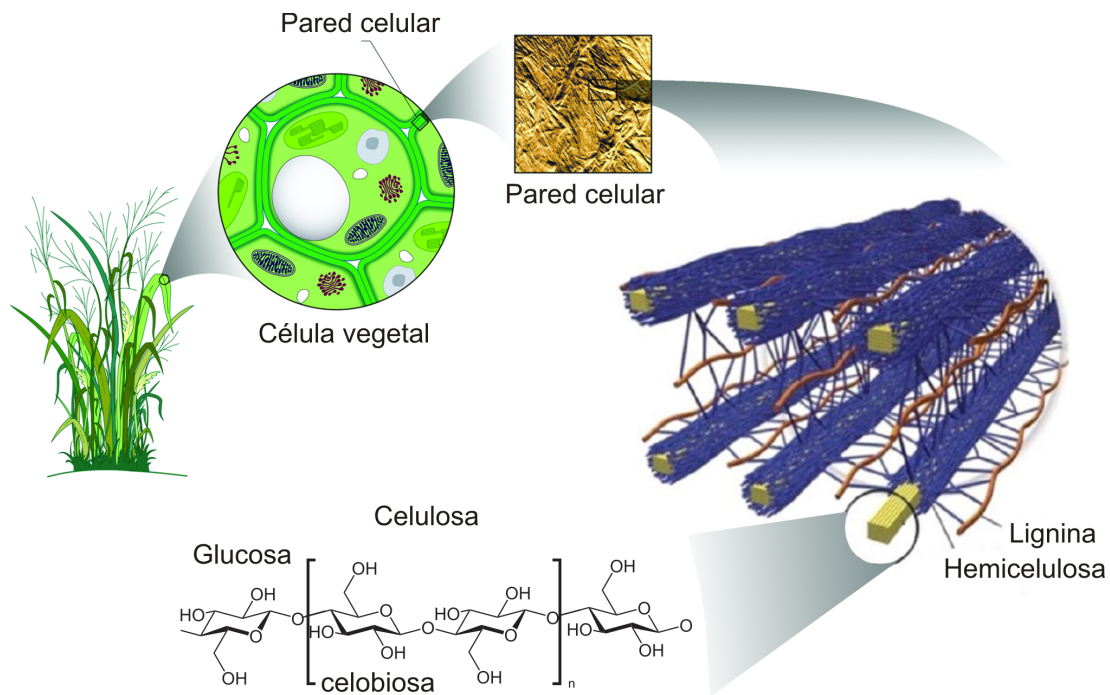


Figura 1: Material lignocelulósico de la pared celular de las plantas. Las cadenas de celulosa se agrupan formando microfibras que contienen regiones cristalinas. Las hemicelulosas interactúan con la superficie de las microfibras de celulosa, mientras que sus cadenas laterales se unen covalentemente a hemicelulosa y lignina formando complejos altamente recalcitrantes. Figura realizada a partir de [King, 2013, Quiroz-Castañeda & Folch-Mallol, 2013].

La composición concreta varía según el tipo de planta, el tipo celular así como el tipo de pared celular y fase de crecimiento [Chundawat et al., 2011, Lynd et al., 2002, Wei et al., 2009].

La celulosa es un polímero lineal formado por unidades de glucosa unidas por un enlace  $\beta$ -1,4-glucosídico (Figura 1). A pesar de que la celulosa se compone exclusivamente de glucosa, la unidad mínima repetida en su estructura es el disacárido celobiosa, ya que cada residuo está rotado  $180^\circ$  respecto al siguiente [Béguin & Aubert, 1994]. Las interacciones por puentes de hidrógeno entre glucosas permiten formar cadenas largas, que suelen contener entre 100 y 10,000 unidades. Estas interacciones son el origen de la rigidez de la estructura y permiten que la celulosa, a diferencia de la mayoría de polisacáridos, pueda formar una estructura cristalina muy poco accesible [Zhang & Lynd, 2004]. Sin embargo, las fibras naturales de celulosa no forman cristales puros y contienen zonas completamente amorfas o con formas intermedias y por tanto presentan un grado de accesibilidad mayor [Lynd et al., 2002].

Envolviendo las fibras de celulosa, se encuentra la hemicelulosa que presenta una estructura amorfa y ramificada. La mayoría de hemicelulosas están formadas por pentosas, siendo la xilosa el azúcar mayoritario (90%) [Bayer et al., 2013]. Su estructura ramificada permite el entrecruzamiento con la lignina y además los azúcares individuales pueden sufrir modificaciones como metilación o acetilación [Bayer et al., 2013].

A diferencia de los anteriores componentes de la lignocelulosa, la lignina no es un polisacárido sino un polímero hidrofóbico de alto peso molecular compuesto por residuos aromáticos fenólicos. Se ha propuesto que las estructuras de lignina se originan por oxidación mediante radicales libres [Chundawat et al., 2011]. La lignificación aporta resistencia mecánica y protección contra patógenos [Wei et al., 2009]. Otra diferencia respecto a los polisacáridos es que la lignina requiere oxígeno para ser degradada, lo que dificulta la utilización de biomasa en ambientes anaerobios [White et al., 2014].

Las hexosas (glucosa) y pentosas (principalmente xilosa) de la lignocelulosa son los azúcares solubles con mayor interés en procesos de fermentación para obtener biocombustibles. Como se ha comentado, estos azúcares forman parte de polisacáridos de la pared celular y para poder ser utilizados es necesario descomponer estos polímeros mediante un proceso conocido como sacarificación. Este proceso es reconocido como el principal cuello de botella para desarrollar procesos rentables y eficientes para poder usar la lignocelulosa como fuente de energía [Demain et al., 2005, Chundawat et al., 2011, Lynd et al., 2002, 2005, Olson et al., 2012, White et al., 2014, Huang et al., 2014]. La dificultad radica en el ataque inicial al sustrato [Bayer et al., 2007, Demain et al., 2005], cuya complejidad y características fisicoquímicas han sido seleccionadas durante la evolución para evitar su degradación y proteger así a la planta [Olson et al., 2012]. La lignocelulosa se comporta como un sólido, lo que limita tanto la accesibilidad como el ataque de enzimas que deben actuar sobre un sustrato insoluble [Himmel et al., 2007], reduciendo la velocidad de hidrólisis varios órdenes de magnitud respecto a reacciones enzimáticas en solución [Chundawat et al., 2011, Lynd et al., 2002].

Para poder desarrollar un proceso industrial capaz de aprovechar la lignocelulosa, es necesario disponer de un método de sacarificación eficiente capaz de superar la barrera impuesta por las características recalcitrantes de la biomasa [Olson et al., 2012]. Para conseguir que la lignocelulosa sea más hidrolizable, en primer lugar se realiza un proceso de pretratado de la biomasa [Lynd et al., 2002]. Este intenta eliminar el sellado de lignina, solubilizar la hemicelulosa, reducir el grado de cristalinidad y aumentar el tamaño de los poros a la vez que se incrementa la superficie accesible [Demain et al., 2005].

## 1.2 SISTEMAS CELULOLÍTICOS

Como se ha comentado, una vez pretratada la biomasa lignocelulósica es necesario realizar una etapa de sacarificación para poder extraer y solubilizar los azúcares que la componen para que estos puedan ser fermentados y generar los productos de interés. En la naturaleza la degradación de estos sustratos la llevan a cabo algunos microorganismos, generalmente en consorcios poco definidos lo que resulta en tiempos muy largos de degradación y procesos poco reproducibles [Bayer et al., 2007]. Se ha observado que para la degradación de lignocelulosa se necesitan enzimas con actividades distintas [Reese et al., 1950] que atacan a los distintos componentes usando distintos

mecanismos de acción. Según la base de datos de enzimas activas sobre carbohidratos [CAZY, 1998], estas enzimas se clasifican en: glucosidasas (catalizan la hidrólisis de enlaces glucosídicos y por tanto actúan sobre la celulosa y la cadena principal de hemicelulosas), liasas de polisacáridos (catalizan la degradación no hidrolítica de enlaces glucosídicos), esterasas (eliminan cadenas laterales y grupos metilo y acetilo de pectinas y xilanos) y actividades asociadas (clase creada recientemente y que engloba enzimas ligninolíticas y otras enzimas capaces de degradar celulosa mediante un mecanismo oxidativo).

Dentro de las glucosidasas se encuentran las celulasas, enzimas capaces de cortar el enlace  $\beta$ -1,4-glucosídico [Lynd et al., 2002, Bayer et al., 2013] y por tanto de degradar celulosa, el polisacárido más recalcitrante del material lignocelulósico. Se ha observado que para una solubilización completa de la celulosa hasta glucosa se requiere la actuación de 3 tipos de celulasas según su modo de acción (Figura 2) [Hasunuma et al., 2013, Lynd et al., 2002]: endocelulasas, que actúan sobre el centro de la fibra, exocelulasas, que cortan desde los extremos y  $\beta$ -glucosidasas, que degradan los oligosacáridos a glucosa.

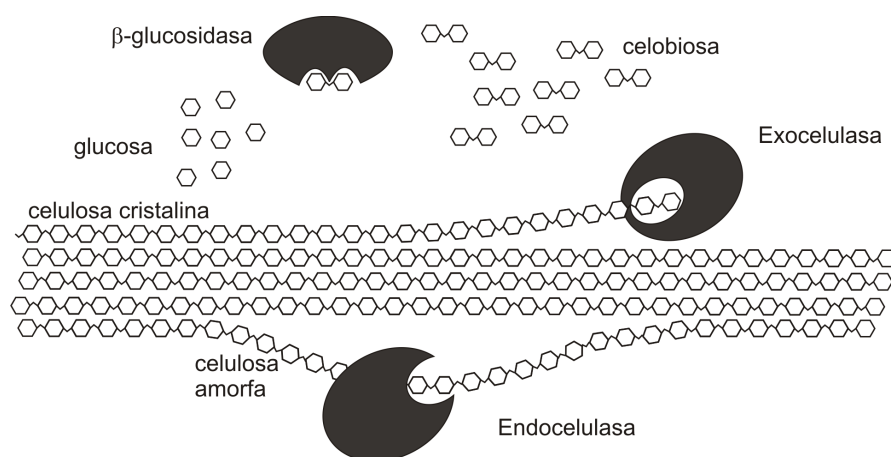


Figura 2: Representación de los tres modos de acción básicos de las celulasas. Las endocelulasas son capaces de cortar las cadenas de celulosa en posiciones centrales, preferentemente en regiones amorfas. Las exocelulasas realizan cortes en la fibra desde los extremos y liberan generalmente celobiosa como producto mayoritario, un disacárido soluble (u otras especies de bajo grado de polimerización). Estas suelen actuar de forma procesiva, degradando progresivamente una misma cadena de celulosa. Las  $\beta$ -glucosidasas son capaces de cortar oligosacáridos solubles (como la celobiosa u otras especies solubles como la celotriosa, celotetraosa, etc.) para producir glucosa libre.

Muchas de las enzimas activas sobre carbohidratos son multimodulares [Tilbeurgh et al., 1986, Tomme et al., 1988, Gilkes et al., 1991], es decir, contienen en un mismo polipéptido varios módulos que pueden presentar distintas funciones: otros módulos catalíticos, módulos de unión u otros con funciones desconocidas [Czjzek et al., 2012, White et al., 2014]. Uno de los módulos más comúnmente encontrados junto con unidades catalíticas es el módulo de unión a carbohidratos (en inglés Carbohydrate Binding Module) (CBM), que en algunas ocasiones incluso se encuentran en tándem

con varios de ellos [Boraston et al., 2004]. Estos módulos suelen ser muy importantes para la actividad enzimática ya que permiten acercar la unidad catalítica a su sustrato específico (se ha descrito un amplio abanico de especificidades en CBM).

Entre los microorganismos con capacidad de degradación de lignocelulosa los hongos han sido los más estudiados. Estos son capaces de producir las distintas enzimas en grandes cantidades lo que facilita su obtención y caracterización. En conjunto, los microorganismos aerobios son los responsables de la degradación de un 90 % de la celulosa [Wei et al., 2009]. Por otro lado, en ambientes anaerobios, se ha observado que muchas bacterias producen enzimas que se ensamblan en grandes complejos, conocidos como celulosomas [Bayer et al., 2008, 2013, Himmel et al., 2010] y que constituyen el objeto de trabajo de esta tesis.

### 1.2.1 Celulosoma

El celulosoma es un complejo proteico producido por algunas bacterias anaerobias para degradar lignocelulosa (Figura 3). Fue descubierto en *Clostridium thermocellum*, donde se identificó la capacidad celulolítica con un complejo de muy alto peso molecular (2.1 MDa) [Lamed et al., 1985]. Ya en los primeros estudios se observó que bajo ciertas condiciones [Lamed et al., 1983] era posible descomponer este complejo en varias unidades que se clasificaron inicialmente según su peso molecular. La mayoría de componentes se identificaron como celulasas a las que se atribuyó la actividad del complejo, sin embargo sorprendentemente se observó que entre los componentes más abundantes había una proteína de alto peso molecular (unos 210 kDa) no catalítica pero capaz de unir las distintas enzimas [Lamed et al., 1983]. Este componente se conoce actualmente como escafoldina [Bayer et al., 1994], es el elemento común a todos los celulosomas y la que distingue este complejo de otros sistemas de degradación de celulosa.

En el celulosoma la escafoldina presenta varias funciones según sus componentes. En el caso de *C. thermocellum* la escafoldina CipA es la encargada de unir las enzimas, dirigir las al sustrato y unir el complejo a la superficie de la bacteria. La escafoldina actúa como andamio donde se unen hasta nueve enzimas. Esto es posible gracias a la presencia de nueve repeticiones [Gerngross et al., 1993] de un módulo llamado cohesina cuya función es la unión de dominios doquerina presentes en las distintas enzimas que se van a integrar en el complejo. En el extremo C-terminal de CipA se encuentra un dominio doquerina que, a diferencia de las doquerinas en las enzimas, no es capaz de unirse a las cohesinas de CipA sino a otras presentes en otras proteínas. Tanto la doquerina de CipA como las cohesinas a las que se une son ligeramente distintas a nivel estructural y se clasifican como tipo II en contraposición a las cohesinas de CipA y doquerinas de las enzimas que se asignan al tipo I. Como se puede ver, la interacción cohesina-doquerina es fundamental en el celulosoma ya que determina la arquitectura global de este. Finalmente, la unión a sustrato la realiza un módulo CBM de familia 3 [Morag et al., 1995, Bayer et al., 1998], que se une preferentemente a celulosa cristalina.

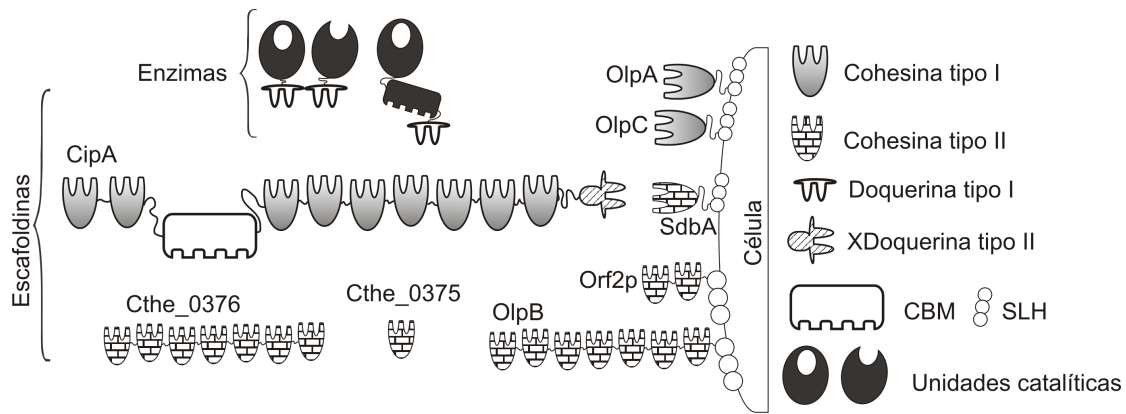


Figura 3: Representación del celulosoma de *C. thermocellum*. El celulosoma de *C. thermocellum* está formado por 8 escafoldinas y más de 70 unidades catalíticas con doquerina. La escafoldina CipA es la principal encargada de unir las distintas enzimas que forman parte del complejo. Esta escafoldina contiene un CBM, 9 cohesinas de tipo I, capaces de unir el dominio doquerina presente en enzimas, y en el extremo C-terminal un módulo X asociado a una doquerina de tipo II, capaz de unirse a las cohesinas del mismo tipo de escafoldinas secundarias ancladas a la pared del microorganismo (SdbA, Orf2p, y OlpB) o libres (Cthe\_0375 y Cthe\_0376). *C. thermocellum* también presenta en su superficie otras dos escafoldinas que contienen una cohesina de tipo I, que podrían estar implicadas en el ensamblado del celulosoma. El anclaje de las escafoldinas secundarias a la pared celular se realiza a través de dominios SLH.

#### 1.2.1.1 Doquerinas

Los dominios doquerina de tipo I son relativamente pequeños y están formados por un segmento duplicado de unos 22 aminoácidos [Bayer et al., 1998] separados por una secuencia de entre 9 y 16 residuos. Cada una de las repeticiones forma 2 hélices  $\alpha$  separadas por un bucle conservado de unos 12 residuos capaz de unir un ion  $Ca^{2+}$  [Lytle et al., 2001]. Esta estructura es una variante de un motivo conocido como mano-EF [Lytle et al., 2000]. Esto resulta en que cada dominio doquerina une 2 iones  $Ca^{2+}$  (Figura 4.C) que son necesarios para su plegamiento [Lytle et al., 2000] y el motivo por el cual la interacción cohesina-doquerina es dependiente de este ion [Johnson & Demain, 1984, Yaron et al., 1995]. La estructura generada presenta simetría cosa que permite que, en algunos casos (en algunas doquerinas de tipo I), estos dominios se puedan unir de dos formas distintas, rotadas  $180^\circ$  sobre la misma superficie de la cohesina [Carvalho et al., 2007], conocido como modo dual de unión (Figura 4.E). Este fenómeno, que se cree que podría aportar flexibilidad durante el ensamblado del celulosoma facilitando la inserción de varias enzimas voluminosas en posiciones cercanas, se ha observado en estudios cristalográficos [Carvalho et al., 2007, Pinheiro et al., 2008, Currie et al., 2012] aunque aún no está completamente claro si ocurre en el complejo natural.

Se han identificado tres tipos de doquerina a partir de su secuencia y estructura [Smith & Bayer, 2013]. En muchos casos se ha observado que los dominios doquerina (de tipo II y III) van acompañados de un módulo hidrofílico llamado X (Figura 4.B)

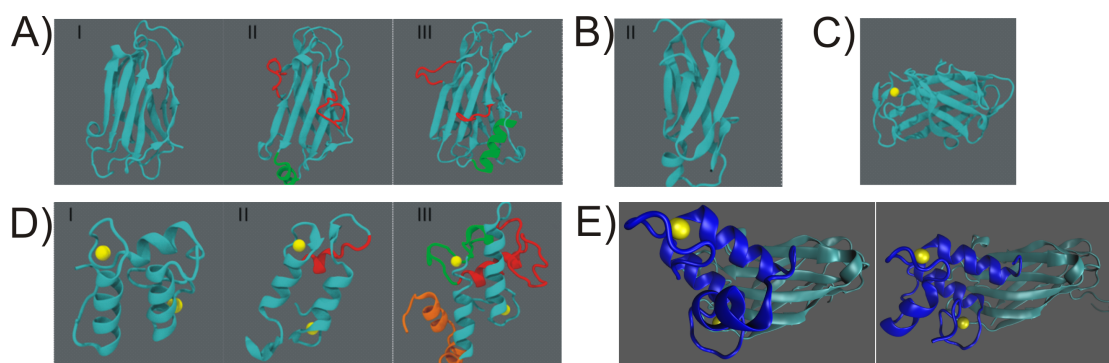


Figura 4: Estructura de distintos componentes del celulosoma. A) Cohesinas de tipo I, II y III. En rojo y verde se resaltan las principales diferencias estructurales entre estos módulos. B) Módulo hidrofílico (módulo X). C) Estructura de un CBM de la familia III. D) Dominios doquerina de tipo I, II y III, con las principales diferencias entre ellos resaltadas. E) Ejemplo del modo dual de unión de algunas doquerinas, en el que una misma doquerina se puede unir de dos formas distintas (rotada 180 °) a una misma cohesina.

[Bayer et al., 1998], y en ausencia de estos módulos las doquerinas correspondientes suelen ser muy poco solubles [Adams et al., 2005]. La estructura cristalográfica de estas doquerinas junto al módulo X ha mostrado que este forma interacciones que estabilizan y rigidifican la doquerina [Xu & Smith, 2010], lo que podría disminuir el coste entrópico de la interacción, y además en algunos casos establecen un contacto directo con la cohesina correspondiente [Adams et al., 2006] resultando en conjunto en constantes de afinidad mayores.

#### 1.2.1.2 Cohesinas

Las cohesinas son módulos de unos 150 residuos que se caracterizan por su estructura tipo  $\beta$ -sándwich con topología *jelly-roll* formada por 9 hebras  $\beta$  organizadas en 2 láminas [Tavares et al., 1997, Shimon et al., 1997]. Todas las hebras son antiparalelas a excepción de la primera y la última (1 y 9) que son paralelas. Se han descrito 3 tipos distintos de cohesinas que presentan ligeras diferencias a nivel estructural. Aunque todas adoptan el mismo plegamiento global [Noach et al., 2005, Adams et al., 2006, Alber et al., 2009, Voronov-Goldman et al., 2014], en las de tipo II y III hay dos hebras interrumpidas por *beta-flaps* y se observa la presencia de una hélice  $\alpha$ .

La interacción con doquerina se da a través de la cara formada por las hebras 8-3-6-5 [Carvalho et al., 2003]. La interacción de dominios doquerina con módulos cohesina suele presentar una afinidad y especificidad extraordinariamente elevadas. Las constantes de afinidad para distintos pares se ha determinado en el rango de  $10^8$ - $10^{11} M^{-1}$  [Slutzki et al., 2012], similar a la de interacciones antígeno-anticuerpo. Por otro lado la interacción cohesina-doquerina es específica por tipo [Bayer et al., 2004], a menudo por especie [Pagés et al., 1997, Haimovitz et al., 2008] pero sin embargo parece que no hay una selección fuerte de la posición dentro de la escafoldina [Yaron et al., 1995].

### 1.2.1.3 Escafoldinas

Las escafoldinas son proteínas que contienen módulos cohesina [Bayer et al., 1994]. A menudo, las escafoldinas presentan varias cohesinas además de otros módulos con distintas funciones. *C. thermocellum* presenta 8 genes correspondientes a escafoldinas, organizado en un clúster. Además de CipA ya descrita, *C. thermocellum* puede producir 5 escafoldinas que en su extremo C-terminal contienen tres dominios SLH capaces de interactuar con los polisacáridos de la matriz extracelular de la célula [Kern et al., 2011] lo que dirige a estas escafoldinas a la superficie bacteriana. Tres de estas escafoldinas (SdbA, ORF2p y OlpB) presentan 1, 2 y 7 cohesinas de tipo II, capaces de unir la doquerina de tipo II presente en la escafoldina CipA, mientras que las otras (OlpA y OlpC) presentan una única cohesina de tipo I, capaz de reconocer doquerinas de enzimas (Figura 3). Esto dicta un orden jerárquico de interacciones que determina la arquitectura global del celulosoma. Los módulos cohesina de una escafoldina están separados entre ellos por secuencias intermodulares de longitud variable y generalmente ricas en prolina y treonina que se encuentra glicosilada [Gerwig et al., 1989, 1991, 1993], aunque el papel de estas modificaciones aún no está completamente claro [Currie et al., 2013]. Al disponer de varias repeticiones capaces de unir doquerinas, ya sea de enzimas o de otras escafoldinas, las escafoldinas se comportan como andamios [Good et al., 2011] capaces de coordinar las distintas moléculas con las que interactúan, permitiendo a la vez aumentar la concentración efectiva de enzimas cerca de la región de interés donde las dirige o ancla. Se ha propuesto que este tipo de estrategia permite al sistema evolucionar fácilmente ya que al reordenar sus componentes o añadir otros permite la aparición de nuevas funciones [Good et al., 2011].

### 1.2.1.4 El celulosoma de *C. thermocellum*

*C. thermocellum* presenta actividad celulásica real, es decir, es capaz de degradar celulosa microcristalina [Demain et al., 2005]. Esta actividad se asocia principalmente a la de las enzimas que incorpora en su celulosoma [Lamed et al., 1983] que incluye celulasas, hemicelulasas, esterases, liasas y proteasas [Fontes & Gilbert, 2010].

En la superficie de *C. thermocellum*, los celulosomas se agrupan formando unas protuberancias conocidas como policelulosomas [Bayer & Lamed, 1986]. En este estado los celulosomas parecen estar ordenados de forma que son menos accesibles a proteasas [Morag et al., 1996, Mayer et al., 1987]. Al entrar en contacto con celulosa cristalina, el CBM de CipA se une a esta y los celulosomas adoptan una conformación mucho más extendida y accesible apareciendo como fibras que unen las células al sustrato [Bayer & Lamed, 1986, Mayer et al., 1987]. En las primeras fases de degradación de celulosa, *C. thermocellum* se une fuertemente al sustrato, sin embargo, al final de la fase exponencial de crecimiento, cuando disminuye la velocidad de obtención de azúcares solubles, las células se liberan de los celulosomas y estos se mantienen unidos al sustrato [Demain et al., 2005].



### 1.2.1.5 Diversidad de celulosomas

Hasta ahora, la descripción del celulosoma se ha centrado en el caso de *C. thermocellum*, sin embargo se han identificado varias especies tanto mesófilas como termófilas capaces de formar celulosomas como es el caso de *Clostridium cellulolyticum*, *Acetivibrio cellulolyticus*, *Bacteroides cellulosolvens*, *Clostridium acetobutylicum*, *Clostridium cellulovorans*, *Clostridium clariflavum*, *Clostridium josui*, *Clostridium papyrosolvens*, *Ruminococcus albus* y *Ruminococcus flavefaciens* [Smith & Bayer, 2013].

En todos estos microorganismos el celulosoma está formado por los mismos componentes elementales: módulos cohesina en escafoldinas capaces de interactuar con dominios doquerina de enzimas u otras escafoldinas. Sin embargo la composición de las escafoldinas, que determina la arquitectura general del complejo es muy variable (Figura 5). Algunos microorganismos como *C. cellulolyticum* (Figura 5.A), *C. cellulovorans*, *C. josui*, o *C. acetobutylicum* forman celulosomas relativamente simples, formados por una sola escafoldina que no se une a la célula. Por otro lado, *C. thermocellum* (Figura 3), *A. cellulolyticus*, *B. cellulosolvens* o *R. flavefaciens* (Figura 5.B, C y D), por ejemplo, producen celulosomas mucho más complejos que involucran varias escafoldinas intermedias lo que podría permitir que hasta 160 enzimas se unan por cada celulosoma anclado a superficie en el caso de *C. clariflavum* [Artzi et al., 2014]. En estos casos las cohesinas suelen estar separadas por secuencias intermodulares más largas y el CBM suele encontrarse en una posición centrada en solo una de las escafoldinas [Bayer et al., 2009]. Para ensamblar estos celulosomas de forma controlada se suelen encontrar cohesinas de distinto tipo en cada escafoldina, por ejemplo, *A. cellulolyticus* produce una escafoldina, ScaA, que contiene módulos cohesina de tipo I y una doquerina de tipo II capaz de reconocer cohesinas del mismo tipo en ScaB. Esta a su vez lleva una doquerina de tipo I con especificidad distinta a la de las enzimas que se puede unir a ScaC que se ancla a la superficie celular mediante tres repeticiones SLH [Xu et al., 2003, Dassa et al., 2012] (Figura 5.B).

El análisis del genoma de varios de estos microorganismos ha permitido observar que existen escafoldinas que incorporan unidades catalíticas en su secuencia [Ding et al., 1999], que presentan cohesinas de varios tipos en una misma escafoldina [Xu et al., 2004] o la posibilidad de que la escafoldina que une las enzimas esté compuesta de cohesinas de tipo II [Ding et al., 2000].

Por otro lado, se ha observado que algunos hongos son capaces de producir sistemas celulolíticos que forman complejos de alto peso molecular [Wilson & Wood, 1992, Gilmore et al., 2015]. Sin embargo, lo que originalmente se identificó como una posible doquerina en sus enzimas, no tiene ningún tipo de relación estructural con las doquerinas que forman parte del celulosoma por lo que parece que se trata de una vía completamente distinta [Peer et al., 2009].



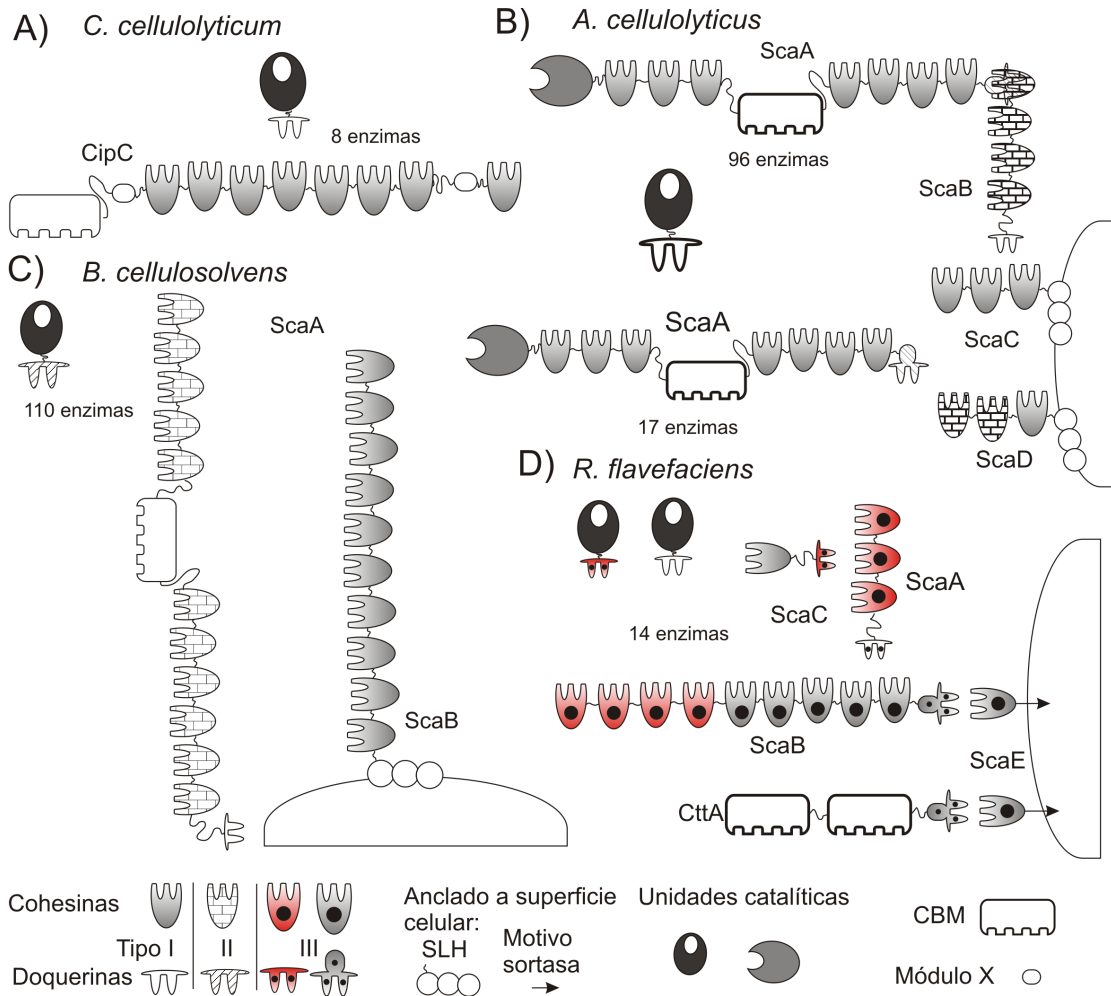


Figura 5: Diversidad de celulosomas. Empleando los mismos bloques básicos, los distintos microorganismos productores de celulosomas son capaces de ensamblar complejos con arquitecturas muy distintas. A) *C. cellulolyticum* produce un celulosoma simple con una única escafoldina libre capaz de ensamblar doquerinas de tipo I presentes en las enzimas. B) *A. cellulolyticus* produce un celulosoma complejo en el que intervienen varias escafoldinas. Sus enzimas llevan doquerinas de tipo I, una de sus escafoldinas contiene una unidad catalítica y presenta una escafoldina que combina cohesinas de varios tipos. C) *B. cellulosolvans* genera un celulosoma donde la escafoldina primaria, ScaA, presenta cohesinas de tipo II, capaces de interactuar con las doquerinas de tipo II presentes en algunas de sus enzimas. D) El celulosoma de *R. flavefaciens* se ancla covalentemente a su superficie mediante el motivo sortasa presente en una de sus escafoldinas. La mayoría de cohesinas en este celulosoma son de tipo III, aunque se encuentran con varias especificidades. Las enzimas de este microorganismo llevan cohesinas de tipo I o tipo III. Los celulosomas mostrados en esta figura representan únicamente parte de la diversidad descubierta hasta el momento, ya que se han descrito multitud de escafoldinas adicionales en algunos de estos microorganismos [Hamberg et al., 2014, Dassa et al., 2014].

### 1.2.2 Sinergia

La capacidad de producir proteínas de los microorganismos que producen celulosomas es muy inferior a la de los hongos o bacterias aerobias [Lynd et al., 2002]. Sin embargo, los complejos enzimáticos de estos microorganismos suelen presentar una actividad específica muy elevada, llegando en el caso de *C. thermocellum* a ser 50 veces mayor comparadas con el sistema de *Trichoderma reesei* [Johnson & Demain, 1984]. Esto se debe tanto a la elevada actividad de las enzimas como a su incorporación en la escafoldina, que puede mejorar su actividad hasta cinco veces [Fierobe et al., 2005]. El aumento de la actividad del conjunto, comparada con la de los componentes por separado se conoce como sinergia y se puede cuantificar como el cociente entre la actividad del conjunto y la suma de actividades de los componentes por separado [Zhang & Lynd, 2004].

En los sistemas de degradación de celulosa se han descrito varias fuentes de sinergia [Lynd et al., 2002, Zhang & Lynd, 2004]: la combinación de endocelulasa y exocelulasa (la primera ataca el centro de una cadena de celulosa generando más sitios donde la segunda puede actuar) que suele ser la mayor fuente de sinergia en estos sistemas, entre exocelulasas (entre enzimas que actúan en el extremo reductor y no reductor), entre exocelulasa y  $\beta$ -glucosidasa (la eliminación de oligosacáridos por parte de una  $\beta$ -glucosidasa evita la inhibición de las exocelulasas) o sinergia intramolecular por el CBM de celulasas (esta dirige y ancla la enzima a su sustrato).

En el caso específico del celulosoma, además se han descrito dos tipos importantes de sinergia derivadas de la incorporación de las enzimas a una escafoldina lo que resulta en sinergia adicional:

- Proximidad espacial de distintas enzimas complementarias.
- Orientación y anclaje al sustrato mediante el CBM de la escafoldina. Además muchas de las enzimas incorporadas llevan su propio CBM que suele estar más especializado para unir el sustrato de la propia enzima.

Para la degradación de sustratos naturales, la presencia de celulasas junto con hemicelulasas y otras enzimas suele ser otra fuente de sinergia [Moraís et al., 2010, Gefen et al., 2012, Arfi et al., 2014]. Esta puede jugar un papel clave en la degradación de estos materiales lignocelulósicos como se refleja en el caso de *C. thermocellum* que produce hemicelulasas celulosomales. Estas liberan pentosas que no son metabolizadas por *C. thermocellum* [Demain et al., 2005], pero le facilitan el acceso a la celulosa. Otro factor que juega un papel importante es la formación del complejo celulosa-enzimas-microorganismo (CEM) [Lu et al., 2006]. Muchos de los microorganismos anaerobios que degradan celulosa suelen adherirse a ella, de hecho, la formación de estos complejos suele resultar en mayores actividades. Este fenómeno se conoce como sinergia CEM y puede tener su origen en el hecho de que asegura que la célula que lo produce es la primera en acceder al producto de degradación. Además puede proteger la célula de ataques y a las enzimas de las proteasas [Lynd et al., 2002].

### 1.2.3 *Mini-celulosomas*

La interacción cohesina-doquerina es la base de la diversidad estructural del celulosoma. La elevada afinidad y especificidad de esta interacción se ha empleado para el diseño de celulosomas quiméricos, donde la escafoldina está compuesta de cohesinas de distintos organismos, de forma que se pueda controlar la incorporación de las distintas enzimas para generar un celulosoma bien definido [Fierobe et al., 2001, 2002, 2005]. Estos son conocidos como mini-celulosomas, o celulosomas de diseño, y permiten estudiar de forma mucho más controlada la acción de los distintos elementos del celulosoma. El uso de mini-celulosomas permite desacoplar el efecto de la proximidad entre las enzimas y la unión al sustrato para estudiar el origen de la sinergia entre las enzimas al unirse a la escafoldina [Fierobe et al., 2001, 2002]. A medida que se dispone de más secuencias de cohesinas y doquerinas, esta aproximación permite construir mini-celulosomas más grandes que permiten la incorporación controlada de más enzimas [Fierobe et al., 2005, Morais et al., 2010, 2012]. Empleando esta estrategia también se han ensamblado celulosomas con geometrías no naturales permitiendo concluir que la movilidad del celulosoma es esencial para una buena actividad [Mingardon et al., 2007b]. También ha permitido estudiar al detalle el efecto sobre la actividad de factores como la longitud de las secuencias intermodulares que enlazan las distintas cohesinas y el orden entre las enzimas y el CBM [Vazana et al., 2013, Stern et al., 2015].

Por otro lado, la incorporación de doquerinas a enzimas permite convertir enzimas libres o de distinta especificidad en enzimas capaces de unirse a un celulosoma. Esto permite la incorporación de enzimas de otros microorganismos al celulosoma lo que a menudo resulta en un aumento de su actividad [Mingardon et al., 2007a, Caspi et al., 2008, 2009, 2010]. Esta aproximación resulta interesante desde un punto de vista aplicado ya que podría permitir mejorar la actividad de cócteles enzimáticos ya definidos.

## 1.3 BIOPROCESOS CONSOLIDADOS

El uso de biomasa lignocelulósica en una biorefinería para producir combustibles u otros productos de interés, es actualmente muy limitado debido a la baja eficiencia de la sacarificación y al elevado coste de la producción de las enzimas necesarias. Para llevar a cabo este proceso son necesarios 4 pasos: producción las enzimas, sacarificación del sustrato, fermentación de hexosas y fermentación de pentosas [Lynd et al., 2005]. Para conseguir que el proceso completo pueda ser rentable se ha propuesto la integración de las 4 fases en un solo proceso, conocido como bioproceso consolidado (en inglés Consolidated BioProcessing) (CBP) [Lynd et al., 2005], lo que podría reducir los costes al emplear cepas capaces de producir sus propias enzimas [Lynd et al., 2005, Hasunuma et al., 2013] para sacarificar el sustrato y luego fermentar los azúcares [Olson et al., 2012]. Esto permitiría limitar la inhibición de las enzimas por los productos de degradación [Hasunuma et al., 2013], aprovechar la sinergia CEM [Lynd et al., 2005]

e incluso reciclar el catalizador si este se encuentra unido al microorganismo [Huang et al., 2014].

El organismo ideal para CBP es aquel capaz de producir un sistema sacarolítico eficiente y de fermentar hexosas y pentosas para producir cantidades elevadas de producto en tiempos cortos [Demain et al., 2005, Lynd et al., 2005, Olson et al., 2012]. Sin embargo, no se conocen organismos que reúnan todas estas características. Para superar este problema se han propuesto varias vías como el uso de cocultivos empleando organismos con capacidades complementarias [Demain et al., 2005] o la modificación de organismos conocidos para dotarlos de todas las características necesarias. Se han propuesto 3 estrategias para la modificación de microorganismos para CBP: adaptar el proceso de fermentación en organismos sacarolíticos (lo que permitiría usar microorganismos productores de celulosoma como *C. thermocellum* que es anaerobio y termófilo, que son características ventajosas para procesos industriales [Demain et al., 2005, Olson et al., 2012]), introducir la capacidad de sacarificación en buenos fermentadores (esto requiere sistemas sacarolíticos muy efectivos [Hyeon et al., 2011, Lynd et al., 2005, Hasunuma et al., 2013]) o alternativamente incorporar la capacidad sacarolítica y rutas de fermentación adecuadas en microorganismos capaces de usar múltiples fuentes de carbono y que toleren compuestos tóxicos. Un ejemplo para este último sería el uso de *E. coli*, que es capaz de metabolizar hexosas y pentosas [Muñoz-Gutiérrez et al., 2012] y para la que ya se han desarrollado cepas para la producción de biocombustibles [Huffer et al., 2012, Hasunuma et al., 2013].

Una opción interesante es la incorporación en estos microorganismos de sistemas celulóliticos anclados en su superficie [Tanaka & Kondo, 2015, Huang et al., 2014, Li et al., 2004]. Esto presenta como ventaja que estos microorganismos tendrán mayor facilidad para acceder al sustrato, lo que limita la posibilidad de contaminación del cultivo [Lynd et al., 2005] y permite aprovechar la sinergia CEM. Se han desarrollado varios sistemas para conseguir anclar proteínas celulóliticas a la superficie celular en *E. coli*, *Saccharomyces cerevisiae*, *Corynebacterium glutamicum* y algunas bacterias del ácido láctico [Huang et al., 2014].

En *E. coli* se han desarrollado varias construcciones que permiten presentar proteínas en su superficie [Huang et al., 2014, Karami et al., 2014, Park et al., 2014]. Una de las más usadas es una construcción híbrida conocida como Lpp-OmpA [Francisco et al., 1992] que fusiona un fragmento de Lpp a otro de OmpA, ambas proteínas de membrana de *E. coli* (Figura 6). Esta estrategia se ha empleado para presentar algunas enzimas celulóliticas o CBM [Francisco et al., 1993, Qu et al., 2015], aunque no se ha descrito la incorporación de celulosomas en su superficie [Huang et al., 2014].

#### 1.4 MECÁNICA EN BIOLOGÍA

La biología moderna compara las células con pequeñas factorías: lugares repletos de máquinas moleculares muy especializadas y organizadas para llevar a cabo todos los procesos celulares de los que depende la vida. Muchas de las funciones que desem-

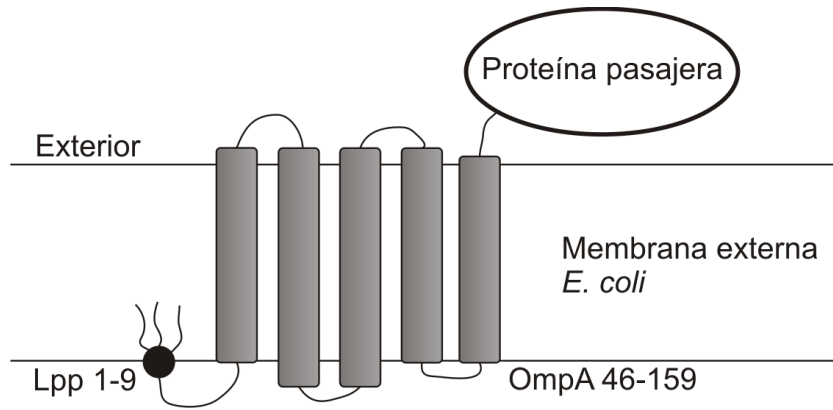


Figura 6: Esquema de la construcción Lpp-OmpA para mostrar proteínas en la superficie de *E. coli*. El péptido señal de Lpp permite dirigir la proteína al espacio periplásmico donde este se corta dejando una cisteína en el extremo N-terminal que se une a la cara interna de la membrana externa de la célula [Ghrayeb & Inouye, 1984]. Las cinco regiones transmembrana de OmpA se insertan en la membrana externa dejando expuesto el extremo C-terminal, donde se encuentra la proteína pasajera de interés, en la superficie celular [Francisco et al., 1992].

peñan estas nanomáquinas requieren de la conversión de energía de un tipo a otro. A menudo, la energía química contenida en ATP se convierte en trabajo mecánico promoviendo los cambios conformacionales responsables de muchas de las actividades de las proteínas [Bustamante et al., 2004]. Muchos de los procesos básicos de una célula, como la replicación, transcripción y traducción pero también el transporte de orgánulos, la adhesión y movilidad celulares, por citar solo algunos, son en el fondo procesos mecánicos, lo que hace necesario el uso de conceptos físicos como fuerza y trabajo para describirlos [Bustamante et al., 2004].

La energía involucrada en estos procesos es de magnitud similar a la energía térmica, que viene dada por la expresión  $K_B T$  donde  $K_B$  es la constante de Boltzmann, y  $T$  la temperatura absoluta. A temperatura ambiente,  $K_B T = 4.1$  pN·nm, y dado que las proteínas tienen tamaños del orden de nanómetros (nm), es de esperar que las fuerzas que actúan sobre las proteínas sean del orden de los picoNewton ( $10^{-12}$ N, pN). 1 pN corresponde aproximadamente con el peso de un eritrocito, la fuerza ejercida por un puntero láser sobre una pantalla o la fuerza para arrastrar una vesícula de 30 nm moviéndose a 500 nm/s a través del citosol [Howard, 2001].

La nanomecánica de proteínas se encarga del estudio de la respuesta de las proteínas al aplicarles una fuerza mecánica. Una cuestión fundamental es la relevancia que puede tener este tipo de información. Por un lado la fuerza se puede usar como método para perturbar el sistema y así obtener información de interés. La mayoría de los experimentos bioquímicos se han realizado usando otros agentes físicos o químicos, por lo que los resultados obtenidos aplicando fuerza pueden aportar un punto de vista adicional con la ventaja de utilizar una coordenada de reacción bien definida: la extensión de la molécula. Por otro lado, muchas proteínas generan o están sometidas a fuerza durante su funcionamiento normal [Howard, 2001, Bustamante et al., 2004]. Por lo tanto, el estudio nanomecánico de proteínas no solo es un método para obtener

información de estos sistemas sino que es un enfoque muy importante, desde el punto de vista fisiológico, que permite estudiar este parámetro, que de otro modo sería inaccesible [Galera-Prat et al., 2011].

Para las proteínas sometidas a estrés mecánico, su capacidad de resistir fuerza sin desplegarse, es decir, su estabilidad mecánica, es muy importante [Bustamante et al., 2004]. Las proteínas cuya función es resistir fuerza suelen presentar varios módulos en tándem unidos por segmentos desestructurados [Bustamante et al., 2004]. Estos elementos suelen presentar estabilidades mecánicas distintas de manera que los que presenten menor estabilidad mecánica tienen mayor probabilidad de desplegarse antes que los de mayor estabilidad. Uno de los casos más estudiados es la proteína titina, responsable de la elasticidad pasiva del sarcómero del músculo donde actúa como muelle y asegura que esta estructura recupere sus dimensiones originales al relajarse [Linke et al., 1999].

Sin embargo, el interés del estudio del comportamiento mecánico de proteínas no solo se limita a aquellas cuya función es resistir fuerza. Estudios recientes indican que el desplegamiento de proteínas dentro de la célula (para su traslocación o degradación por maquinaria especializada por ejemplo) es un proceso mecánico [Aubin-Tam et al., 2011, Maillard et al., 2011]. Esto implica que las propiedades mecánicas, no solo de las proteínas sometidas a fuerza, sino de la mayoría de las proteínas, son importantes durante su ciclo de vida.

## 1.5 ESPECTROSCOPIA DE FUERZA DE MOLÉCULAS INDIVIDUALES

Dada la importancia de la fuerza mecánica en biología, es necesario disponer de técnicas que permitan estudiar de forma cuantitativa el comportamiento mecánico de las proteínas. Hay que tener en cuenta que la fuerza es una magnitud vectorial y, además, la respuesta de las proteínas al someterlas a fuerza es altamente anisotrópica. Por tanto, no solo es importante conocer su magnitud sino también su dirección y punto de aplicación [Carrion-Vazquez et al., 2003]. Esto supone una enorme dificultad técnica ya que implica controlar con precisión el punto y dirección de aplicación de fuerza sobre cada proteína presente en la muestra, pues de lo contrario se estarían mezclando procesos distintos.

Las técnicas de molécula individual permiten estudiar el comportamiento de moléculas una a una. Estas técnicas permiten obtener de forma directa la distribución de eventos observables y evitan la pérdida de información debido al promediado de los datos. Otra ventaja es que permiten estudiar los procesos en tiempo real, por lo que se puede obtener información dinámica de la molécula y además permiten capturar eventos raros y resolver heterogeneidades que, de otro modo, quedarían ocultos en el promediado de una población [Basché et al., 2001].

Existen varias técnicas capaces de manipular proteínas individuales aplicándoles fuerza de forma controlada y que permiten medir la respuesta mecánica de estas siendo las más usadas el microscopio de fuerza atómica (en inglés Atomic Force

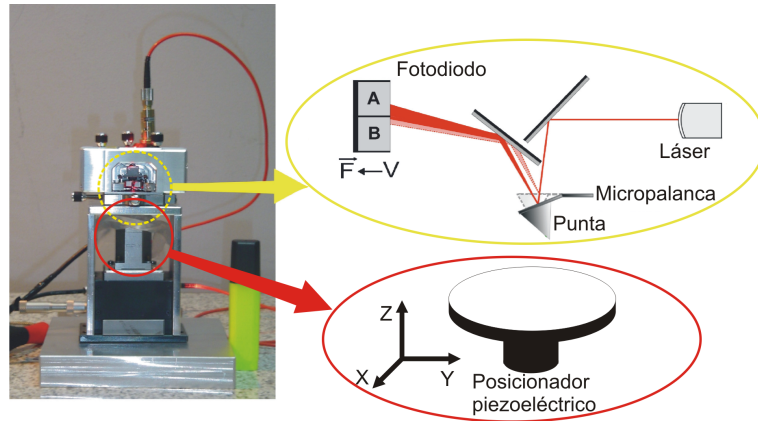


Figura 7: Elementos principales del microscopio de fuerza atómica (AFM). El AFM (izquierda) está formado por una cabeza (en amarillo) donde se encuentra la micropalanca que es el elemento que interactúa con la muestra a través de una pequeña punta en su extremo. El desplazamiento de la punta, que se relaciona con la fuerza a través de su constante de muelle, se puede medir siguiendo el reflejo de un haz láser empleando un fotodiodo partido en dos regiones. Por otro lado, un posicionador piezoeléctrico (en rojo) permite el movimiento controlado y muy preciso en los tres ejes espaciales. Figura adaptada de [Carrión-Vázquez et al., 2006].

Microscope) (AFM), las pinzas ópticas y las pinzas magnéticas [Ritort, 2006, Neuman & Nagy, 2008]. El AFM es el que presenta la mejor resolución espacial de los tres y es capaz de medir en el rango alto de fuerzas, por lo que es bastante adecuado para el estudio nanomecánico de proteínas.

El AFM fue desarrollado en 1986 por Binnig, Quate y Gerber [Binnig et al., 1986] como una variante del microscopio de efecto túnel. El AFM consta de dos elementos principales: un posicionador piezoeléctrico, que permite el movimiento muy preciso y controlado de la muestra y una punta muy fina, montada sobre una micropalanca sensora, que se usa para escanear la superficie (Figura 7). Para realizar una imagen topográfica, se puede registrar el cambio en la altura de la punta a medida que se recorre la muestra.

Además de ser una técnica de imagen de alta resolución, el AFM se puede usar como herramienta de alta precisión para estudiar las propiedades mecánicas de distintas biomoléculas en el régimen alto de fuerzas [Burnham & Colton, 1989]. En este tipo de experimentos se mide el comportamiento de una molécula individual al someterla a fuerza, modo que se conoce como espectroscopía de fuerza de moléculas individuales (en inglés Single Molecule Force Spectroscopy) (SMFS).

En un experimento de SMFS mediante AFM, a partir de ahora abreviado como SMFS, la muestra, que habitualmente es una proteína purificada, se ancla a una superficie (sustrato). Esta se acerca a la punta de la micropalanca permitiendo así que se forme un circuito mecánico entre los dos elementos a través de la proteína a estudiar. A continuación, empleando el posicionador piezoeléctrico se aleja el sustrato de la punta de forma que se estira la proteína mientras se mide la respuesta de esta. Dependiendo del

protocolo de actuación del posicionador, se pueden distinguir los modos de velocidad constante (sección 1.5.1) o fuerza constante.

### 1.5.1 *Modo de velocidad constante*

Como su propio nombre indica, en el modo de velocidad constante el posicionador piezoeléctrico se desplaza a una velocidad determinada (normalmente en el orden de 1 nm/ms) y se registra la fuerza medida respecto a la extensión de la molécula (Figura 8).<sup>1</sup> Una vez formado el circuito mecánico y estirando a velocidad constante, a medida que la punta y el sustrato se alejan, la elasticidad entrópica del polímero se opone a su extensión [Evans & Ritchie, 1999] resultando en una fuerza de restauración que aumenta de forma no lineal doblando la micropalanca. Al proseguir el estirado, la fuerza aumenta hasta que se rompen una serie de enlaces críticos que son los responsables de la resistencia de la proteína a ser estirada. Esta región se conoce como parche mecánico y es la principal responsable de mantener la proteína plegada en presencia de fuerza. En muchos casos, una vez roto el parche mecánico, la proteína se despliega completamente. Esto resulta en un aumento repentino de la longitud de la molécula expuesta a la fuerza, lo que hace que la fuerza se reduzca bruscamente y que en la curva de fuerza-extensión aparezca un pico de fuerza. La altura de este pico se toma como medida de la estabilidad mecánica de la proteína. A continuación, siempre que el circuito mecánico se mantenga, el proceso descrito se repite a medida que se estira el resto de la molécula y hasta que se rompe el circuito mecánico, momento en el que la fuerza vuelve al nivel basal y el experimento ha terminado (Figura 8).

Existen limitaciones asociadas a la técnica para estudiar moléculas individuales. Por un lado, la deriva térmica suele limitar la duración de un experimento por lo que impone un límite inferior a la velocidad de estiramiento sobre los 0.01 nm/ms mientras que la resistencia generada por la viscosidad del medio líquido donde se hace el experimento puede introducir errores en la medida a velocidades superiores a 10 nm/ms [Ng et al., 2007]. Desafortunadamente, las velocidades de estirado en este rango suelen ser superiores a las encontradas en muchos procesos biológicos [Merkel et al., 1999].

### 1.5.2 *Interpretación*

En base a resultados obtenidos en ensayos en solución, el proceso de desplegamiento de una proteína se considera habitualmente bien descrito como un proceso de dos estados [Borgia et al., 2008]. En experimentos de SMFS, donde a menudo las proteínas se despliegan sin estados intermedios, parece también adecuado modelarlo como un proceso de dos estados como primera aproximación. Este tipo de procesos se caracteriza por tener un diagrama de energía con dos mínimos correspondientes a los estados

---

<sup>1</sup> La extensión de la molécula se calcula a partir de la posición del piezoeléctrico y la fuerza en cada momento, que mediante la constante de muelle de la punta permite calcular su desplazamiento.



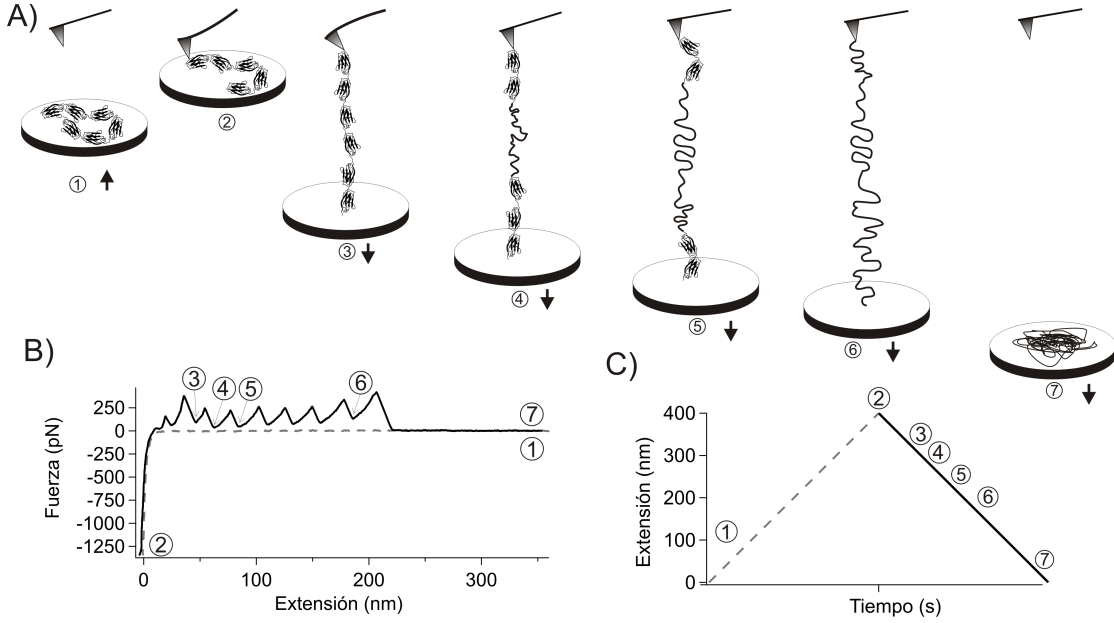


Figura 8: Representación de un experimento de SMFS en el modo de velocidad constante. A) representación esquemática de la proteína y el AFM durante el experimento. B) Diagrama Fuerza-extensión resultante de un experimento a velocidad constante. C) Extensión del posicionador piezoeléctrico durante el experimento. Al principio la superficie donde se encuentra la proteína se encuentra lejos de la punta (1). El posicionador piezoeléctrico se aproxima a velocidad constante hacia la punta y esta se dobla al entrar en contacto (2). Esto puede permitir la unión a la punta de alguna proteína presente en el sustrato. La superficie se aleja a velocidad constante estirando la proteína que se encuentra anclada entre ambos elementos (3). Esto produce un aumento de la fuerza hasta que la proteína se despliega, lo que resulta en una caída de la fuerza al aumentar su longitud efectiva (4). El proceso continua hasta desplegar el resto de elementos (6) o hasta que se desancla la proteína (7). Figura adaptada de [Galera-Prat et al., 2011].

estables (plegado y desplegado) separados por una barrera energética con una altura y distancia respecto el estado plegado características de cada proteína.

Los datos de los experimentos de SMFS se suelen interpretar mediante un modelo de dos estados basado en la teoría del estado de transición [Bell, 1978, Evans & Ritchie, 1997]. En este modelo los estados nativo y desplegado están separados por una barrera energética cuya altura relativa se reduce por la fuerza aplicada como se describe en la siguiente ecuación:

$$\alpha(F) = A \cdot \exp\left(\frac{-(\Delta G^\ddagger - F \cdot \Delta x_D)}{K_B T}\right) = \alpha_0 \cdot \exp\left(\frac{F \cdot \Delta x_D}{K_B T}\right) \quad (1)$$

Donde  $\alpha(F)$  es la tasa de desplegamiento a una fuerza  $F$ ,  $A$  representa el factor preexponencial,  $\Delta G^\ddagger$  es la altura de la barrera energética entre el estado nativo y el estado de transición,  $\Delta x_D$  corresponde a la distancia en la que se aplica la fuerza (distancia entre el estado nativo y el estado de transición en la coordenada de reacción),  $K_B T$  la energía térmica del sistema y  $\alpha_0$  la tasa de desplegamiento sin fuerza aplicada.

De la ecuación se desprende que la fuerza aplicada reduce la altura de la barrera energética en proporción al trabajo mecánico realizado, lo que facilita la transición al estado desplegado (Figura 9). La fuerza de desplegamiento ( $F_D$ ) más probable se puede predecir a partir de la ecuación 1 como:

$$F_D = \left( \frac{K_B T}{\Delta x_D} \right) \ln \left( \frac{r \cdot \Delta x_D}{\alpha_0 \cdot K_B T} \right) \quad (2)$$

donde  $r$  corresponde a la tasa de carga (la velocidad a la que cambia la fuerza). Para experimentos de velocidad constante,  $r$ , se aproxima al producto de la velocidad de estirado por la constante de muelle de la micropalanca. Por lo tanto, es de esperar que la fuerza observada en un experimento dependa del logaritmo de la velocidad a la que se ha estirado la proteína.

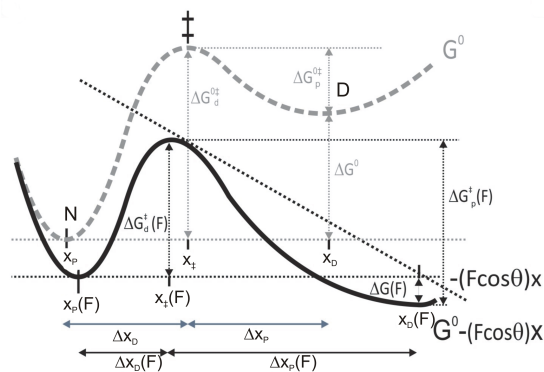


Figura 9: Representación de un diagrama energético en dos estados para una proteína cualquiera y el efecto que tiene la aplicación de fuerza sobre esta. En línea discontinua se muestra el diagrama energético de la proteína en ausencia de fuerza. El estado nativo (N) se encuentra en un mínimo energético profundo separado por una barrera energética del estado desplegado (D). El máximo de esta barrera energética corresponde con el estado de transición ( $\ddagger$ ) que se caracteriza por su altura ( $\Delta G_D^\ddagger$  respecto al estado nativo) y distancia a los estados nativo ( $\Delta x_D$ ) y desestructurado ( $\Delta x_P$ ). Según la ecuación 1 al aplicar fuerza entre los extremos de la proteína el paisaje energético se inclina de forma que se reduce la barrera, facilitando que esta sea superada a la vez que se estabiliza el estado desplegado. Figura adaptada de [Carrión-Vázquez et al., 2006].

Las fluctuaciones térmicas permiten a una proteína explorar diferentes conformaciones y, eventualmente, cruzar la barrera energética que la separa del estado desplegado. En un experimento de SMFS la aplicación de fuerza a una proteína reduce su barrera energética (aumentando la probabilidad de desplegamiento) a la vez que estabiliza el estado desplegado respecto al nativo (Figura 9). Por lo tanto, la fuerza de desplegamiento (el máximo en el pico de fuerza en un experimento a velocidad constante) de una proteína varía de forma estocástica dependiendo del estado energético en el momento de estirarla. Esto produce que la distribución de fuerza de desplegamiento para una proteína sea relativamente ancha y dependa tanto de la energía térmica como de la distancia al estado de transición [Ng et al., 2007]. Por eso la estabilidad mecánica

de una proteína se define como el promedio de las fuerzas de desplegamiento a una determinada velocidad de estiramiento.

La dependencia de la rigidez de la proteína con la fuerza aplicada hace que la tasa de carga varíe durante un experimento a velocidad constante, lo que impide aplicar directamente la ecuación 2. Para poder obtener los parámetros cinéticos ( $\alpha_0$  y  $\Delta x_D$ ) del proceso, es necesario medir la estabilidad mecánica de una proteína a distintas velocidades para luego, mediante simulaciones de Monte Carlo modelando el comportamiento de una proteína al ser estirada, obtener una estimación de estos [Oberhauser et al., 1998, Rief et al., 1998, Carrion-Vazquez et al., 1999b].

Una pregunta interesante es si las barreras energéticas estudiadas mediante SMFS son las mismas que las estudiadas por otros métodos (por ejemplo químicos o térmicos) habituales. A diferencia de los parámetros termodinámicos, que son variables de estado y por tanto solo dependen del estado inicial y final, la estabilidad mecánica de una proteína es un parámetro cinético y por tanto depende de la ruta seguida [Bustamante et al., 2004]. Los experimentos de SMFS imponen una coordenada de reacción bien definida y distinta a la de los experimentos tradicionales con las proteínas libres y en solución [Fowler et al., 2002, Bustamante et al., 2004, Javadi et al., 2013] por lo que en general los resultados no tienen por que ser comparables [Carrion-Vazquez et al., 2000, Best et al., 2001, Brockwell et al., 2002, Valbuena et al., 2009]. Además, tanto la ruta de desplegamiento como la barrera que se supera dependen de la geometría de estiramiento [Brockwell et al., 2003, Carrion-Vazquez et al., 2003]. Es importante destacar que por estas mismas razones, no es posible inferir la estabilidad mecánica a partir de la termodinámica, por lo que la única forma de acceder a esta información es mediante los experimentos de SMFS.

### 1.5.3 Identificación de las proteínas

Un punto clave en todos los experimentos de molécula individual consiste en asegurar que los datos obtenidos provienen realmente de una sola molécula y no de un conjunto y que además se trata de la molécula que se quiere estudiar.

La presencia de contaminantes o de proteína parcialmente desplegada o agregada puede enmascarar los eventos que realmente son originados por el proceso de desplegamiento que se quiere estudiar. Esto hace necesario usar métodos que permitan identificar la proteína estudiada y establecer unos criterios de selección para distinguir los registros que contienen información del sistema estudiado para evitar contaminar los datos con ruido. Estos criterios se basan en la identificación de características propias de las moléculas a estudiar que permiten seleccionar con seguridad los datos obtenidos en un experimento [Carrión-Vázquez et al., 2006].

Al estirar una proteína a velocidad constante se obtienen patrones de fuerza contra extensión no lineales que pueden ser ajustados por modelos de elasticidad de polímeros semi-flexibles. El modelo más ampliamente usado es la forma interpolada del modelo de la cadena gusaniforme (en inglés Worm Like Chain) (WLC) [Bustamante

et al., 1994, Marko & Siggia, 1995] (ecuación 3) que también se ha aplicado a otros biopolímeros al someterlos a fuerza. Este modelo considera un polímero como una cadena flexible con una longitud de contorno ( $L_C$ ) (la longitud de la molécula completamente estirada, a fuerza infinita) y permite predecir la fuerza de restauración del polímero en función de la extensión de sus extremos y su longitud de persistencia ( $p$ ), un parámetro que se relaciona con su rigidez.

$$F_{WLC} = \frac{K_B T}{p} \left[ \frac{1}{4 \left(1 - \frac{x}{L_C}\right)^2} - \frac{1}{4} + \frac{x}{L_C} \right] \quad (3)$$

Si consideramos el proceso de desplegamiento de una proteína mediante SMFS se observa que al aumentar la distancia entre punta y sustrato la fuerza registrada sigue el modelo WLC con una  $L_C$  igual a la longitud de la molécula plegada más la de las secuencias espaciadoras que contenga (Figura 10). Una vez superada la resistencia mecánica de la proteína, esta se despliega y expone a la fuerza la región de la proteína que antes estaba protegida. Si se sigue el proceso de estirado, la fuerza nuevamente vuelve a aumentar siguiendo el modelo WLC pero esta vez con una  $L_C$  correspondiente a la longitud de la molécula desplegada y las secuencias espaciadoras. Ajustando el modelo WLC a los dos picos de fuerza, se puede calcular un nuevo parámetro, el incremento de longitud de contorno ( $\Delta L_C$ ), que es una medida de la longitud de secuencia<sup>2</sup> escondida tras la barrera que protege a la proteína ante el desplegamiento por fuerza (Figura 10). El  $\Delta L_C$  es un parámetro específico de cada molécula y depende de su plegamiento<sup>3</sup>, por lo que representa una huella identificativa de esta. Por otro lado, una  $p$  única también se ha usado para la identificación de moléculas individuales [Watanabe et al., 2002].

En un experimento de SMFS, cuando la punta se encuentra cerca del sustrato, suelen producirse interacciones no específicas con otros elementos de la superficie que también suelen generar picos de fuerza. Debido a esto, la región proximal del espectro (primeros 30 - 70 nm) es habitualmente muy ruidosa (aparecen picos de fuerza no identificables, ver Figura 8.C) y puede enmascarar fácilmente la señal de la proteína de interés. Además, ha de tenerse en cuenta que distintas proteínas pueden presentar  $\Delta L_C$  similares.

Para superar este problema, se desarrolló una estrategia basada en la construcción de proteínas de fusión con múltiples repeticiones en tándem de la proteína de interés [Carrion-Vazquez et al., 1999b] conocidas como poliproteínas. Para una poliproteína, el desplegamiento de cada módulo se observa como un pico de fuerza similar y que

<sup>2</sup> Se puede calcular el número de aminoácidos escondidos de la fuerza como  $\Delta L_C / (\Delta L_C \text{ de un aminoácido})$  [Ainavarapu et al., 2007].

<sup>3</sup> El  $\Delta L_C$  es un parámetro característico de una molécula con un plegamiento definido y al ser estirada en una geometría concreta. En el caso de cambiar la geometría el  $\Delta L_C$  [Dietz et al., 2006] puede ser distinto a pesar de tratarse de la misma molécula. Por otro lado, en algunas proteínas intrínsecamente desordenadas, este parámetro puede ser muy variable incluso al estirar en la misma geometría, lo que refleja una amplia variedad de conformaciones donde las barreras al desplegamiento mecánico se pueden encontrar en múltiples sitios [Hervás et al., 2012].

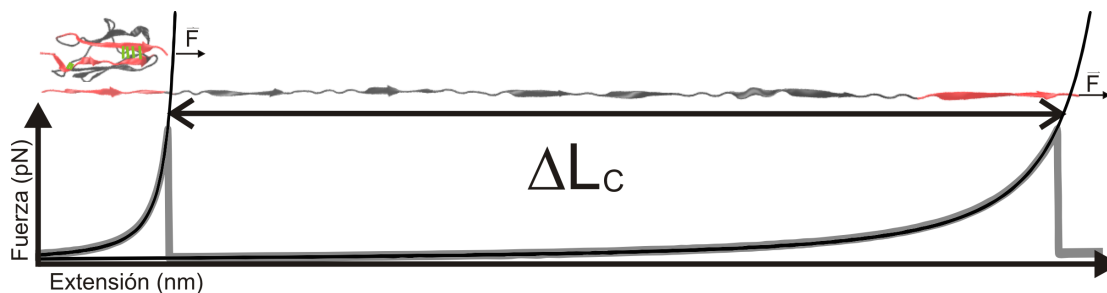


Figura 10: Al desplegar una proteína y ajustar el WLC al pico de desplegamiento y al siguiente se puede calcular  $\Delta L_C$  entre ambos ajustes. Este parámetro es característico de una molécula y su plegamiento, lo que puede permitir su identificación. Se muestra un módulo I27 (código PDB: 1TIT) antes (arriba) y después (centro) de su desplegamiento mecánico. En rojo se resalta el parche mecánico, la región que se encuentra sometida a fuerza antes de su desplegamiento. La estructura secundaria en la molécula desplegada se mantiene en la figura para facilitar la comparación.

libera la misma longitud de proteína (el  $\Delta L_C$  del módulo). Esto resulta en un patrón periódico conocido como patrón en dientes de sierra, que permite identificar la molécula.

Una alternativa menos laboriosa consiste en fusionar la proteína interés entre varias repeticiones de un módulo marcador como el I27, son las llamadas hetero-poliproteínas [Li et al., 2001, Steward et al., 2002]. Las repeticiones del módulo marcador actúan como poliproteína generando un patrón repetitivo que permite identificar la molécula [Steward et al., 2002]. Además se han construido vectores que contienen varias repeticiones de un módulo marcador separados cada uno de ellos por distintas dianas de restricción, con lo que generar este tipo de proteínas es más sencillo [Steward et al., 2002, Oroz et al., 2012]. Los módulos marcadores adecuados pueden además promover la solubilidad de la proteína de interés facilitando la expresión de esta [Steward et al., 2002], sin embargo se limita la cantidad de datos que se obtienen en cada registro.

El uso de poliproteínas, por lo tanto, permite identificar aquellos registros generados a partir del estiramiento de una molécula individual al producir patrones reconocibles.

## 1.6 SIMULACIONES DE DINÁMICA MOLECULAR

El  $\Delta L_C$  obtenido en los experimentos de SMFS proporciona información de la longitud de secuencia protegida de la fuerza, sin embargo suele ser difícil asignar con seguridad cual es la región de la proteína responsable de esta resistencia. Para obtener información estructural de como la fuerza promueve el desplegamiento de una proteína es necesario estudiar el proceso mediante simulaciones. La combinación de simulaciones de dinámica molecular (en inglés Molecular Dynamics) (MD) con los experimentos de SMFS puede proporcionar información detallada a nivel atómico de la posible estructura o los contactos presentes en el estado de transición [Evans & Ritchie, 1997], así como de la evolución de la estructura durante el proceso de estirado.

Las simulaciones por MD permiten calcular el movimiento de partículas individuales a partir de las ecuaciones de movimiento. Empleando los potenciales adecuados, estas simulaciones se pueden aplicar a moléculas como proteínas y además, introduciendo potenciales externos en una técnica conocida como dinámica molecular dirigida (en inglés Steered Molecular Dynamics) (SMD), se puede emular el estirado de la proteína como ocurre en un experimento de SMFS. Esta estrategia se ha usado ampliamente en estudios de moléculas individuales donde habitualmente reproduce (e incluso en ocasiones predice [Lu et al., 1998, Carrion-Vazquez et al., 1999b]) muy bien los resultados experimentales [Valbuena et al., 2009], proporcionando además los detalles atómicos del proceso.

Las simulaciones de MD han revelado que al aplicar fuerza a una proteína desde sus extremos, esta actúa en primer lugar en una región muy concreta de la proteína. Esto es debido a la existencia de elementos locales de estructura secundaria capaces de soportar la fuerza aplicada y de oponerse al desplegamiento de la proteína completa. Estas regiones se conocen como parche mecánico y suelen estar conectadas por puentes de hidrógeno cuyo número [Valbuena et al., 2009] además del grado de empaquetamiento de sus cadenas laterales [Ng et al., 2005], suelen ser los determinantes principales de la estabilidad mecánica de la proteína.

Las simulaciones de MD suelen ser costosas a nivel computacional pues en principio se tiene que calcular el movimiento de cada átomo del sistema y las fuerzas que actúan sobre este en cada paso de tiempo, que suele ser del orden de 1 femtosegundo, para asegurar la estabilidad de la simulación [Amber, 2012]. Por lo tanto, realizar simulaciones durante tiempos razonables incluso para proteínas no muy grandes suele ser un proceso lento. Por ello, para simular mediante SMD el desplegamiento de una proteína en tiempos razonables se suelen emplear velocidades de estiramiento muy superiores a las experimentales [Lu et al., 1998].

Para poder emplear este método con proteínas más grandes y/o tiempos de simulación más largos se pueden usar aproximaciones que reduzcan los cálculos necesarios. Una posible aproximación consiste en eliminar las moléculas del solvente, que suelen ser una fracción muy importante de los átomos del sistema, y reemplazarlas por un modelo continuo que de forma implícita aporta el efecto del solvente como el Generalizado de Born [Still et al., 1990].

Otra alternativa se basa en el uso de modelos aún más simplificados conocidos como modelo de grano grueso (en inglés Coarse-Grain) (CG), donde distintos grupos de átomos se consideran como una sola entidad de forma que el número de cálculos necesarios se reduce drásticamente, permitiendo simular sistemas mucho más grandes y durante tiempos más largos. Este tipo de aproximación ha permitido simular durante microsegundos una estructura grande como el celulosoma de *C. thermocellum*, modelando la escafoldina y varias enzimas como elementos formados por pocas esferas que reproducen el volumen de la proteína, y estudiar como distintos parámetros de enzimas con doquerina influyen su unión a las cohesinas de la escafoldina [Bomble et al., 2011].

Modelando cada aminoácido de una proteína como una sola partícula y considerando la estructura cristalográfica depositada en el PDB, se ha simulado el estirado de todas las proteínas de menos de 250 aminoácidos de las que se conocía estructura [Sikora et al., 2009]. Gracias al uso de modelos simplificados, en este estudio a gran escala se pudo estudiar la estabilidad mecánica de 17,134 proteínas, identificando varios módulos cohesina entre las proteínas más estables mecánicamente.

### 1.7 MECÁNICA DEL CELULOSOMA

En un estudio pionero se planteó la hipótesis de que el celulosoma estuviese sometido a estrés mecánico [Valbuena et al., 2009] ya que este puede actuar como sistema de adhesión entre la célula bacteriana que lo produce y un sustrato de celulosa [Bayer et al., 1983]. Como se ha comentado (sección 1.4), es habitual que los sistemas de adhesión estén sometidos a estrés mecánico [Bustamante et al., 2004]. Esto es debido a que el movimiento relativo entre los dos cuerpos unidos puede generar fuerzas que se transmiten a través del elemento que los mantiene unidos, que en el caso del celulosoma se trata de la escafoldina.

Se puede hipotetizar que si la aplicación de fuerza sobre las cohesinas llega a producir su desplegamiento, estas perderían su estructura nativa, incluyendo la superficie de unión a doquerina, de manera que se podrían liberar las enzimas con lo que se perdería la sinergia que aporta su incorporación en la escafoldina [Valbuena et al., 2009]. Por lo tanto, en este caso las propiedades mecánicas de los módulos cohesina serían determinantes para la actividad del complejo y por lo tanto la estabilidad mecánica sería una propiedad que podría haber sido seleccionada durante la evolución de este complejo.

Es interesante observar que en la escafoldina CipA también se encuentran 2 módulos cohesina en el extremo N-terminal que están fuera de la zona de anclaje [Gerngross et al., 1993]. Dado que se espera que las fuerzas actúen entre los puntos de anclaje, estos módulos no estarían sometidos a estrés mecánico y por tanto sus propiedades mecánicas no serían tan importantes ni habrían sufrido la misma presión de selección. En base a esto se denominó región conectora a la región de la escafoldina que se espera que esté sometida a estrés mecánico y región externa a la que se encuentra fuera de los puntos de anclaje [Valbuena et al., 2009].

Para verificar esta hipótesis se realizó un primer estudio de SMFS para caracterizar las propiedades mecánicas de varios módulos cohesina [Valbuena et al., 2009]. En concreto se analizaron las cohesinas 2 y 7 de CipA de *C. thermocellum* y la 1 de CipC de *C. cellulolyticum* y se observó que la cohesina 7, proveniente de la región conectora, presentaba una estabilidad mecánica muy elevada. Mientras que la cohesina 2, de la región externa, presentaba una estabilidad mecánica mucho más reducida. Además, se identificó el parche mecánico de estos módulos, formado por la primera y última hebra  $\beta$  formando un parche compuesto por 10 puentes de hidrógeno en la cohesina 7, lo que se utilizó como argumento para explicar el origen de la elevada estabilidad

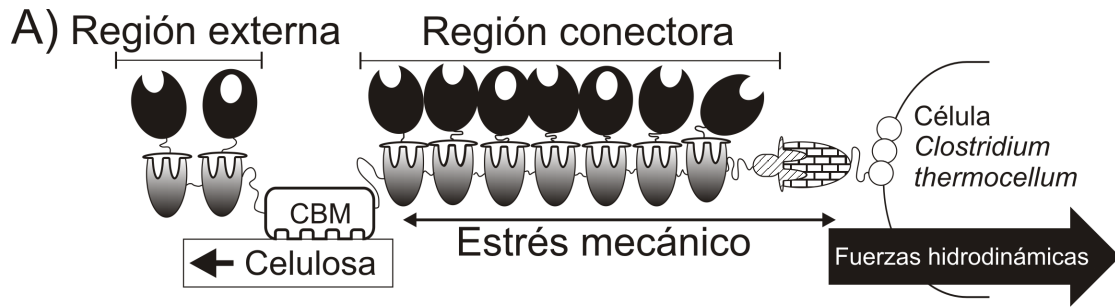


Figura 11: La hipótesis mecánica del celulosoma. El movimiento de la célula respecto al sustrato puede resultar en la aplicación de estrés mecánico a la región de la escafoldina situada entre los dos puntos de anclaje (región conectora). Por otro lado, las cohesinas que no estén situadas entre los dos puntos de anclaje (región externa) no sentirían dicho estrés.

mecánica de este módulo [Valbuena et al., 2009]. Estos resultados son compatibles con la hipótesis planteada pero no permiten demostrarla.

En esta tesis nos planteamos profundizar en el estudio de la nanomecánica del celulosoma. Hasta el momento se conoce la estabilidad mecánica de 3 módulos cohesina por lo que nos planteamos estudiar otros módulos con secuencias más divergentes y también de otros celulosomas más complejos para estudiar si el comportamiento observado es generalizable a todos los celulosomas. Además nos planteamos cómo distintos factores que se encuentran en los celulosomas naturales, como la presencia de secuencias intermodulares de distintos tamaños y composiciones entre las cohesinas o la interacción con doquerina, pudieran afectar a estas propiedades mecánicas.



Parte II

OBJETIVOS

## OBJETIVOS

---

El objetivo principal de esta tesis es el estudio de la hipótesis mecánica del celulosoma. Según esta, el celulosoma está sometido a estrés mecánico y por tanto sus propiedades mecánicas pueden ser relevantes para entender su funcionamiento y para el diseño de celulosomas más eficientes para su posible uso industrial. Para ello se ha planteado como objetivo estudiar en primer lugar las propiedades mecánicas de módulos cohesina aislados, y a continuación incorporando otros elementos encontrados en los celulosomas naturales. Además se ha planteado el objetivo de estudiar la relación entre la actividad de los celulosomas y la estabilidad mecánica de sus cohesinas y finalmente la posibilidad de incorporar celulosomas en *E. coli*.

Para alcanzar estos objetivos generales se han planteado los objetivos concretos siguientes:

1. Estudiar la estabilidad mecánica de módulos cohesina divergentes del celulosoma modelo producido por *C. thermocellum*.
2. Determinar la estabilidad mecánica de varios módulos cohesina de celulosomas complejos como el de *A. cellulolyticus*, en particular tanto de la región externa como conectora de las escafoldinas ScaA, ScaB y ScaC.
3. Caracterizar el efecto de las secuencias intermodulares naturales en la estabilidad mecánica de los módulos cohesina.
4. Estudiar la estabilidad mecánica de cohesina al estar interaccionando con el dominio doquerina de una celulasa.

Parte III

MATERIALES Y MÉTODOS

## MATERIALES Y MÉTODOS

---

### 3.1 NOMENCLATURA EMPLEADA PARA LAS COHESINAS

Para unificar la nomenclatura de los distintos módulos cohesina empleados a lo largo de la tesis estos se referirán con nombres del tipo «GeSN», donde Ge indican el género y la especie del microorganismo del que proviene, S indica la escafoldina en la que se encuentra y N es un número que indica la posición dentro de esta. Así, el módulo 3 de la escafoldina ScaC de *A. cellulolyticus* se indica como AcC3 y el módulo 7 de la escafoldina CipA de *C. thermocellum* se indica como CtA7. Para las construcciones donde además de la cohesina se incluyen las secuencias intermodulares naturales de la misma escafoldina se añade al final del nombre la letra l (del inglés *linker*). Por ejemplo, para CtA7l indica que se trata de la cohesina CtA7 con sus correspondientes secuencias intermodulares.

### 3.2 CLONACIÓN MOLECULAR

Los métodos de clonación molecular empleados en esta tesis se basan en los descritos en [Sambrook & Russell, 2001]. Todos los oligonucleótidos se sintetizaron por Sigma-Aldrich (St. Louis, MO), los dNTPs se adquirieron a Invitrogen (Thermo Fisher Scientific, Waltham, MA) y las enzimas usadas se adquirieron a distintos proveedores: enzimas de restricción, polimerasa de ADN *taq*, polinucleótido quinasa y fosfatasa en New England Biolabs (Ipswich, MA), enzimas de restricción y T4-ligasa a Fermentas (Thermo Fisher Scientific, Waltham, MA) y polimerasa de ADN *phusion* de Fynnzymes (Thermo Fisher Scientific, Waltham, MA). Las secuencias clonadas tras un paso de PCR se secuenciaron en el servicio de secuenciación de Secugen S. L. (Madrid, España). La transformación de los vectores se realizó en bacterias «z-competentes» elaboradas mediante el procedimiento de Zymo Research (Irvine, CA). En las etapas de clonación, se usaron las cepas XL1-Blue y DH5 $\alpha$ .

Para determinar los extremos de los módulos cohesina cuya estructura no está resuelta se realizaron alineamientos de su secuencia con la de otras cohesinas de estructura conocida para estimar si realmente se había seleccionado el módulo completo y minimizar así la posibilidad de trabajar con estructuras incompletas, lo que podría afectar a las propiedades mecánicas estudiadas.

### 3.2.1 Clonación de hetero-poliproteínas en pAFM

Para el análisis por SMFS, se generaron hetero-poliproteínas donde la cohesina de interés se insertó en el vector pAFM [Steward et al., 2002], que fue amablemente cedido por la Dra. Jane Clarke (Universidad de Cambridge). Este vector está basado en el vector de expresión pRSET-A (Invitrogen Thermo Fisher Scientific, Waltham, MA) al que se han insertado 8 repeticiones del módulo marcador I27 de titina cardíaca humana, cada una de ellas separada por una diana de restricción distinta. Esto permite seleccionar tanto el número como la posición de los módulos I27 que se van a reemplazar por otra secuencia de interés para generar una hetero-poliproteína de forma directa. Además, este vector incorpora una etiqueta de histidinas que se utiliza tanto para la purificación de la poliproteína como para la unión a la superficie de cubreobjetos funcionalizados con ácido nitrilotriacético (NTA)-Ni<sup>2+</sup>, y también incorpora 2 residuos cisteína en el extremo C-terminal de la proteína resultante que permiten su unión a superficies de oro usadas en experimentos de SMFS, en nuestro caso habitualmente como recubrimiento de las puntas de las micropalanca.

En la mayoría de los casos se optó por las dianas KpnI y SpeI, lo que resulta en una construcción del tipo (I27)<sub>3</sub>-Cohesina-(I27)<sub>2</sub>. De esta forma, dado el tamaño de un módulo cohesina (aproximadamente 15 kDa) y el de los módulos I27 (aproximadamente 10 kDa), las construcciones resultantes no son excesivamente grandes (unos 70 kDa) como para comprometer su expresión en *E. coli* y el número de marcadores mono-moleculares que es necesario observar en un experimento de SMFS es razonable. Esta estrategia se empleó para generar las siguientes construcciones:

- pAFM-CtA1: El módulo cohesina 1 de CipA de *C. thermocellum* (CtA1) se amplificó a partir de ADN genómico obtenido anteriormente [Valbuena et al., 2009]. Este se clonó entre las dianas KpnI/SpeI.
- pAFM-CtA9: El módulo cohesina 9 de CipA de *C. thermocellum* (CtA9) se amplificó a partir del plásmido con la región conectora de la misma escafoldina [Valbuena et al., 2009].
- pAFM-CtA7I: El módulo 7 de CipA de *C. thermocellum* con las secuencias intermodulares (CtA7I) se amplificó a partir del plásmido pAFM-c7A [Valbuena et al., 2009] que contiene el módulo deseado y se utilizaron unos oligonucleótidos que incluían la secuencia codificante de las regiones intermodulares de ambos extremos.
- pAFM-CcC1I: El módulo 1 de CipC de *C. cellulolyticum* con sus secuencias intermodulares en ambos extremos (CcC1I) se amplificó a partir del plásmido pET-Coh1B [Spinelli et al., 2000], amablemente cedido por el Dr. Henri-Pierre Fierobe (Aix-Marseille Université-CNRS, Francia). Este plásmido contiene el módulo CcC1, que se amplificó empleando unos oligonucleótidos que contenían las secuencias intermodulares correspondientes.

- pAFM-CcC3l: El módulo 3 de CipC de *C. cellulolyticum* con sus secuencias intermodulares (CcC3l) se obtuvo a partir de un vector generado en [Valbuena et al., 2009] que contenía un fragmento de la escafoldina CipC que abarcaba desde el módulo cohesina 1 hasta la mitad del 4 y que se había generado a partir de ADN genómico. CcC3 presenta una elevada identidad de secuencia con otras cohesinas de la escafoldina (incluyendo CcC2) por lo que se clonó en dos pasos: Se realizó una PCR con un oligonucleótido capaz de reconocer el extremo 5' de las cohesinas 2, 3 y 4 y otro capaz de unirse a la región central de la cohesina 4. Esto resultó en 3 productos de amplificación de tamaños distintos, correspondientes a la región CcC2-4 (1.1 kb), CcC3-4 (650 pb) y CcC4 (200 pb). El fragmento de 650 pb se purificó mediante un gel de agarosa y se empleó como molde para una segunda PCR en la que se emplearon oligonucleótidos específicos para el inicio y final del módulo CcC3l. El producto de esta PCR se clonó entre las dianas KpnI/SpeI.
- pAFM-CcC8: El módulo 8 de CipC de *C. cellulolyticum* (CcC8) se amplificó a partir de ADN genómico obtenido anteriormente [Valbuena et al., 2009].
- pAFM-CcC123: La construcción con la región que contiene los módulos 1 a 3 de CipC de *C. cellulolyticum* se realizó en varios pasos. La cohesina CcC3 se amplificó a partir de pAFM-CcC3l utilizando un oligonucleótido que introducía una mutación puntual silenciosa (sin efecto en la secuencia de la proteína) en la región de la secuencia intermodular de modo que incorporase una diana KpnI y un oligonucleótido 3' con la diana SpeI. La construcción se clonó en el vector pAFM entre las dianas KpnI/SpeI. A continuación se amplificó la cohesina CcC1 a partir de pAFM-c1C [Valbuena et al., 2009] incorporando una diana BssHIII en el extremo 5' y una mutación silenciosa en el extremo 3' de modo que se generase la diana KpnI. Esta secuencia se clonó en el vector mencionado entre las dianas BssHIII/KpnI con lo que se generó el plásmido pAFM-CcC1-intermodular-3. Finalmente la secuencia de CcC2 con la secuencia intermodular en 5' se amplificó a partir del plásmido generado en [Valbuena et al., 2009] con el fragmento entre las cohesinas 1 y 4 de CipC. Se empleó una aproximación similar a la seguida para pAFM-CcC3l seleccionando la banda de 1.1 kb en la PCR y empleando este fragmento como molde de una segunda PCR con los oligonucleótidos adecuados (ambos con dianas KpnI introducidas por mutación silenciosa y manteniendo la secuencia intermodular del extremo 5'). El fragmento resultante se clonó por extremos romos en el vector pT7-blue (Novagen, Merck Millipore, Darmstadt, Alemania) cortado previamente con la diana SmaI (que también deja extremos romos). El fragmento de la PCR contenía tanto CcC2 como CcC3, que se pudieron distinguir mediante la secuenciación de varios clones. Finalmente, se digirió el vector pT7-blue-CcC2 con KpnI y el fragmento se clonó en el vector pAFM-CcC1-intermodular-3 para generar la construcción deseada pAFM-CcC123. La construcción resultante se secuenció de nuevo para asegurar que era la correcta y era de la forma (I27)<sub>2</sub>-CcC123-(I27)<sub>2</sub>.

- pAFM-AcA3: El módulo cohesina 3 de ScaA de *A. cellulolyticus* (AcA3) se amplificó a partir del plásmido pET28-AcA3-CBM [Hamberg et al., 2014], amablemente cedido por el Dr. Edward A. Bayer (Weizmann Institute of Science, Israel), y se clonó entre las dianas KpnI/SpeI.
- pAFM-AcA4: El módulo cohesina 4 de ScaA de *A. cellulolyticus* (AcA4) se amplificó a partir del plásmido pET28-AcA4-CBM [Hamberg et al., 2014], amablemente cedido por el Dr. Edward A. Bayer, y se clonó entre las dianas KpnI/SpeI.
- pAFM-AcB4: El módulo cohesina 4 de ScaB de *A. cellulolyticus* (AcB4) se amplificó a partir del plásmido pET28-AcB4-CBM [Hamberg et al., 2014], amablemente cedido por el Dr. Edward A. Bayer, y se clonó entre las dianas KpnI/SpeI.
- pAFM-AcC3: El módulo cohesina 3 de ScaC de *A. cellulolyticus* (AcC3) se amplificó a partir del plásmido pET28-AcC3-CBM [Hamberg et al., 2014] amablemente cedido por el Dr. Edward A. Bayer empleando unos oligonucleótidos diseñados para insertar una diana KpnI en 5' y NheI en 3'. El producto de PCR purificado se digirió con las enzimas KpnI/NheI y se clonó en el vector pAFM previamente digerido con las dianas KpnI/SpeI, aprovechando que las dianas NheI y SpeI generan extremos compatibles.

### 3.2.2 Homo-poliproteína de cohesina

La homo-poliproteína de 8 repeticiones de CtA1 se generó en el vector de expresión pET28a (Novagen, Merck Millipore, Darmstadt, Alemania). En primer lugar se amplificó por PCR la secuencia de CtA7 empleando unos oligonucleótidos diseñados para introducir en el extremo 5' la diana NheI y en el extremo 3' la diana SpeI seguida de dos codones para cisteína, dos codones de terminación y una diana XhoI. Este fragmento se clonó entre las dianas NheI/XhoI en el vector pET28a resultando un vector que contenía una etiqueta de histidinas (para la purificación y unión de la proteína en los cubreobjetos usados en los experimentos de SMFS), la secuencia de la cohesina entre las dianas NheI-SpeI y dos cisteínas para el anclaje de la proteína a superficies de oro ((CtA7)<sub>1</sub>). A continuación se amplificó la secuencia de CtA1 empleando como molde la construcción pAFM-CtA1 y añadiendo las dianas NheI y SpeI en los extremos 5' y 3', respectivamente. El fragmento resultante se clonó en el vector (CtA7)<sub>1</sub> entre las dianas NheI/SpeI, sustituyendo la cohesina CtA7, para generar el vector (CtA1)<sub>1</sub>. Aprovechando que NheI y SpeI forman extremos compatibles, se construyó la poliproteína a base de clonar reiteradamente el fragmento entre las dianas NheI/SpeI a otro vector igual cortado únicamente en la diana SpeI, sin generar una diana reconocible por enzimas de restricción. De esta forma se puede generar un vector con 2, 4 y finalmente 8 repeticiones de la cohesina de interés ((CtA1)<sub>8</sub>).

### 3.3 EXPRESIÓN

Las proteínas se expresaron en las cepas de *E. coli* BL21 (DE3), BLR (DE3) (Novagen, Merck Millipore, Darmstadt, Alemania), BL21 Star (DE3) (Invitrogen) o C41 (DE3) [Miroux & Walker, 1996]. Los cultivos se crecieron en medio LB durante 2-4 h a 37°C hasta alcanzar una densidad óptica medida a 595 nm ( $OD_{595}$ ) entre 0.5 y 0.8. A continuación se añadió Isopropil- $\beta$ -D-tiogalactopiranosido (IPTG) para inducir la expresión de las distintas proteínas. En el caso de las proteínas para el análisis por SMFS la expresión se llevó a cabo durante 4 h a 37°C añadiendo 1 mM de IPTG. Cel8A (amablemente cedida por el Dr. Edward A. Bayer) a 37°C 4 h añadiendo de 1 mM IPTG y 1 mM  $CaCl_2$ .

### 3.4 PURIFICACIÓN

La purificación de las proteínas expresadas de forma recombinante en *E. coli* se inició con la lisis del cultivo mediante el método de lisis enzimática [Sambrook & Russell, 2001] empleando 1 mg/ml de lisozima y 1 % Tritón X-100. El tampón de lisis empleado fue 50 mM de tampón fosfato de sodio/500 mM NaCl/50 mM imidazol, pH 7.4 para todas las proteínas a excepción de las que contenían doquerina (Cel8A, GFP-XDoquerina y mini-escafoldinas) además de pAFM-CtA7 y pAFM-CtA2 [Valbuena et al., 2009] para las que se empleó el tampón 20 mM Tris/500 mM NaCl/1 mM  $CaCl_2$ /50 mM Imidazol, pH 7.5.

Los distintos pasos del proceso de purificación se siguieron mediante PAGE en condiciones desnaturalizantes por la presencia de SDS (PAGE-SDS), que se utilizó también para evaluar la pureza de las muestras. La concentración de proteína se midió a partir de su densidad óptica a 280 nm mediante NanoDrop (ThermoFisher Scientific, Waltham, Massachusetts, USA) y su coeficiente de extinción molar, calculado mediante ProtParam (<http://web.expasy.org/protparam/>).

#### 3.4.1 Afinidad por $Ni^{2+}$

Los clarificados se purificaron mediante cromatografía de afinidad por iones  $Ni^{2+}$  en columnas HisTrap HP 5 ml (GE Healthcare, Uppsala, Suecia) en un aparato FPLC (ÄKTA Purifier, GE Healthcare). Los clarificados se cargaron en la columna empleando como tampón de unión el mismo usado en la etapa de lisis y que también se usó para lavar la columna. Finalmente las proteínas unidas se eluyeron empleando el mismo tampón pero con una mayor concentración de imidazol (500 mM) para desplazar la interacción entre la etiqueta de histidinas de las proteínas y la resina.

#### 3.4.2 Purificación por intercambio iónico

En los casos en los que se estimó necesario (para Cel8A, pAFM-AcA3, pAFM-AcA4, pAFM-AcB4 y pAFM-AcC3), se realizó otro paso de purificación empleando una co-



lumna de cromatografía de intercambio aniónico HiTrap Q (GE Healthcare). Para ello las muestras se concentraron en el tampón adecuado mediante ultrafiltración empleando Amicon 3k o 10K (Merck Millipore, Darmstadt, Alemania). Los tampones usados fueron 50 mM tampón fosfato de sodio/100 mM NaCl/1 mM DTT/1 mM EDTA a pH 8 para las muestras sin doquerina o alternativamente 50 mM Tris/100 mM NaCl/1 mM CaCl<sub>2</sub>, pH 8, para las que sí contenían doquerina. Las proteínas se cargaron a la columna y esta se lavó en el mismo tampón de carga. A continuación se eluyeron las proteínas realizando un gradiente desde el tampón de carga hasta un segundo tampón con la misma composición pero con mayor concentración de NaCl (1 M).

### 3.4.3 *Purificación por exclusión molecular*

En la mayoría de las proteínas para análisis por SMFS se realizó una purificación por cromatografía de exclusión molecular después del paso de purificación por afinidad. Esta purificación se realizó empleando una columna Hiloal 16/60 (GE Healthcare) y usando un tampón que contenía 100 mM NaP/100 mM NaCl/1 mM DTT/1 mM EDTA, pH 7.4. La muestra se concentró en un volumen inferior a 1.5 ml mediante ultrafiltración y se cargó en la columna adaptando el flujo de circulación del tampón al tamaño de la proteína.

## 3.5 ANÁLISIS DE PROTEÍNAS

### 3.5.1 *Western-Blot*

El análisis por Western-Blot se realizó siguiendo protocolos habituales. Se emplearon membranas Hybond ECL (GE Healthcare) a las que se transfirieron las proteínas, que se detectaron empleando un anticuerpo monoclonal capaz de reconocer la etiqueta de histidinas (Novagen) a una dilución 1:2000. Como anticuerpo secundario se usó anti-IgG de ratón (animal del que proviene el anticuerpo primario) conjugado a peroxidasa de rábano (Novagen). El resultado se reveló mediante el sistema de detección ECL (GE Healthcare).

### 3.5.2 *Electroforesis en condiciones nativas*

Las electroforesis en condiciones nativas (PAGE-nativa) se prepararon como los geles PAGE-SDS pero empleando soluciones sin SDS y usando un tampón a pH 9.5 para el gel separador. Una vez polimerizados, las muestras se cargaron en tampón de carga para geles nativos (Laemli sin SDS [Sambrook & Russell, 2001]) y el gel se corrió a 90 V durante 3 h y 30 min. Como en los geles desnaturalizantes, una vez terminada la electroforesis el gel se tiñó en solución de tinción con azul de Coomassie (40% v/v metanol, 10% v/v ácido acético glacial, 0.25% p/v Coomassie Brilliant Blue R-250) y

posteriormente se incubó en solución de destinción (20 % v/v metanol, 5 % v/v ácido acético glacial).

Estos geles se emplearon principalmente para comprobar la formación del complejo cohesina:doquerina. Para este tipo de experimentos, las proteínas a estudiar se incubaron en tampón de interacción (50 mM Tris/300 mM NaCl/1 mM CaCl<sub>2</sub>, pH 7.5) durante 1 h a 37°C. A continuación las muestras se analizaron por PAGE-nativa usando las mismas proteínas separadas como referencia.

### 3.6 ESPECTROSCOPIA DE FUERZA DE MOLÉCULAS INDIVIDUALES

#### 3.6.1 *Cubreobjetos*

Para los experimentos de SMFS se emplearon cubreobjetos funcionalizados con NTA-Ni<sup>2+</sup>, un grupo capaz de anclar la etiqueta de histidinas del extremo N-terminal de las proteínas empleadas en esta tesis.

Para su preparación, los cubreobjetos se sumergieron durante una noche en una solución 20 M de KOH y luego se colocaron bajo un flujo de agua MilliQ durante 1 h. A continuación se incubaron en una solución al 2 % de 3-mercaptopropil-trimetoxisilano (Sigma-Aldrich)/0.02 % ácido acético glacial a 90°C durante 1 h. Después se lavaron con agua MilliQ durante 1 h y posteriormente se incubaron 15 min en un horno a 120°C. Tras dejarlos enfriar a temperatura ambiente durante aproximadamente 10 min, los cubreobjetos se transfirieron a una solución 100 mM DTT durante 15 min y después se lavaron con agua MilliQ durante 1 h. A continuación, se depositó sobre cada cubreobjetos una gota de aproximadamente 50  $\mu$ l de una solución de 20 mg/ml de maleimida-C3-NTA (Laboratorios Dojindo), disuelta en 10 mM de ácido 3-(N-morfolino) propanosulfónico, pH 7.0 durante 30 min. Después de lavar con agua MilliQ, se añadió a cada cubreobjetos una gota (60  $\mu$ l) de 10 mM NiCl<sub>2</sub> y se incubaron durante 10 min. Por último, los cubreobjetos se lavaron brevemente en agua MilliQ antes de su almacenamiento [Hervás et al., 2012].

Alternativamente, en los experimentos para estudiar la estabilidad mecánica de la cohesina en presencia de doquerina se emplearon superficies de oro (Arrandee, Werther, Alemania).

#### 3.6.2 *Aparato y micropalancas*

La mayoría de los experimentos de SMFS realizados en esta tesis se empleó un AFM no comercial cuyo modo de funcionamiento se ha descrito anteriormente [Valbuena et al., 2007]. Las mediciones para las construcciones pAFM-AcA4, pAFM-AcB4 y pAFM-AcC3 se realizaron en un AFM comercial Atomic Force Spectrometer (AFS, Luigs and Neumann, Ratingen, Alemania).

Antes de cada experimento, la micropalanca se limpió durante 10-15 min con una lámpara de UV/ozono (UV/Ozone ProCleaner<sup>TM</sup> Plus, Bioforce Nanoscience Inc.).

La constante de muelle de cada micropalanca individual se calculó usando el teorema de equipartición. Los experimentos se realizaron empleando micropalancas BL-RC (Olympus, Tokyo, Japón) con una constante de muelle nominal de  $\sim 30$  pN/nm.

Las muestras empleadas para AFM consistían en proteínas purificadas a concentraciones variables de 0.5 a 5 mg/ml. La cantidad de proteína usada en cada experimento se determinó empíricamente, aumentando la cantidad en caso de no observar curvas buenas o disminuyéndola en caso de observar curvas saturadas (con muchos picos no identificables). La muestra se incubó en tampón adecuado (PBS/0.2 mM EDTA/5 mM DTT, pH 7.4 para todas las proteínas a excepción de los experimentos para estudiar la estabilidad mecánica de cohesinas en presencia de doquerinas en los que se empleó el tampón de interacción) durante 10-15 min. A continuación se lavó la superficie 3 veces con el mismo tampón y se inició el experimento.

Al acercar el posicionador piezoeléctrico a la micropalanca a lo largo del eje Z, las proteínas, depositadas sobre el sustrato, pueden anclarse por adhesión específica (a través de dos residuos cisteína incorporadas en el extremo C-terminal de la proteína) o adsorción a la punta de la micropalanca. El actuador piezoeléctrico fue retraído varios cientos de nm para estirar las proteínas previamente ancladas. Excepto en los experimentos de dependencia de velocidad, todos los experimentos de SMFS se realizaron a una velocidad de estiramiento de 400 nm/s en el modo de velocidad constante.

### 3.6.3 *Análisis de SMFS*

Los datos obtenidos se analizaron utilizando el programa Igor Pro 6 (Wavemetrics) empleando procedimientos previamente establecidos Valbuena et al. [2009]. La elasticidad de las moléculas estiradas se analizó utilizando el modelo de elasticidad de polímeros WLC. En las figuras se muestran las distribuciones de fuerza de desplegamiento (altura del pico de fuerza) y  $\Delta L_C$  obtenidos con el ajuste del modelo, y se dan los valores como media  $\pm$  desviación estándar.

El criterio para seleccionar las curvas adecuadas para el análisis se basó en las construcciones realizadas. Para las construcciones con 5 módulos marcadores I27 (3 en el extremo N-terminal y 2 en el C-terminal, para casi todas las construcciones excepto pAFM-CcC123 y (CtA1)<sub>8</sub>) se identificó el patrón de la cohesina en cuestión a partir del análisis de curvas con 4 ó 5 módulos marcadores (en esta geometría, 4 es el número mínimo de marcadores que garantiza haber estirado la cohesina). Para pAFM-CcC123, que consta de 4 módulos marcadores se seleccionaron registros con 3 ó 4 picos provenientes de estos (3 es el número mínimo en esta construcción).

Se rechazaron aquellos registros con más de 5 marcadores, en los que no se observaba el desanclado de la proteína o con longitud total claramente superior a la esperada para la construcción. Una vez identificado el patrón de cada molécula se relajó el criterio del número de marcadores incluyendo aquellos registros en los que se observaba el desplegamiento de cohesina (con su patrón característico de  $\Delta L_C$  y un número de módulos marcadores inferior (2 ó 3)).

### 3.7 MODELADO ESTRUCTURAL

Los modelos estructurales usados para las simulaciones de dinámica molecular y estirado por SMD se generaron usando el programa MODELLER (versión 9.1) [Sali & Blundell, 1993]. Las secuencias de las proteínas se obtuvieron de la base de datos Uniprot ([www.uniprot.org](http://www.uniprot.org)) y se emplearon como secuencia pregunta en pBLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) [Altschul et al., 1990] limitando la búsqueda sobre la base de datos de proteínas del PDB. La o las coincidencias con menor valor-E se seleccionaron como moldes para generar el modelo estructural. El conjunto de secuencias (moldes y objetivo) se alinearon mediante Clustal (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) [Sievers & Higgins, 2014] y se usaron como entrada para MODELLER. En caso de disponer de más de una estructura o conformación de residuos dentro de un mismo fichero PDB, se escogió por defecto la primera. Solo se generó un modelo en cada caso.

La calidad de los modelos estructurales generados mediante MODELLER se validó mediante el análisis de distintas características empleando PROSAweb [Wiederstein & Sippl, 2007] y Procheck [Laskowski et al., 1993]. PROSAweb permite evaluar la calidad global del modelo, mientras que Procheck analiza la estereoquímica de la estructura. En concreto se tubo en cuenta el valor Z proporcionado por PROSAweb, que permite su comparación con otras estructuras de tamaño similar, y se analizó el mapa de Ramachandran para comprobar que la mayor parte de residuos se encontraban en regiones favorables.

En particular se empleó esta técnica para la obtención de modelos estructurales de las cohesinas CtA9, CtA1, CcC3 y CcC8, empleando como molde las estructuras 1AOH, 1ANU, 1G1K y 1G1K respectivamente.

Para modelar la estructura de la cohesina CtA7 interaccionando con doquerina, se realizó un alineamiento estructural de dicha cohesina (código PDB: 1AOH) con la cadena correspondiente a la cohesina CtA2 interaccionando con la doquerina de Xyn10B resuelta previamente (código PDB: 1OHZ) Carvalho et al. [2003]. A continuación se eliminó la estructura de CtA2 del conjunto de forma que se obtuvo el modelo deseado con CtA7 interaccionando con la doquerina de Xyn10B.

### 3.8 SIMULACIONES

La simulaciones de dinámica molecular de las distintas estructuras de cohesinas se realizaron utilizando un modelo de solvente implícito, realizando correcciones a la energía debido a la exposición al solvente mediante el generalizado de Born con correcciones de superficie [Tsui & Case, 2000]. Esto permite reducir el número de átomos presentes en la simulación al eliminar las moléculas de agua pero aún así predice cualitativamente las propiedades mecánicas de forma comparable a las simulaciones donde se incluye el agua de forma explícita [Valbuena et al., 2009]. Las coordenadas atómicas iniciales se obtuvieron de la base de datos del PDB [Berman et al., 2000] o

a partir de modelos estructurales. Todos los pasos de simulación se realizaron con el programa SANDER del paquete AMBER10 [Case et al., 2008] y se utilizó el campo de fuerza AMBER99SB Cornell et al. [1995]. En el modelo de solvente se incluyó la presencia de 0.1 M de iones monovalentes. La superficie expuesta al solvente se calculó mediante el método LCPO [Hawkins et al., 1996]. El potencial electrostático se truncó a 10 Å disminuyendo hasta cero en 12 Å.

Para todas las estructuras, se realizaron en primer lugar 5000 pasos de minimización energética y restringiendo la posición de los  $C_\alpha$  (20 kcal/mol·Å<sup>2</sup>). A continuación, empleando un paso de tiempo de 1 fs se aumentó la temperatura de la molécula desde 0 a 300 (ó 340) K a razón de 15 K/ps manteniendo la misma restricción que en el paso de minimización. Una vez alcanzada la temperatura se eliminaron gradualmente las restricciones en los  $C_\alpha$  de la molécula durante 8 ps. A partir de este momento se realizaron simulaciones de dinámica libre durante 10 ns.

La estructura obtenida cada 2 ns se empleó como punto de partida para simular el proceso de estirado. Para ello se realizaron simulaciones de SMD aplicando una restricción en la distancia de los  $C_\alpha$  N- y C-terminales (5 kcal/mol·Å<sup>2</sup>) y dicha distancia se aumentó a razón de 1 Å/ps.

Las trayectorias tanto de la dinámica libre como de los estirados por SMD se analizaron mediante VMD [Humphrey et al., 1996] y los datos obtenidos se representaron con Igor Pro 6 (Wavemetrics). La raíz de la desviación media cuadrática (RMSD) durante la dinámica libre se calculó tomando como referencia la estructura al inicio de la dinámica libre y calculando el valor de RMSD para los átomos pesados del esqueleto peptídico de la proteína. En el caso de la raíz de la fluctuación media cuadrática (RMSF), se calculó la fluctuación de cada  $C_\alpha$  durante los últimos 5 ns de la dinámica libre (donde la molécula ya está estabilizada). Las curvas fuerza-extensión se calcularon a partir de las trayectorias de las simulaciones de SMD a partir de la distancia entre los  $C_\alpha$  de los extremos N- y C-terminales. La fuerza se calculó como:

$$F = -k \cdot (X_{N-C_t} - t \cdot v - X_{N-C_0}) \quad (4)$$

Donde F es la fuerza, k es la constante de la restricción empleada en las simulaciones (5 kcal/mol·Å<sup>2</sup>),  $X_{N-C_t}$  es la distancia entre los  $C_\alpha$  en el instante t, v es la velocidad de estiramiento y  $X_{N-C_0}$  la distancia entre los  $C_\alpha$  al iniciar el estirado.

Para determinar la presencia de puentes de hidrógeno del parche mecánico se consideró únicamente los átomos de la cadena peptídica y se emplearon los valores de 0.38 nm y 35 ° como puntos de corte de la distancia y ángulo de los átomos candidatos.

### 3.8.1 Simulación del desplegamiento mecánico mediante simulaciones de Monte Carlo

Las simulaciones de Monte Carlo del proceso de desplegamiento mecánico de proteínas para estimar sus parámetros cinéticos  $\alpha_0$  y  $\Delta x_D$  se basaron en el método descrito previamente [Rief et al., 1997, 1998]. Brevemente, se asume un modelo Markoviano de desplegamiento en dos estados debido al efecto de la fuerza que obedece a la ecuación 1 y donde la probabilidad de observar el desplegamiento de cualquier módulo

en cada paso es  $dP = N\alpha_F\Delta t$ . Donde  $N$  representa el número de módulos plegados,  $\alpha_F$  la tasa de desplegamiento (al aplicar fuerza) y  $\Delta t$  el intervalo de sondeo [Rief et al., 1998].  $\alpha_F$  se calcula a partir de la ecuación 1 (donde  $\alpha$  y  $\Delta x_D$  son los parámetros de interés) y la fuerza se calcula mediante el modelo WLC (asumiendo una  $p$  de 0.4 nm, y  $K_B T = 4.1$  pN·nm a temperatura ambiente y  $\Delta L_C$  según los datos experimentales). Se realizaron simulaciones a las mismas velocidades que en los experimentos empleando  $\Delta t$  entre ms y  $\mu$ s para distintos valores de  $\alpha$  y  $\Delta x_D$  y se determinaron los mejores valores comparando las fuerzas calculadas con las obtenidas experimentalmente (menor  $\chi^2$ ).

Parte IV

RESULTADOS

## RESULTADOS

---

Los resultados obtenidos durante la tesis se presentan en dos partes: la primera centrada en el análisis nanomecánico de módulos cohesina provenientes de varios celulosomas y la segunda en el estudio de la relación entre la estabilidad mecánica de los módulos cohesina y la actividad de los celulosomas de los que forman parte. Los primeros constituyen evidencias indirectas de la hipótesis mecánica, mientras que los segundos proveen evidencias directas al respecto.

### 4.1 NANOMECAÁNICA DE COHESINAS INDIVIDUALES

El trabajo de esta tesis se inició con la caracterización de módulos cohesina individuales. En concreto se trabajó con componentes de los celulosomas de *C. thermocellum*, *C. cellulolyticum* y *A. cellulolyticus* con el objetivo de profundizar en el conocimiento del comportamiento mecánico de estos, así como estudiar la generalidad de la hipótesis mecánica al trabajar con celulosomas muy diversos.

#### 4.1.1 Nanomecánica de cohesinas de *C. thermocellum*

Como se ha comentado anteriormente (sección 1.7), el módulo cohesina 2 de CipA de *C. thermocellum* (CtA2), perteneciente a la región externa, presenta una estabilidad mecánica inferior a la del módulo 7 (CtA7), de la región conectora [Valbuena et al., 2009]. Esto es coherente en el marco de la hipótesis mecánica del celulosoma que es la hipótesis de trabajo asumida en esta tesis. Estos módulos se escogieron por tratarse de los únicos de los que se disponía de estructura cristalográfica para esa escafoldina. La primera pregunta que nos planteamos abordar en esta tesis es si las propiedades descritas para estas cohesinas específicas son representativas de las cohesinas de cada región de la escafoldina como predice la hipótesis de trabajo.

Los módulos cohesina de CipA presentan un alto grado de identidad [Gerngross et al., 1993], especialmente entre las cohesinas de la región conectora [Valbuena et al., 2009]. En concreto, los módulos 3 a 8 presentan un 92 % de residuos idénticos mientras que el módulo 9 comparte solo un 78 % de sus residuos con el módulo 8 (Figura 12). Dada la elevada conservación de los módulos 3 a 8, es de esperar que su comportamiento mecánico sea muy parecido al de CtA7, por lo que se puede asumir que el comportamiento ya descrito probablemente sea representativo de todos estos módulos. Sin embargo, parece interesante estudiar el módulo 9 (CtA9) pues se encuentra en la región conectora y su estabilidad mecánica podría ser muy diferente a la de CtA7 ya estudiado. Por otro lado, el módulo 1 (CtA1) y CtA2, de la región externa, presentan



una identidad bastante menor (especialmente CtA1) por lo que también es interesante su estudio.

CtA1	1	ATMTVEIGKVTAAVCSKVEITPHTLKGVPESKGMANCDFVLGYDPNVLEIVTEVKPGSITIKDPDPSKSFDSATVPDRKMI
CtA2	1	-GVVVEIGKVTGSGVCTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEITIGTDPGDIIVDPNFKSFDTAIYPDRKII
CtA3	1	NAIKIKVDTVNAKPGDTRIPVRFSGIPSKGIANCDFVYSYDPNVLEIIEIEIPGDIIVDPNPKSFDTAIYPDRKII
CtA4	1	-AVRIKVDTVNAKPGDTRIPVRFSGIPSKGIANCDFVYSYDPNVLEIIEIEIPGDIIVDPNPKSFDTAIYPDRKII
CtA5	1	-AVRIKVDTVNAKPGDTRIPVRFSGIPSKGIANCDFVYSYDPNVLEIIEIEIPGDIIVDPNPKSFDTAIYPDRKII
CtA6	1	-AVRIKVDTVNAKPGDTRIPVRFSGIPSKGIANCDFVYSYDPNVLEIIEIEIPGDIIVDPNPKSFDTAIYPDRKII
CtA7	1	-AVRIKVDTVNAKPGDTRIPVRFSGIPSKGIANCDFVYSYDPNVLEIIEIEIPGDIIVDPNFKSFDTAIYPDRKMI
CtA8	1	-AVRIKVDTVNAKPGDTRIPVRFSGIPSKGIANCDFVYSYDPNVLEIIEIEIPGDIIVDPNPKSFDTAIYPDRKII
CtA9	1	-KLTTEIKIGRAEGRPGDTRIPVNLVGVPOKGIASGDFVYSYDPNVLEIIEIEIPGDIIVDPNFKSFDTAIYPDRKMI
CtA1	78	VFLFAEDSGRCITYAITQDGVFATIVAKVKSAAAAP--ITLLEVGAFADNDLVEISTTFVAGGVNLG
CtA2	77	VFLFAEDSGTGAYAITKDGVFARIIRATVKSAPG--YITFDEVGGFADNDLVEQKVSFIIDGGVNVG
CtA3	78	VFLFAEDSGTGAYAITKDGVFATIVAKVKSAPNGLSVIKFVEVGGFANNDLVEQRTQFFDGGVNVG
CtA4	77	VFLFAEDSGTGAYAITKDGVFATIVAKVKSAPNGLSVIKFVEVGGFANNDLVEQRTQFFDGGVNVG
CtA5	77	VFLFAEDSGTGAYAITKDGVFATIVAKVKSAPNGLSVIKFVEVGGFANNDLVEQRTQFFDGGVNVG
CtA6	77	VFLFAEDSGTGAYAITKDGVFATIVAKVKSAPNGLSVIKFVEVGGFANNDLVEQRTQFFDGGVNVG
CtA7	77	VFLFAEDSGTGAYAITEDGVFATIVAKVKSAPNGLSVIKFVEVGGFANNDLVEQRTQFFDGGVNVG
CtA8	77	VFLFAEDSGTGAYAITKDGVFATIVAKVKSAPNGLSVIKFVEVGGFANNDLVEQRTQFFDGGVNVG
CtA9	77	VFLFAEDSGTGAYAITEDGVFATIVAKVKSAPNGLSVIKFVEVGGFANNDLVEVEITDLINGGVNVG

Figura 12: Alineamiento de las secuencias de los módulos cohesina de CipA de *C. thermocellum*. La línea azul separa los módulos de la región externa (1 y 2) de los de la región conectora (3 a 9). CtA1, CtA2 y CtA9 son los módulos que presentan más diferencias comparado con el resto de secuencias, que presentan un elevado grado de identidad entre ellas.

Para examinar experimentalmente la estabilidad mecánica de los módulos cohesina de CipA se generaron dos construcciones similares a las descritas en [Valbuena et al., 2009] que contenían CtA1 (pAFM-CtA1) o CtA9 (pAFM-CtA9) flanqueados por 3 y 2 módulos I27 marcadores (Figura 13.D).

El análisis de SMFS mediante AFM se realizó a 400 nm/s, con el fin de obtener resultados comparables con los ya publicados [Valbuena et al., 2009]. Los resultados obtenidos se resumen en la Figura 13.

Se puede observar que la estabilidad mecánica observada para CtA1 fue incluso menor ( $124 \pm 25$  pN) que la de CtA2 ( $214 \pm 41$  pN) ya estudiado a la misma velocidad. Por otro lado, CtA9 mostró una estabilidad mecánica muy similar a la de CtA7 ( $480 \pm 77$  pN) y claramente mucho mayor que la de las cohesinas de la región externa. Estos resultados indicaron que a pesar de las diferencias en secuencia, la estabilidad mecánica de los módulos cohesina de la región conectora de CipA es claramente superior a la medida para los de la región externa, resultado que está plenamente de acuerdo con la hipótesis de trabajo.

La distribución de  $\Delta L_C$  de estos módulos (Figura 13.C) indicó que todas las cohesinas de CipA liberaban, tras las rotura del módulo, una longitud de secuencia similar ( $49.2 \pm 1.3$  nm para CtA1 y  $49 \pm 0.9$  nm para CtA9). Esto indicó que muy probablemente el parche mecánico de estas cohesinas se encuentra en la misma posición ya que tienen estructuras muy similares. El  $\Delta L_C$  máximo que se puede esperar para una proteína se puede estimar a partir del  $\Delta L_C$  de un aminoácido [Ainavarapu et al., 2007] y el número de residuos presentes en la proteína. Dado que estos módulos cohesina tienen unos 140 residuos, el  $\Delta L_C$  de un aminoácido es de 0.4 nm y teniendo en cuenta que estos módulos plegados miden unos 4 nm, el máximo  $\Delta L_C$  esperado es de unos 52 nm. Este valor corresponde muy bien con los valores medidos lo que es indicativo de que el

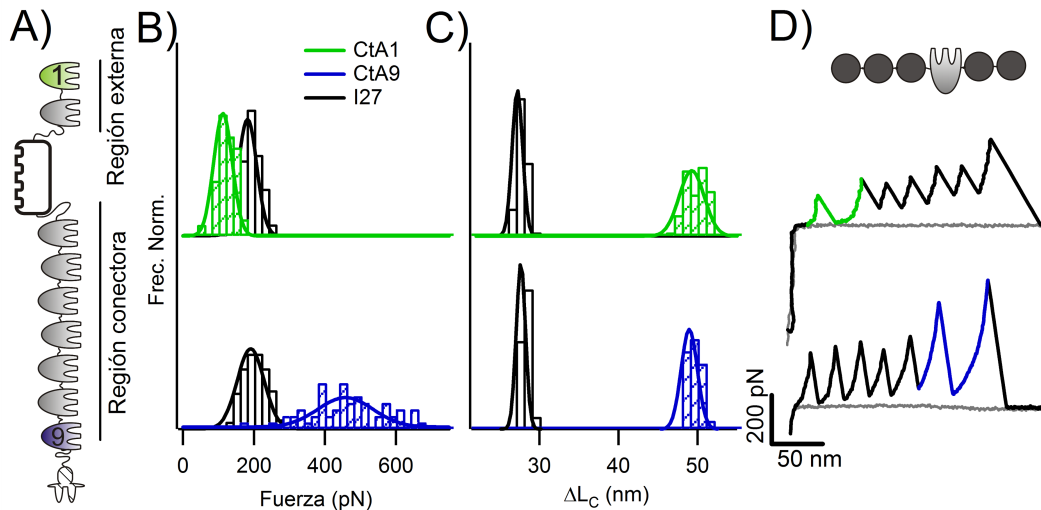


Figura 13: Resultados del análisis por SMFS de CtA1 y CtA9. A) Representación de la escafolina CipA donde se indican las regiones conectora y externa y se destacan en el correspondiente color los módulos estudiados. B) Distribución de fuerzas de desplegamiento observadas para CtA1 (arriba en verde,  $F_D = 124 \pm 25$  pN;  $n = 70$ ) y CtA9 (abajo en azul,  $F_D = 460 \pm 110$  pN;  $n=59$ ). También se muestran la distribución de fuerzas de desplegamiento para el marcador I27 obtenidos en cada caso ( $F_D = 192 \pm 27$  para CtA1;  $F_D = 208 \pm 43$  pN para CtA9). C) Histogramas de  $\Delta L_C$  liberados tras el desplegamiento de CtA1 ( $\Delta L_{C_{CtA1}} = 49.2 \pm 1.3$  nm,  $\Delta L_{C_{I27}} = 27.4 \pm 0.6$  nm, arriba en verde) y CtA9 ( $\Delta L_{C_{CtA9}} = 49 \pm 0.9$  nm,  $\Delta L_{C_{I27}} = 27.8 \pm 0.6$  nm, abajo en azul). D) Esquema de las construcciones empleadas en estos experimentos (arriba). Debajo, curvas fuerza-extensión obtenidas en experimentos con cada construcción. El pico correspondiente al desplegamiento de la cohesina se indica en su color correspondiente.

parche mecánico se encuentra cerca de los extremos del módulo cohesina, tal y como se había identificado para los módulos ya estudiados [Valbuena et al., 2009].

Por otra parte, se puede observar también (Figura 13.B y C) que los resultados obtenidos para el marcador I27 coinciden entre ellos y con los descritos previamente [Carrion-Vazquez et al., 1999b]. El módulo I27 se usó en estas construcciones como marcador para poder identificar las moléculas estiradas pero a la vez, dado que sus propiedades mecánicas son conocidas, se empleó como control interno de que las medidas eran comparables.

A continuación, para profundizar en la caracterización mecánica de estos módulos, se determinaron los parámetros cinéticos del desplegamiento mecánico de CtA1 y CtA9 realizando experimentos de SMFS a distintas velocidades (Figura 14.A) [Rief et al., 1997, 1998].

Se realizaron simulaciones de Monte Carlo variando los parámetros  $\alpha_0$  y  $\Delta x_D$  hasta obtener unos valores que reproducían las fuerzas medidas experimentalmente a cada velocidad (Figura 14.B). Los resultados obtenidos indicaron que  $\Delta x_D$  de CtA1 (0.31 nm) es similar al de I27 (0.25 nm) pero mayor que el de CtA9 (0.17 nm).

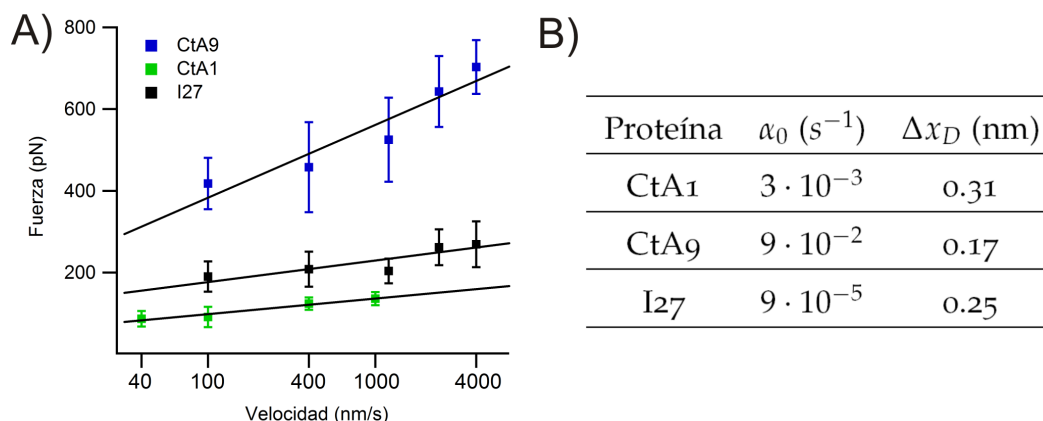


Figura 14: Determinación de los parámetros cinéticos del desplegamiento mecánico de CtA1 y CtA9. A) Fuerza de desplegamiento medida en función de la velocidad de estiramiento en los experimentos de SMFS para CtA1 (en verde), CtA9 (en azul) e I27 (en negro). Los datos de I27 corresponden a los obtenidos con los marcadores en la construcción pAFM-CtA9. Para CtA1 se empleó la construcción  $(CtA1)_8$  que contiene 8 repeticiones del módulo CtA1. B) Tabla resumen de los valores obtenidos para  $\alpha_0$  y  $\Delta x_D$  para cada módulo estudiado. Estos valores se estimaron mediante simulaciones de Monte Carlo (sección 3.8.1).

Para caracterizar el parche mecánico de estos dos módulos se han realizado simulaciones de SMD para determinar a nivel atómico los determinantes de la estabilidad mecánica de cada módulo. Para poder realizar las simulaciones de MD es necesario disponer de la estructura de la molécula [Amber, 2012]. La estructura cristalográfica de la cohesina 9 se había resuelto previamente junto a otros elementos del celulosoma de *C. thermocellum* (código PDB: 3KCP [Adams et al., 2010]) y para realizar las simulaciones se seleccionó la parte de la estructura correspondiente al módulo cohesina (residuos 9 a 149 de la cadena A).

Sin embargo, la estructura de cohesina 1 no está disponible en PDB por lo que para poder realizar el estudio mediante SMD se optó por emplear modelado por homología para obtener un modelo atómico de su estructura. Esta aproximación se basa en que la estructura de una proteína está más conservada durante su evolución que su secuencia por lo que en general, cuando 2 proteínas presentan más de un 40% de residuos idénticos, suelen presentar la misma estructura global [Eswar et al., 2006]. En el caso de las cohesinas es habitual encontrar identidades superiores al 40% [Gerngross et al., 1993] y además, todos los módulos cohesina cristalizados hasta el momento muestran que su plegamiento es bastante similar (sobre todo entre las del mismo tipo). Por todo ello el uso de modelado estructural por homología pareció muy adecuado para obtener la estructura de estos módulos.

Para estudiar la viabilidad de esta aproximación se usó MODELLER [Sali & Blundell, 1993] para calcular en primer lugar un modelo estructural de CtA9 empleando la estructura de CtA7 (código PDB: 1AOH [Tavares et al., 1997]) como molde ya que es con la que presenta mayor identidad (76.4%). La calidad global del modelo se estudió mediante el valor-Z obtenido en PROSAweb [Sippl, 1993, Wiederstein & Sippl, 2007],

que es una medida global de la calidad de un modelo. El valor obtenido fue de -6.15 que es un valor similar al calculado para estructuras de tamaño comparable resueltas experimentalmente. A continuación se empleó Procheck [Laskowski et al., 1993] para estudiar la calidad estereoquímica de la estructura generada. El análisis indicó que la mayoría de los residuos (94.8 %) se encontraban en conformaciones favorables mientras que el resto lo estaban en regiones permitidas. El modelo de CtA9 se comparó con la estructura cristalográfica observando que el RMSD entre los átomos pesados del esqueleto peptídico era de 0.71 Å, valor que está por debajo de la resolución de la estructura cristalográfica, indicando que ambas estructuras eran muy similares (Figura 15.A y B).

El modelo y la estructura cristalográfica de CtA9 se emplearon como punto de partida para realizar simulaciones de MD. Las estructuras se sometieron a varios pasos de preparación (ver sección 3.8) antes de realizar 10 ns de dinámica libre. El análisis del RMSD durante la dinámica libre muestra que ambas estructuras alcanzaron rápidamente una conformación estable (Figura 15.C). Por otro lado, el análisis de las fluctuaciones de cada  $C_{\alpha}$  (Figura 15.D) durante los últimos 5 ns de dinámica libre mostró que ambas estructuras presentaban un comportamiento muy similar: coincidían la posición de los máximos y mínimos en las dos estructuras y estos coincidían con los bucles y regiones de estructura secundaria de la proteína, respectivamente. La mayor diferencia se encontró sobre la posición 60 donde el modelo presenta mayores fluctuaciones que la estructura, sin embargo esta región está alejada del posible parche mecánico de la molécula por lo que es improbable que tuviese mucho efecto sobre este (Figura 15.B).

La estructura de cada módulo cada 2 ns de dinámica libre se usó como punto de partida para simular el estirado de la proteína como en el AFM mediante SMD. Las curvas fuerza-extensión del modelo y la estructura cristalográfica de la cohesina 9 resultaron muy parecidas entre ellas (Figura 15.E). Al principio del estirado apareció un pico de fuerza alto que coincidía con la rotura de los puentes de hidrógeno entre la primera y la última hebra  $\beta$  del módulo. Este pico representa la rotura del parche mecánico y aparece en una posición similar a la del pico observado en los experimentos de AFM por lo que se asume que describe este mismo proceso. Además, este coincidía con la posición de este elemento mecánico en las otras cohesinas descritas [Valbuena et al., 2009]. Se observó también la presencia de otros picos atribuidos a la rotura de intermedios aunque estos no se detectaron en los experimentos de AFM. En concreto, se observó un primer intermedio (segundo pico en la curva) cuando la extensión entre los extremos era de unos 15 nm en la que las hebras 1, 8 y 9 se han estirado y separado del resto del módulo que sigue plegado. A continuación, a una extensión cercana a 30 nm se observó un tercer pico originado al romper los contactos entre las hebras 3 y 7, y que está más definido en las curvas provenientes de la estructura cristalográfica que en las del modelo (Figura 16). Estos resultados indican que ambas estructuras se despliegan siguiendo una ruta similar.

La estabilidad mecánica (fuerza promedio de rotura del parche mecánico) del modelo ( $760 \pm 130$  pN) fue un poco mayor que la de la estructura cristalográfica ( $650 \pm 52$

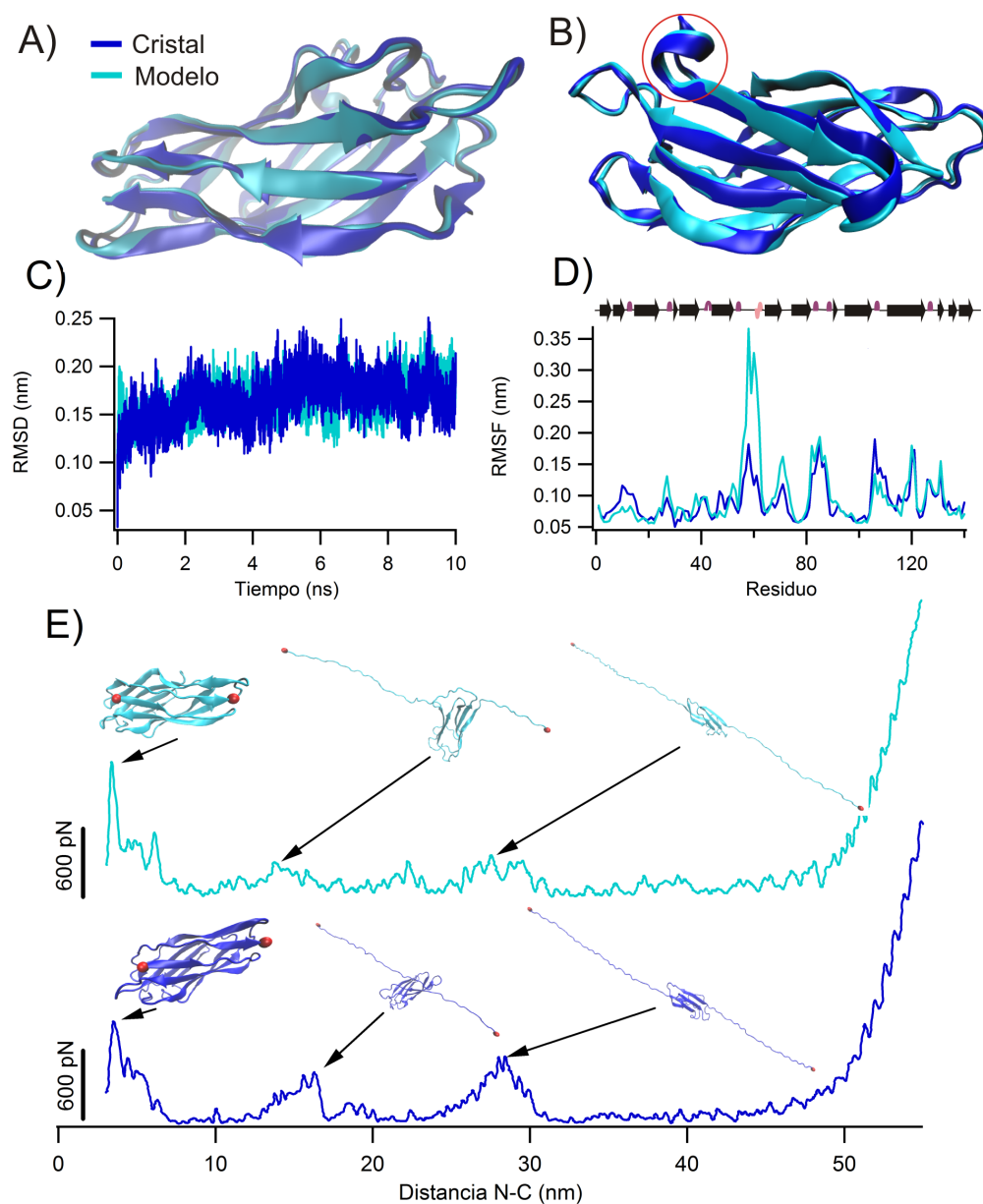


Figura 15: Validación del modelado molecular para la realización de simulaciones de SMD con cohesinas. A) La superposición de la estructura modelada (azul claro) y la estructura cristalográfica (azul oscuro) muestra que ambas estructuras son muy parecidas ( $\text{RMSD} = 0.71 \text{ \AA}$ ). B) Otra vista de la superposición de la estructura modelada y la cristalográfica. En rojo se indica el bucle de la posición 60. C) Evolución del RMSD del esqueleto peptídico durante la etapa de dinámica libre. Se puede observar que ambas estructuras alcanzan rápidamente una conformación estable. D) RMSF de los  $C_{\alpha}$  durante los últimos 5 ns de dinámica libre. Ambas estructuras presentan un patrón similar donde los mínimos coinciden con las zonas con estructura secundaria. E) Curva fuerza-extensión calculada a partir de una simulación del estirado de cada estructura tras 10 ns de dinámica libre. Se muestra la estructura en distintos puntos de la simulación correspondientes a los picos de fuerza. En ambas simulaciones la estructura presente en los distintos picos de fuerza son similares y presentan los mismos elementos de estructura secundaria formados. En rojo se representan los  $C_{\alpha}$  terminales, que corresponden a los puntos desde los que se estira la molécula.

pN) (Figura 16). Esto podría ser un reflejo de que algunas cadenas laterales se encontraban en conformaciones ligeramente distintas (el RMSD de todos los átomos era de 3.05 Å) lo que podría haber originado contactos adicionales. Las fuerzas obtenidas en las simulaciones fueron claramente más altas que las observadas en los experimentos de SMFS ya que la velocidad de estiramiento en las simulaciones era muy superior.

En conjunto, estos resultados indicaron que el modelo presentaba un comportamiento en simulaciones muy similar al de la estructura cristalográfica, incluyendo la presencia del parche mecánico y la misma ruta de desplegamiento, con lo que se validó la estimación de la estabilidad mecánica de módulos cohesina mediante esta aproximación.

El mismo proceso de SMD se llevó a cabo con el modelo de CtA1 empleando la estructura de CtA2 (código PDB: 1ANU [Shimon et al., 1997]) como molde, ya que es con la que presenta mayor identidad de secuencia (69%). El valor-Z para este modelo fue de  $-5.5$ , también muy parecido a los obtenidos para estructuras de este tamaño mientras que el análisis con Procheck indicó que el 90.5% de los residuos se encontraban en las regiones más favorables y el resto en permitidas. En la Figura 16 se muestran los resultados de las simulaciones de los estiramientos obtenidos para CtA1 y CtA9. En todas ellas se observó la presencia de un pico de fuerza al principio del proceso de estirado (flechas en la Figura 16) que coincidía en su posición con el pico observado en los experimentos. Para CtA1, se calculó una estabilidad mecánica de  $560 \pm 80$  pN, inferior a la calculada para CtA9. Como se puede observar, las estabilidades mecánicas obtenidas por SMD siguen un comportamiento similar al observado en los experimentos: las cohesinas de la región conectora presentan mayor estabilidad mecánica que las de la región externa, aunque en las simulaciones la estabilidad mecánica de CtA1 es mayor que la calculada para CtA2 ( $470 \pm 80$  pN) [Valbuena et al., 2009].

El análisis detallado de los determinantes moleculares del primer pico de fuerza en las tres estructuras indicó que este coincidía con la rotura de varios puentes de hidrógeno (CtA1: 8, CtA9 modelo: 7, CtA9 cristal: 8) entre residuos de las hebras 1 y 9, región que se identificó como el parche mecánico de estos módulos. Al estirar estos módulos, los puentes de hidrógeno del parche mecánico permitirían resistir la fuerza aplicada pero, una vez rotos estos puentes, las hebras 1 y 9 se desplazan una respecto a la otra y el módulo se abre quedando expuestos a la fuerza los residuos que antes estaban protegidos.

Finalmente, según la ecuación 2, es de esperar que la temperatura tenga un efecto sobre la estabilidad mecánica de la proteína. Las simulaciones anteriores se realizaron a 300 K, una temperatura similar a la de los experimentos que se hacen a temperatura ambiente. Sin embargo, *C. thermocellum* es un organismo termófilo con una temperatura óptima de crecimiento de unos 60°C ( $\sim 340$  K) [McBee, 1954], por lo que las simulaciones (dinámica libre y estiramientos) se repitieron a esa temperatura para obtener resultados más realistas. En la Figura 16 se puede observar que, como era de esperar, al aumentar la temperatura las fuerzas observadas fueron sistemáticamente inferiores pero la relación de estabilidades mecánicas se mantuvo, siendo los módulos de la región conectora de mayor estabilidad mecánica que los de la región externa.

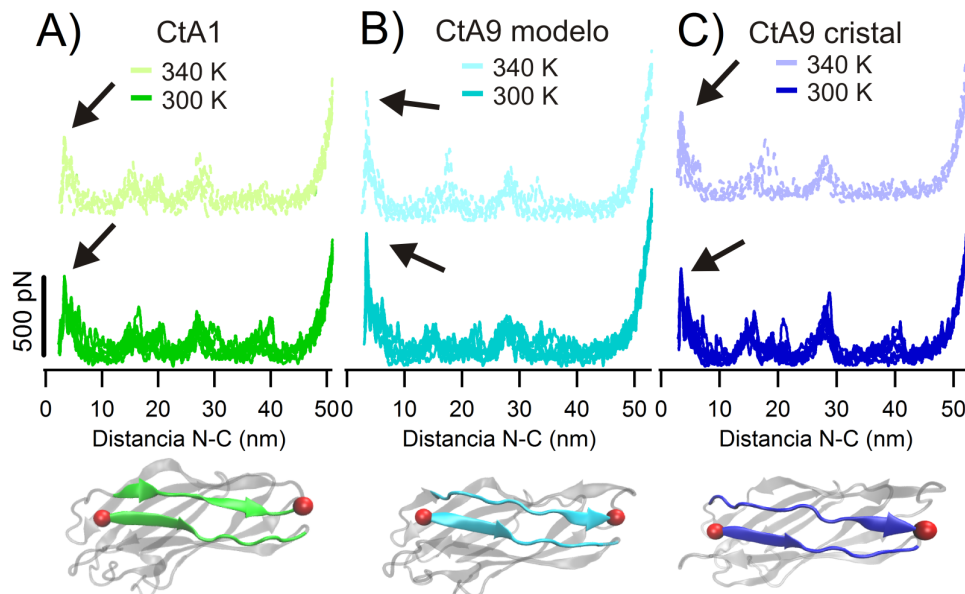


Figura 16: Simulaciones del estirado de cohesinas de *C. thermocellum*. Se muestran las 6 curvas fuerza-extensión calculadas a partir de las trayectorias de las simulaciones del estirado de cada estructura a 300 K (línea continua) y 340 K (línea discontinua) para A) CtA1 en verde, B) CtA9 (estructura modelada) en azul claro y C) CtA9 (estructura cristalográfica) en azul oscuro. La fuerza promedio de desplegamiento de CtA1 ( $560 \pm 80$  pN a 300 K y  $375 \pm 48$  pN a 340 K) era inferior a la de CtA9 tanto en la estructura cristalográfica ( $650 \pm 52$  pN a 300 K y  $497 \pm 87$  pN a 340 K) como en el modelo ( $760 \pm 130$  pN a 300 K y  $480 \pm 150$  pN a 340 K). Las fuerzas de desplegamiento se vieron reducidas en todos los casos al aumentar la temperatura aunque se mantuvo el orden relativo entre las distintas estructuras. En la parte inferior se muestra la estructura de cada módulo en el máximo del primer pico de fuerza donde se ha destacado la primera y última hebra de cada módulo, que componen el parche mecánico de estas proteínas.

#### 4.1.2 Nanomecánica de cohesinas de *C. cellulolyticum*

Para *C. cellulolyticum* no se ha identificado ningún elemento concreto que sugiera que su escafoldina esté unida a la célula [Ferdinand et al., 2013]. Esto descartaría *a priori* que sus cohesinas estén sometidas a estrés mecánico, sin embargo se ha comprobado que *C. cellulolyticum* se une a celulosa y que en mutantes deficientes en CipC esta unión está muy afectada [Ferdinand et al., 2013]. Por tanto, no se puede descartar que las cohesinas de CipC estén sometidas a estrés mecánico mediante algún tipo de interacción entre algún componente de superficie todavía no identificado.

El análisis de la secuencia de las cohesinas de CipC mostró, como en el caso de *C. thermocellum*, que la mayoría de módulos (1 a 7) son muy parecidos (más de 80% de residuos idénticos entre cualquier par de módulos) mientras que la última cohesina (módulo 8, CcC8) presenta una secuencia que difiere bastante de la del resto (Figura 17). Para estudiar la estabilidad mecánica de CcC8 se realizó una construcción ade-



```

CcC1 1 LKVTVGTANGKPGDVTVPVTFADVAKMKNVGTCNFYLCYDASLLEVVSVDDAGPIVKNAAVNFSS--SASNGTISFL
CcC2 1 LKVAVGTAANGKPGDVTVPVTFADVNVGNVGTTCNFYLAIDASLLEVVSVDDAGPIVKNAAVNFSS--SASNGTISFL
CcC3 1 LKVAVGTAANGKPGDVTVPVTFADVASAGNVGTTCNFYLAIDASLLEVVSVDDAGPIVKNAAVNFSS--SASNGTISFL
CcC4 1 LKVAVGTAANGKPGDVTVPVTFADVASAGNVGTTCNFYLAIDASLDDVVSVAAGPIVKNAAVNFSS--SASNGTISFL
CcC5 1 LKVAVGTAANGKPGDVTVPVTFADVASAGNVGTTCNFYLAIDASLDDVVSVAAGPIVKNAAVNFSS--SASNGTISFL
CcC6 1 LKVAVGTAANGKPGDVTVPVTFADVASAGNVGTTCNFYLAIDASLDDVVSVAAGPIVKNAAVNFSS--SASNGTISFL
CcC7 1 LKVAVGTAANGKPGDVTVPVTFADVATVGNVGTTCNFYLYDASLLEVVSVDDAGPIVKNAAVNFSS--SASNGTISFL
CcC8 1 LGVKTASVTGKTDGTTTVPVTLNSNVKSGNVGTTCNFYLYDASMLQAVSATAGDLVLNAPVNFSSSINATTTCTISIL

CcC1 76 FLDNTITDELITADGVFANIKFKLKSVAKTTPVTFKDGGAFFGDGTMKIASVTKTNGSVTIDPG 141
CcC2 76 FLDNTITDELITSDGVFANIKFKLKSVAKTTPVTFKDGGAFFGDGTMKIASVTKTNGSVTIDPG 141
CcC3 76 FLDNTITDELITADGVFANIKFKLKSVAKTTPVTFKDGGAFFGDGTMKIASVTKTNGSVTIDPG 141
CcC4 76 FLDNTITDELITADGVFANITFKLKSVAKTTPVTFKDGGAFFGDGTMKIASVTKTNGSVTIDPG 141
CcC5 76 FLDNTITDELITADGVFANITFKLKSVAKTTPVTFKDGGAFFGDGTMKIASVTKTNGSVTIDPG 141
CcC6 76 FLDNTITDELITADGVFANITFKLKSVAKTTPVTFKDGGAFFGDGTMKIASVTKTNGSVTIDPG 141
CcC7 76 FLDNTITDQLIKTDFEATFKLKSVAKTTPVTFKDGGAFFGDGTMKIASVTKTNGSVTIDPG 141
CcC8 78 FLDNTITDQLITSDGVFANITFKVVGTSSTTTPVTFKDGGAFFGDGTMKIASVTKTNGSVTIDPG 140

```

Figura 17: Alineamiento de las secuencias de los módulos cohesina de CipC de *C. cellulolyticum*. Se puede observar que CcC8 es la que presenta más diferencias (~60% respecto a cualquiera de las otras cohesinas) mientras que el resto presentan una identidad mucho mayor (~90% entre cualquiera de las cohesinas 1-7).

cuada para el análisis por SMFS en el vector pAFM (pAFM-CcC8, Figura 18.D), que se analizó en las mismas condiciones que los módulos anteriores (sección 4.1.1).

Las curvas de fuerza-extensión (Figura 18.D) mostraron un único pico atribuible a la cohesina que liberaba una longitud de  $48.1 \pm 0.8$  nm (Figura 18.C). Este comportamiento es muy similar a la de las cohesinas de *C. thermocellum*, sin embargo el histograma de fuerzas (Figura 18.B) indica que su estabilidad mecánica es inferior ( $346 \pm 58$  pN) a la de la cohesina 1 de la misma escafoldina (CcC1,  $425 \pm 53$  pN [Valbuena et al., 2009]), pero aún así superior a la de las cohesinas de la región externa de *C. thermocellum*.

La estructura de CcC8 tampoco está resuelta por lo que se empleó de nuevo el modelado por homología para obtener un modelo estructural de la misma. Se empleó como molde la estructura de CcC1 (código PDB: 1G1K [Spinelli et al., 2000]) ya que es la estructura conocida con la que presenta mayor identidad (60.7%). La calidad global del modelo se estudió como en los modelos de cohesinas de *C. thermocellum* analizando su valor-Z (-5.64, de nuevo similar a los valores obtenidos para proteínas de tamaño parecido) mientras que la calidad estereoquímica se evaluó con Procheck, que indicó que el 89.2% de los residuos se encontraban en regiones favorables y el 10% en regiones permitidas, con solo 1 residuo fuera de esta región. El RMSD de los átomos pesados del esqueleto peptídico del modelo durante la dinámica libre (Figura 18.E) aumenta rápidamente hasta alcanzar un valor estable indicando que se encuentra en un mínimo energético.

El análisis de las simulaciones de SMD mostró de nuevo la presencia de un pico al inicio del proceso de estirado, como en el caso de los experimentos, que se atribuyó a la rotura de 7 puentes de hidrógeno entre las hebras 1 y 9, identificado como el parche mecánico de la molécula. La estabilidad mecánica calculada en las simulaciones fue de  $568 \pm 86$  pN, similar al obtenido para CcC1 ( $630 \pm 70$  pN) [Valbuena et al., 2009]. Se observó la presencia de picos de fuerza correspondientes a la rotura de intermedios equivalentes a los observados en las simulaciones de SMD con módulos de *C. thermocellum*. Esto indica que, a pesar de las diferencias a nivel de secuencia entre las cohesinas de ambos organismos, las cohesinas estudiadas hasta el momento siguen una ruta de desplegamiento similar.



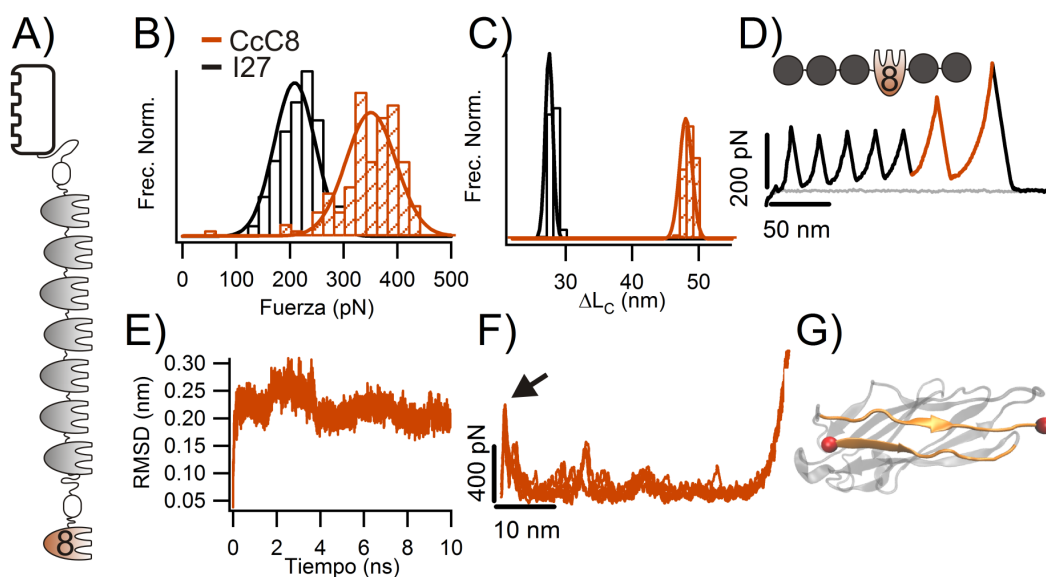


Figura 18: Resultados del estirado de CcC8. A) Esquema de la escafoldina CipC de *C. cellulolyticum* donde se resalta CcC8 en marrón. B) Distribución de fuerzas de desplegamiento observadas para el módulo CcC8 en marrón ( $346 \pm 58$  pN;  $n = 119$ ) y para el marcador I27 en negro ( $219 \pm 40$  pN). C) Distribución de  $\Delta L_C$  de CcC8 ( $48.1 \pm 0.8$  nm) y el marcador I27 ( $27.6 \pm 0.6$  nm). D) Esquema de la construcción empleada para el estudio y registro fuerza-extensión obtenido con la construcción. E) Evolución del RMSD del modelo estructural de CcC8 durante la dinámica libre. F) Superposición de 6 curvas fuerza-extensión calculadas a partir de las trayectorias de las simulaciones del estirado del modelo. La estabilidad mecánica calculada para el primer pico (marcado con una flecha) fue de  $568 \pm 86$  pN. G) Estructura de CcC8 en el máximo del primer pico de fuerza (flecha en F) durante una de las simulaciones de estirado. Se destaca la primera y última hebra del módulo que forman el parche mecánico de la molécula.

#### 4.1.3 Nanomecánica de cohesinas de *A. cellulolyticus*

Para completar el estudio de módulos cohesina individuales se decidió abordar el análisis nanomecánico de módulos cohesina de *A. cellulolyticus*. El celulosoma de este microorganismo se ancla a su superficie pero es más complejo que el de *C. thermocellum* ya que la escafoldina que une las enzimas, ScaA, se une a la escafoldina de superficie ScaC a través de una tercera escafoldina ScaB (Figura 5.B) [Hamberg et al., 2014]. En ScaA se distinguen de nuevo la región externa, que incluye las 3 cohesinas del N-terminal de la proteína, y la región conectora, que presenta 4 cohesinas. La identidad de secuencia entre las cohesinas de este organismo es muy variable (41-97%) [Dassa et al., 2012]. El alineamiento de estas secuencias mostró que, a diferencia de *C. thermocellum* las cohesinas de la región externa de ScaA presentan una elevada proporción de residuos idénticos ( $\sim 90\%$ ) mientras que las de la región conectora varían bastante entre si ( $\sim 60\%$ ) y respecto a las de la región externa. Se seleccionaron las cohesinas 3 y 4 (AcA3 y AcA4) como representantes de la región externa y conectora,

respectivamente, aunque en el caso de la región conectora hay otros módulos bastante diferentes (especialmente el 7).

La escafoldina ScaB está compuesta por módulos cohesina de tipo II, mientras que en ScaC son de tipo I [Dassa et al., 2012]. Se puede ver que al unirse el celulosoma al sustrato y a la célula, las cohesinas 2 a 4 y 2 a 3 de ScaB y ScaC respectivamente se encuentran entre los puntos de anclaje del sistema. La cohesina 1 de las dos escafoldinas también podría estarlo pero con una geometría distinta a las otras que se estirarían desde los extremos N- y C-terminales, como en el AFM. Para obtener información del comportamiento mecánico de las cohesinas de este celulosoma, se escogieron las cohesinas 4 de ScaB (AcB4) y 3 de ScaC (AcC3) para su análisis nanomecánico, por tratarse de las más conservadas dentro de la propia escafoldina.

Las cuatro cohesinas seleccionadas se clonaron en el vector pAFM y se procedió al análisis de las proteínas resultantes por SMFS. En todos los casos las cohesinas se desplegaron mostrando un único pico de fuerza (Figura 19.D), incluyendo la AcB4 que es la primera cohesina de tipo II que se estudia por SMFS. Como se puede observar en la Figura 19, AcA4 muestra una elevada estabilidad mecánica ( $442 \pm 49$  pN) similar a la otras cohesinas de la región conectora de CipA de *C. thermocellum* y claramente superior a la de AcA3 ( $209 \pm 54$  pN) que se encuentra en la región externa. Este resultado es similar al obtenido para las cohesinas de CipA de *C. thermocellum* y de nuevo está de acuerdo con la hipótesis mecánica planteada.

Sorprendentemente, el análisis nanomecánico de AcB4 reveló que esta presenta una estabilidad mecánica ( $259 \pm 39$  pN) comparable a la de AcA3 a pesar de encontrarse entre los puntos de anclaje, región en la que según la hipótesis de trabajo se esperaba encontrar módulos de alta estabilidad mecánica. Finalmente, AcC3 presenta una estabilidad mecánica elevada ( $455 \pm 51$  pN), similar a la de AcA4. Hay que destacar que para AcA4, AcB4 y AcC3 los resultados son aún preliminares ya que el número de datos es bajo para poder hacer una buena estimación de su estabilidad mecánica. Sin embargo, la distribución de fuerzas de desplegamiento en ambos casos está bastante centrada por lo que no es de esperar que al aumentar el número de moléculas analizadas el valor presentado cambie drásticamente.

#### 4.2 NANOMECAÁNICA DE FRAGMENTOS DEL CELULOSOMA

Los resultados presentados hasta el momento sobre la nanomecánica del celulosoma se limitan a dominios cohesina aislados y provenientes de distintas especies y escafoldinas. En un siguiente paso se planteó incluir en el estudio otros elementos presentes en el celulosoma y que podrían influenciar su respuesta mecánica. De esta forma se pretende obtener una visión más completa, lo más cercana posible al complejo natural.

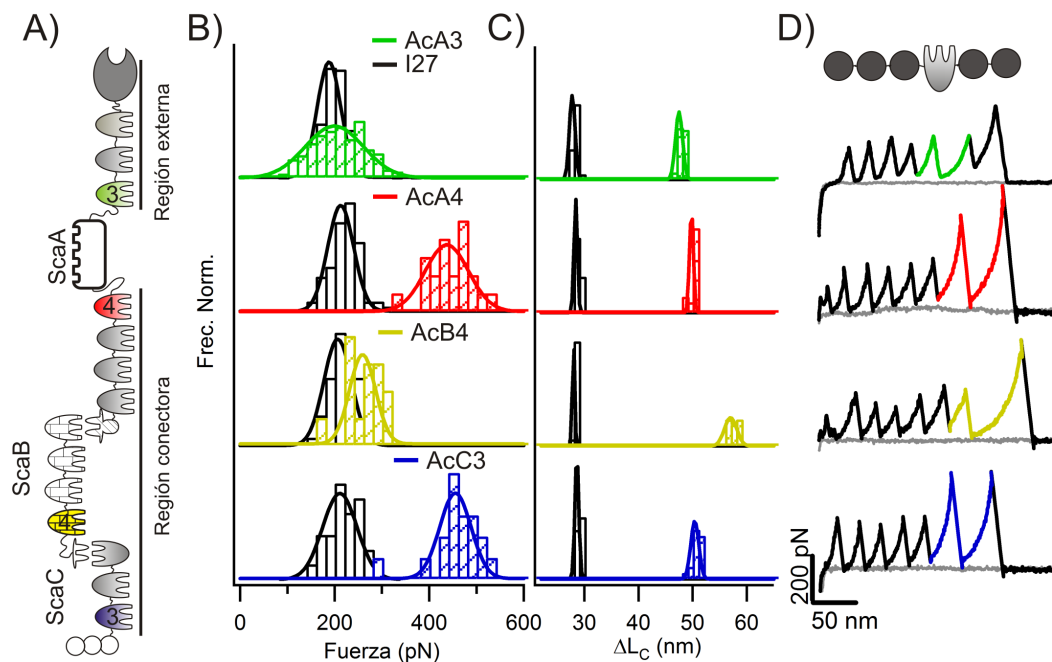


Figura 19: Análisis nanomecánico de varias cohesinas representativas de *A. cellulolyticus*. A) Representación de parte del celulosoma de *A. cellulolyticus* donde se muestran las escafoldinas ScaA, ScaB y ScaC en las que se destacan los módulos cohesina estudiados. B) Distribución de fuerzas de desplegamiento para AcA3 (arriba en verde,  $F_D = 209 \pm 54$  pN;  $n = 188$ ), AcA4 (en rojo,  $F_D = 442 \pm 49$  pN;  $n = 21$ ), AcB4 (en amarillo,  $F_D = 259 \pm 39$  pN;  $n = 14$ ) y AcC3 (en azul,  $F_D = 455 \pm 51$  pN;  $n = 18$ ). En cada histograma se incluye también la distribución de fuerzas de desplegamiento de los módulos marcadores correspondientes ( $F_{D_{AcA3}} = 197 \pm 31$  pN;  $F_{D_{AcA4}} = 220 \pm 31$  pN;  $F_{D_{AcB4}} = 215 \pm 27$  pN;  $F_{D_{AcC3}} = 218 \pm 32$  pN). C) Distribución de  $\Delta L_C$  para AcA3 ( $47.4 \pm 1.5$  nm), AcA4 ( $49.6 \pm 0.7$  nm), AcB4 ( $57 \pm 0.9$  nm), AcC3 ( $50.2 \pm 0.9$  nm) y de los correspondientes módulos marcadores ( $27.9 \pm 0.5$  nm;  $28.4 \pm 0.5$  nm;  $28 \pm 0.3$  nm;  $28.6 \pm 0.5$  nm). D) Esquema de la construcción empleada para el análisis nanomecánico de estas proteínas y ejemplos de curvas de fuerza-extensión obtenidas para cada una de ellas. El desplegamiento de cada módulo se indica con su color correspondiente.

#### 4.2.1 Nanomecánica de cohesinas con secuencias intermodulares

En [Valbuena et al., 2009] se había abordado por primera vez el estudio nanomecánico de componentes del celulosoma. En ese estudio, además de módulos cohesina independientes tanto de la escafoldina CipA de *C. thermocellum* como CipC de *C. cellulolyticum*, se realizaron experimentos de SMFS para medir la estabilidad mecánica de la región entre cohesina 3 y 8 de CipA (CtA3-8) y entre cohesina 1 y 7 de CipC (CcC1-7). Los resultados mostraban que en el caso de CipA la estabilidad mecánica de CtA7 ( $480 \pm 77$  pN) era similar a la observada promediando los distintos módulos cohesina de la región conectora de esta proteína (CtA3-8,  $495 \pm 108$  pN). Sin embargo, en el caso de CipC, se observó que la estabilidad mecánica promedio de CcC1-7 ( $592 \pm$

111 pN) era superior a la del módulo 1 (CcC1,  $425 \pm 53$  pN). Hay varias posibilidades que podrían explicar estos resultados:

- La estabilidad mecánica de CcC1 es inferior a la de la mayoría de cohesinas de esa escafoldina. Esto parece poco probable dado el elevado grado de identidad entre la secuencia de CcC1 y el resto de cohesinas de la escafoldina (Figura 17).
- A diferencia de la construcción con CcC1, en CcC1-7 se incluyen las secuencias intermodulares que separan cada módulo cohesina. Por un lado esto podría permitir la formación de nuevos contactos entre los residuos adicionales y los módulos adyacentes que estabilizasen a las cohesinas. Por otro lado, la presencia de múltiples repeticiones cohesina podría promover que se formen contactos entre estas que estabilizasen a las cohesinas.

Para indagar en el origen de estos resultados, se realizaron dos construcciones similares a las descritas anteriormente [Valbuena et al., 2009] y en los experimentos de las secciones 4.1.1, 4.1.2 y 4.1.3 pero incorporando esta vez las secuencias intermodulares que flanquean en ambos lados al módulo CtA7 y al módulo CcC1, generando las construcciones pAFM-CtA7I (Figura 20.A) y pAFM-CcC1I (Figura 21.A) que se analizaron por SMFS a una velocidad constante de 400 nm/s, de forma que los resultados fuesen comparables a los publicados [Valbuena et al., 2009].

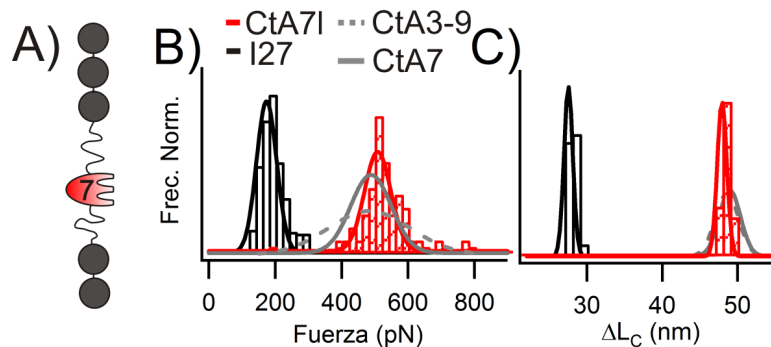


Figura 20: Análisis nanomecánico de la construcción pAFM-CtA7I. A) Esquema de la construcción empleada en los experimentos, que incluye además de CtA7 las secuencias intermodulares de 22 residuos en ambos extremos. B) Distribución de fuerzas de desplegamiento de CtA7I en rojo ( $529 \pm 82$  pN,  $n = 95$ ) y del marcador I27 en negro ( $193 \pm 39$  pN). En gris se incluye el ajuste a la distribución obtenida para CtA7 (línea continua) y CtA3-8 (línea discontinua) [Valbuena et al., 2009] como referencia para facilitar la comparación. C) Distribución de  $\Delta L_C$  para CtA7I ( $48.0 \pm 0.7$  nm) y el marcador I27 ( $27.6 \pm 0.6$  nm). Se concluyó que la presencia de las secuencias intermodulares no genera ningún efecto sobre el comportamiento mecánico de CtA7 que a su vez es similar al observado para la mayoría de los módulos de la región conectora de esta escafoldina.

Los resultados obtenidos (Figuras 20.B y 21.B) mostraron que la estabilidad mecánica de los dos módulos cohesina estudiados ( $529 \pm 82$  pN para CtA7I y  $398 \pm 47$  pN para CcC1I) no se ve afectada por la presencia de las secuencias intermodulares, resultando en ambos casos en distribuciones de fuerza de desplegamiento similares a

las de las respectivas cohesinas sin las secuencias intermodulares. Esto permitió descartar la formación de contactos adicionales entre los residuos añadidos y CcC1 como origen de la elevada estabilidad mecánica medida en CcC1-7. El  $\Delta L_C$  medido para los módulos con secuencias intermodulares ( $48 \pm 0.7$  nm para CtA7I y  $48.4 \pm 1.1$  nm para CcC1I) es muy similar al obtenido para los módulos sin ellas ( $49.3 \pm 0.3$  nm para CtA7 y  $48.5 \pm 0.2$  nm para CcC1 [Valbuena et al., 2009, Chwastyk et al., 2014]) lo que también reforzó la idea de que la posición del parche mecánico no se vio alterada en esta construcción.

Para estudiar la posibilidad de que otros módulos cohesina de CipC fuesen de mayor estabilidad mecánica o de que se formase una superestructura por contactos entre cohesinas de la misma escafoldina se prepararon dos construcciones adicionales: pAFM-CcC3I donde se usó el módulo cohesina 3 (CcC3I) para generar una construcción análoga a las de la sección 4.1.2 y pAFM-CcC123 donde se incluye la región que comprende las tres primeras cohesinas de CipC (CcC123) flanqueadas por dos módulos marcadores I27 a cada lado.

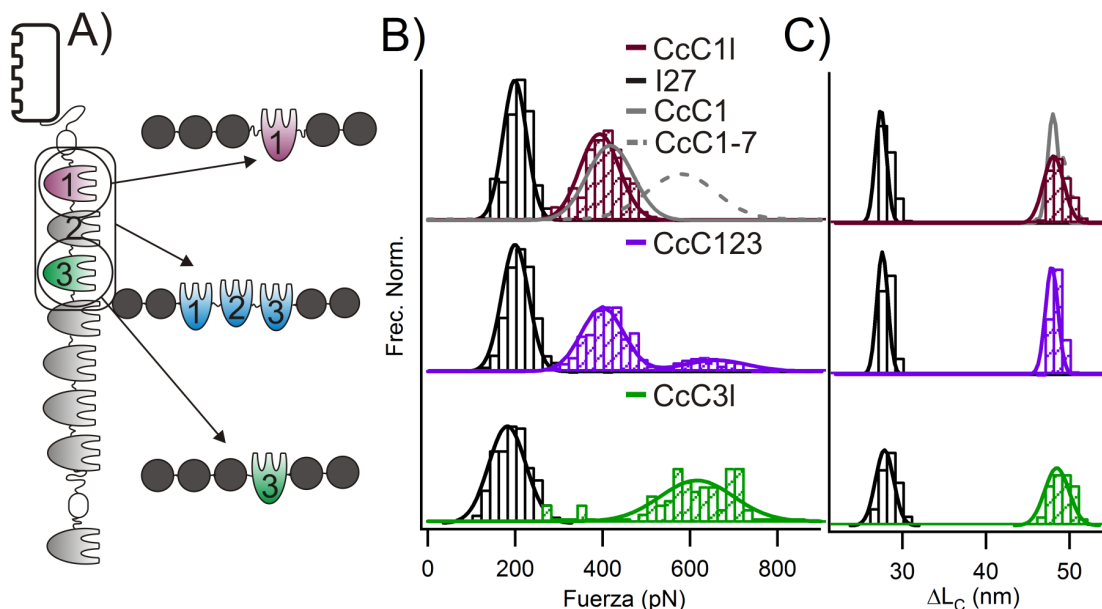


Figura 21: Estudios de nanomecánica de cohesinas de *C. cellulolyticum* en presencia de secuencias intermodulares. A) Esquema de la escafoldina CipC donde se resaltan los módulos 1 y 3. Se incluye también un esquema de las construcciones empleadas indicando en la misma Figura su origen. B) Distribución de fuerzas de desplegamiento de CcC11 (arriba en violeta,  $F_D = 398 \pm 47$  pN,  $n = 105$ ), de CcC123 (en el centro en azul, con dos picos centrados en  $404 \pm 64$  pN y  $627 \pm 79$  pN,  $n = 586$ ) y CcC3I (abajo en verde,  $F_D = 592 \pm 125$  pN,  $n = 110$ ) y para los marcadores I27 de cada construcción (en negro,  $F_{DCcC11} = 207 \pm 34$  pN;  $F_{DCcC123} = 212 \pm 35$  pN;  $F_{DCcC3I} = 192 \pm 39$  pN). Se incluye también el ajuste a la distribución de fuerzas observadas para CcC1 (línea continua gris  $F_D = 425 \pm 53$ ) y CcC1-7 (línea gris discontinua  $F_D = 592 \pm 111$  pN [Valbuena et al., 2009]) para facilitar la comparación. C) Distribución de  $\Delta L_C$  para las tres construcciones ( $48.4 \pm 1.1$  nm para CcC1I,  $48.5 \pm 1.3$  nm para CcC123I y  $48.5 \pm 1.3$  nm para CcC3I). Para el módulo marcador, se observó un  $\Delta L_C$  de  $27.8 \pm 0.7$  nm para CcC1I,  $27.8 \pm 0.6$  nm para CcC123I y  $28 \pm 1$  nm para CcC3I.

Como se puede observar en la Figura 21 el histograma de fuerzas de desplegamiento de los módulos cohesina de la construcción pAFM-CcC123 mostró dos poblaciones centradas en  $404 \pm 64$  pN y  $627 \pm 79$  pN. La estabilidad mecánica de la primera población es similar a la observada para CcC1 ( $425 \pm 53$  pN [Valbuena et al., 2009] o  $403 \pm 64$  pN [Chwastyk et al., 2014] y  $390 \pm 62$  pN para CcC11) de la misma escafoldina. Por otro lado en el histograma de fuerzas de pAFM-CcC31 (Figura 21.B) se pudo observar que este módulo presentaba una estabilidad mecánica muy elevada ( $620 \pm 118$  pN) valor muy similar a la fuerza promedio de la segunda población observada para CcC123 y a la obtenida para la región CcC1-7 ( $592 \pm 111$  pN [Valbuena et al., 2009]). Esto indicó que la estabilidad mecánica de cada módulo cohesina está determinada por el propio módulo y no depende de la presencia de secuencias intermodulares ni de otros módulos cohesina cercanos.

La distribución de  $\Delta L_C$  de CcC123 presentaba un solo pico centrado en  $48.5 \pm 1.3$  nm lo que indicó que el parche mecánico en los tres módulos de la construcción se encontraba en una posición similar ya que todos tienen una longitud parecida. Para CcC31 la distribución de  $\Delta L_C$  se centró en  $48.5 \pm 1.3$  nm, observándose un solo pico atribuible al desplegamiento de la cohesina.

La estructura del módulo 3 tampoco se ha resuelto experimentalmente por lo que se empleó el modelado por homología para obtener un modelo de esta. En este caso se empleó como molde la estructura del módulo 1 (código PDB: 1G1K) ya que es la cohesina con estructura resuelta con la que tiene mayor identidad (94%). El análisis en PROSAweb indicó un valor-Z de -5.21, adecuado para una estructura de esta longitud, mientras que el análisis mediante Procheck reveló que la mayoría de los residuos (94.2%) se encontraban en regiones favorables, mientras que un 5% se encontraba en regiones permitidas. El análisis del RMSD durante los 10 ns de dinámica libre (Figura 22.A) mostró que la molécula alcanzaba una conformación estable que se mantenía durante el tiempo de simulación. Los estiramientos mediante SMD permitieron calcular que la estabilidad mecánica de este módulo ( $733 \pm 133$  pN) es comparable a la observada para el módulo 1 ya descrito ( $630 \pm 70$  pN) [Valbuena et al., 2009]. El pico de fuerza inicial, atribuido a la rotura del parche mecánico de este módulo, coincide con la pérdida de 8 puentes de hidrógeno entre las cadenas 1 y 9, como en las otras cohesinas estudiadas.

Los resultados obtenidos indicaron que la mayor estabilidad mecánica observada para la región de cohesinas 1 a 7 de CipC comparada con la del módulo 1 era debida a la presencia de módulos con mayor estabilidad mecánica que este. Además los resultados indican que los principales determinantes de la estabilidad mecánica en una escafoldina son sus módulos cohesina, mientras que las secuencias intermodulares parecen no tener efecto.

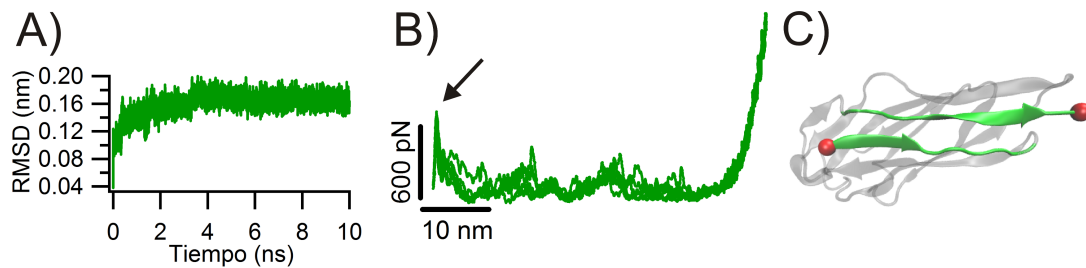


Figura 22: Resultados de simulaciones del estirado de CcC3. A) Evolución del RMSD de los átomos del esqueleto peptídico durante la dinámica libre. B) Superposición de las 6 curvas fuerza-extensión obtenidas a partir de las trayectorias de SMD. La estabilidad mecánica calculada fue de  $733 \pm 133$  pN, referido al primer pico, marcado con una flecha. Se observa la presencia de intermedios en el desplegamiento mecánico de la molécula siendo el que se encuentra cuando la distancia N-C terminal es de unos 25 nm el más claro. C) Imagen de la estructura en el máximo del primer pico de fuerza. Se destacan la primera y última hebra que forman el parche mecánico de la molécula.

#### 4.2.2 Nanomecánica de cohesinas unidas a doquerina

La interacción proteína-ligando es un factor que suele modificar significativamente la estabilidad termodinámica de las proteínas. A pesar de que la estabilidad termodinámica y la mecánica no tienen porqué estar relacionadas en absoluto, incluyendo el caso de las cohesinas [Valbuena et al., 2009], se ha observado que la interacción de una proteína con su ligando (incluso en el caso de moléculas pequeñas) también puede modificar su estabilidad mecánica [Junker et al., 2009, Oroz et al., 2011]. En el caso del celulosoma, se ha visto que algunas interacciones cohesina-doquerina pueden aumentar la estabilidad termodinámica de los componentes individuales [Salama-Alber et al., 2013]. En base a esto decidimos estudiar si la unión de doquerina pudiera afectar a la estabilidad mecánica de los módulos cohesina. Es importante resaltar que el dominio doquerina se une a cohesina a través de la cara formada por las hebras 8-3-6-5 [Carvalho et al., 2003], mientras que el parche mecánico está formado por las hebras 1 y 9 que forman parte de la otra hoja  $\beta$  (secciones 4.1 y 4.2.1 y [Valbuena et al., 2009]). Sin embargo, se ha observado que en el caso de la proteína GB1, capaz de unir la región constante de anticuerpos por una región de la proteína alejada de su parche mecánico, la interacción con este ligando puede aumentar su estabilidad mecánica [Cao et al., 2007, 2008].

Para estudiar el posible efecto de la unión a doquerina sobre la estabilidad mecánica de cohesina se realizaron experimentos de SMFS empleando las construcciones pAFM-CtA2 y pAFM-CtA7 ya descritas [Valbuena et al., 2009] en presencia de la endocelulasa Cel8A de *C. thermocellum*. En primer lugar se comprobó que las cohesinas presentes en las construcciones retenían su capacidad de interactuar con la doquerina de la celulosa, a pesar de estar flanqueadas por los módulos marcadores I27. Para ello se realizó un análisis mediante PAGE-nativa con muestras de ambas proteínas incubadas



previamente a 37°C durante 1 h en tampón de interacción (Figura 23), confirmando que la interacción se mantiene.

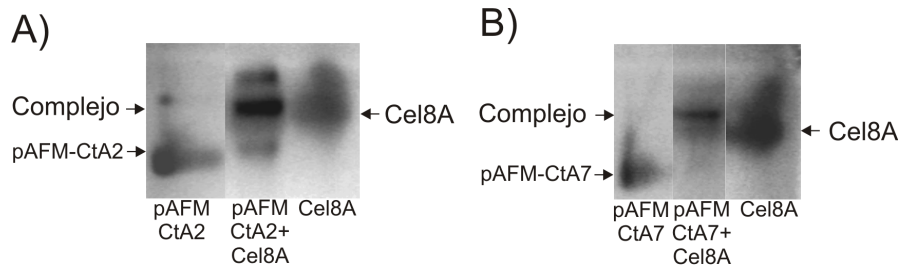


Figura 23: Control de la interacción entre la doquerina de Cel8A y las cohesinas en la construcción para análisis nanomecánico. A) Gel nativo donde se muestra que al aumentar la concentración de la enzima Cel8A aparece una nueva banda atribuida al complejo pAFM-CtA2: Cel8A de forma concomitante a la desaparición de la banda de pAFM-CtA2. B) Experimento análogo empleando pAFM-CtA7.

Tras confirmar que la interacción cohesina-doquerina se mantenía en la construcción para SMFS, se realizó el análisis nanomecánico de las dos construcciones en presencia de Cel8A en exceso molar a 400 nm/s en el tampón de interacción. Para poder realizar una comparación directa, también se estudió la estabilidad mecánica de las dos construcciones en el mismo tampón pero en ausencia de Cel8A. Los resultados se muestran en la Figura 24 donde se puede observar que la estabilidad mecánica de ambas cohesinas se mantuvo inalterada al añadir la celulasa Cel8A en la construcción, tanto para la cohesina de la región externa CtA2 ( $179 \pm 33$  pN para CtA2 sola y  $195 \pm 42$  pN en presencia de Cel8A) como en la cohesina de la región conectora CtA7 ( $515 \pm 69$  pN para CtA7 sola y  $506 \pm 87$  pN en presencia de Cel8A). El histograma de  $\Delta L_C$  era idéntico en presencia y ausencia de Cel8A para ambas construcciones (24.C y 24.E) lo que sugirió que no ha hubo cambios en el parche mecánico de ninguna de las cohesinas y por tanto que la estabilidad mecánica de las cohesinas en el celulosoma está determinada únicamente por la estabilidad del propio módulo.

Para profundizar en los detalles del proceso de desplegamiento mecánico de los módulos cohesina en presencia de doquerina y estudiar en qué momento durante el proceso de estiramiento se pierde la interacción se realizaron simulaciones de MD. Como punto de partida se empleó la estructura de un complejo cohesina-doquerina resuelto experimentalmente (código PDB: 1OHZ) [Carvalho et al., 2003] y también el modelo de CtA7 interaccionando con la misma doquerina. Se calcularon 10 ns de dinámica libre para ambos complejos y a continuación se realizaron simulaciones del estirado de las cohesinas en presencia de las doquerinas mediante SMD. En la Figura 25 se muestra la superposición de las curvas de fuerza-extensión calculadas a partir de las trayectorias.

En general se puede observar que las curvas son similares en ambas construcciones, mostrando intermedios de desplegamiento mecánico en posiciones similares. La estabilidad mecánica calculada ( $489 \pm 451$  pN para 1OHZ y  $621 \pm 60$  pN para CtA7:doquerina) son similares a las obtenidas con las respectivas cohesinas sin la doquerina en simulaciones realizadas en las mismas condiciones ( $470 \pm 80$  pN para CtA2



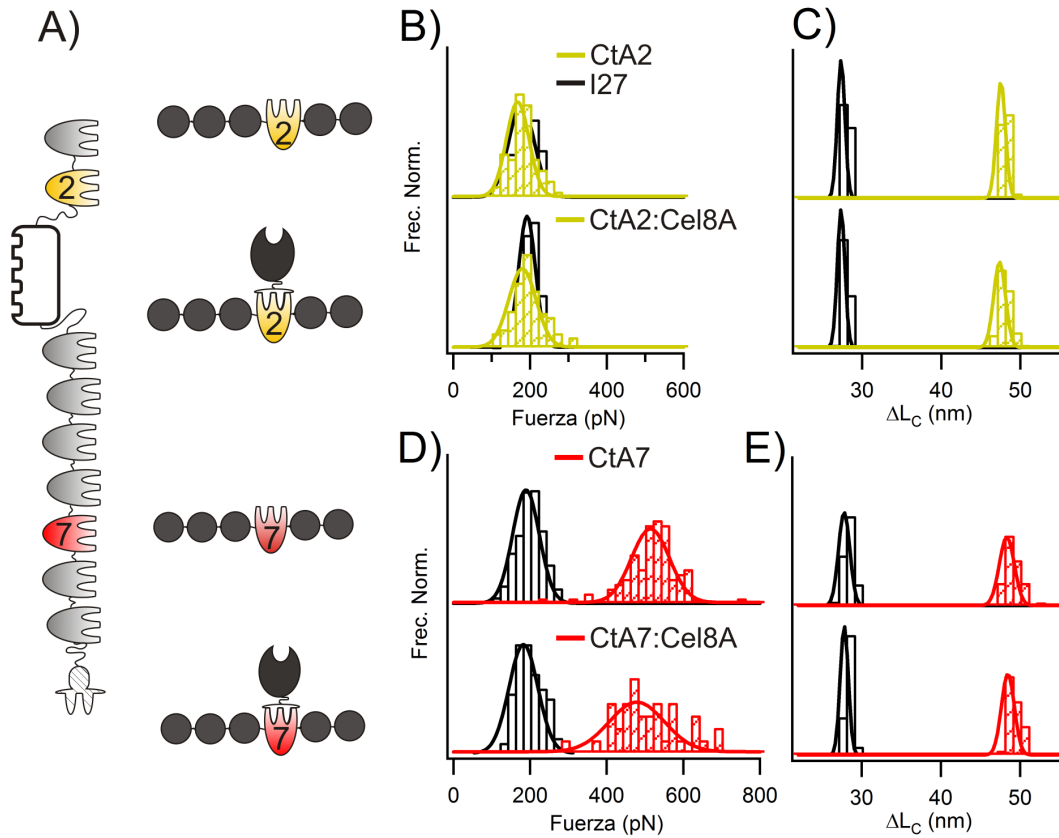


Figura 24: Resultados del estudio de nanomecánica de las cohesinas CtA2 y CtA7 en presencia de Cel8A. A) Representación de la escafoldina CipA de *C. thermocellum* donde se destacan los módulos estudiados y se muestra un esquema de las construcciones empleadas. B) En amarillo, distribución de fuerzas de desplegamiento para CtA2 sola (arriba,  $179 \pm 133$  pN,  $n = 69$ ) y en presencia de Cel8A en exceso ( $195 \pm 42$  pN,  $n = 185$ ). C) Distribución de  $\Delta L_C$  para las mismas construcciones ( $47.6 \pm 0.6$  nm para la cohesina sola y  $47.5 \pm 0.8$  nm en presencia de Cel8A). D) como en B) pero para CtA7 (en rojo arriba,  $515 \pm 59$  pN,  $n = 102$  para la cohesina sola y  $506 \pm 87$  pN,  $n = 40$  en presencia de Cel8A abajo). E) como en C) para CtA7 (en rojo arriba,  $48.6 \pm 1.1$  nm para la cohesina sola y  $48.7 \pm 0.8$  nm en presencia de Cel8A abajo). En negro se muestran los resultados obtenidos para los marcadores I27 de cada construcción:  $I27_{CtA2}$   $189 \pm 31$  pN,  $27.4 \pm 0.5$  nm;  $I27_{CtA2: Cel8A}$   $202 \pm 25$  pN,  $27.4 \pm 0.5$  nm;  $I27_{CtA7}$   $198 \pm 34$  pN,  $27.9 \pm 0.7$  nm;  $I27_{CtA7: Cel8A}$   $197 \pm 34$  pN,  $27.9 \pm 0.5$  nm.

(1ANU) y  $710 \pm 120$  pN para CtA7 (1AOH)). El análisis más detallado de las trayectorias de desplegamiento permitió observar que la doquerina perdía la interacción específica (pérdida de la superficie con la que contacta en la estructura nativa) con el desplegamiento del intermedio que se observa cuando la extensión N-C es de unos 30 nm en las curvas de fuerza-extensión. No se observaron diferencias significativas en las trayectorias de desplegamiento lo que indica que en presencia de doquerina, la cohesina sigue una ruta de desplegamiento similar.

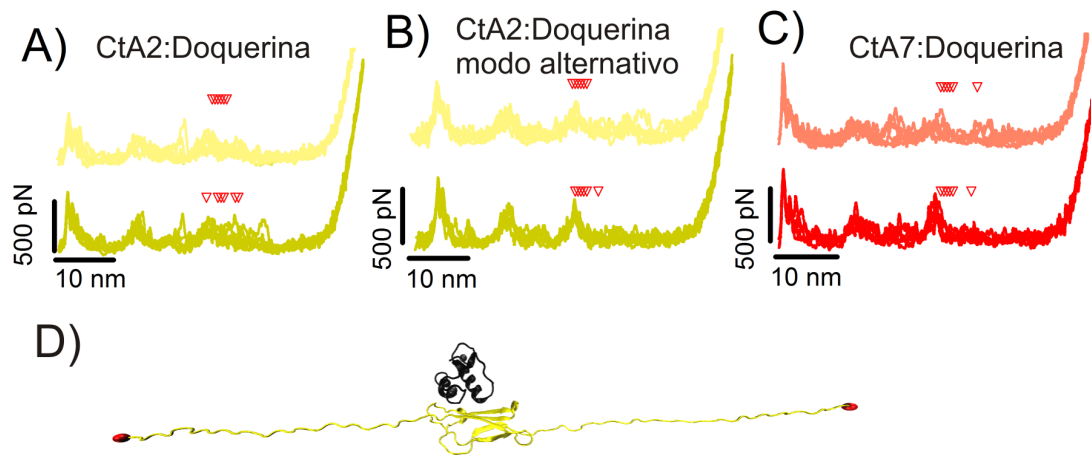


Figura 25: Resultados de las simulaciones del estirado de módulos cohesina en presencia de doquerina de Xyn10B. A) Superposición de curvas de fuerza-extensión calculadas a partir de las trayectorias del estirado de CtA2 en presencia de doquerina de Xyn10B empleando las coordenadas depositadas en PDB con el código 1OHZ a 300 K (abajo) y 340 K (arriba). Los triángulos indican el momento de la simulación en la que se pierde la interacción específica entre el módulo cohesina y el dominio doquerina. La estabilidad mecánica calculada para el primer pico es de  $489 \pm 51$  pN a 300 K y  $375 \pm 72$  a 340 K. B) Como en A) pero empleando la estructura con código PDB 2CCL que incluye CtA2 interaccionando con un doble mutante de la doquerina Xyn10B dispuesta en el modo de unión alternativo. La estabilidad mecánica calculada es de  $476 \pm 91$  pN a 300 K y  $456 \pm 63$  pN a 340 K. C) Como en A) pero empleando la estructura generada al sustituir CtA7 por CtA2 en 1OHZ. La estabilidad mecánica en este caso es de  $621 \pm 60$  pN a 300 K y  $567 \pm 105$  pN a 340 K. D) Muestra de la estructura presente justo antes de que se pierda la interacción específica entre la cohesina y la doquerina en un estirado de 1OHZ. Esta estructura es representativa de la que se observa en el resto de simulaciones y en las otras estructuras.

Como se ha comentado, algunas doquerinas son capaces de unirse a cohesina de dos modos distintos (rotados  $180^\circ$  uno respecto al otro). Para estudiar esta configuración, se realizaron simulaciones del estirado de cohesina 2 en presencia de doquerina unida en el modo alternativo (código PDB: 2CCL) [Carvalho et al., 2007]. Los resultados obtenidos en estas simulaciones (Figura 25.B) fueron muy similares a los anteriores, y se observó que la estabilidad mecánica calculada ( $476 \pm 91$  pN) era muy similar tanto a la de la cohesina sola ( $470 \pm 80$  pN) como en presencia de doquerina en el modo de unión habitual ( $489 \pm 51$  pN). Además se observó un patrón de desplegamiento muy similar al observado en las otras simulaciones lo que sugiere que la unión a

doquerina, independientemente de la orientación, no afecta a la estabilidad mecánica de la cohesina a la que se une.

Finalmente, se realizaron las mismas simulaciones a 340 K, temperatura óptima de crecimiento del *C. thermocellum*, y se observó en todos los casos un comportamiento similar al descrito a 300 K (Figura 25). Las estabilidades mecánicas calculadas, sin embargo, fueron más bajas en los tres casos, como cabía esperar.

Parte V

DISCUSIÓN

## DISCUSIÓN

---

El análisis nanomecánico de módulos cohesina individuales se ha centrado en los celulosomas de *C. thermocellum*, *C. cellulolyticum* y *A. cellulolyticus*. Tanto en los experimentos de SMFS como en las simulaciones por SMD la dirección en la que se ha aplicado fuerza a los módulos cohesina ha sido desde sus extremos terminales en sentidos opuestos. Esta es la misma geometría en la que se piensa que puede actuar la fuerza en el sistema natural, por lo que los resultados obtenidos deberían reflejar las propiedades del sistema en condiciones similares a las fisiológicas.

Al comparar los resultados experimentales obtenidos mediante SMFS con las simulaciones de SMD se ha identificado que el pico de fuerza observado por AFM corresponde con la rotura de varios puentes de hidrógeno establecidos entre la primera y última hebra  $\beta$  de los módulos cohesina. Estos elementos de estructura secundaria forman el parche mecánico del módulo, es decir, la principal región responsable de resistir fuerza [Carrion-Vazquez et al., 1999a, Valbuena et al., 2009]. Como se ha comentado, en la estructura de los módulos cohesina la primera y última hebra  $\beta$  están dispuestas en paralelo de forma que los extremos N- y C-terminales se encuentran en lados opuestos de la molécula. Por lo tanto al estirar desde los extremos se produce una fuerza en cizalla sobre el parche mecánico, donde los puentes de hidrógeno se encuentran perpendiculares a la dirección de la fuerza. Esto impone la necesidad de romper varios puentes de hidrógeno a la vez antes de poder desplegar el módulo completo, lo que se traduce en fuerzas de desplegamiento relativamente altas [Brockwell et al., 2003, Carrion-Vazquez et al., 2003, Valbuena et al., 2009]. Esta geometría se ha descrito para el parche mecánico de otras proteínas sometidas a estrés mecánico [Carrion-Vazquez et al., 1999b, Sułkowska & Cieplak, 2008, Valbuena et al., 2009] y generalmente proporciona una mayor resistencia mecánica en comparación con el parche en cremallera (donde los puentes de hidrógeno se encuentran paralelos a la dirección de la fuerza) [Carrion-Vazquez et al., 2000], o con proteínas donde el parche mecánico se compone de hélices  $\alpha$ , que suelen desplegarse a fuerzas bajas debido a la diferente geometría de los puentes de hidrógeno en esta estructura [Rief et al., 1999, Bustamante et al., 2004].

La diferencia entre la estabilidad mecánica observada en simulaciones de SMD y en los experimentos de SMFS se debe principalmente a la diferencia de velocidades de estiramiento empleadas en ambos casos, que implican tasas de carga diferentes [Evans & Ritchie, 1997, Lee et al., 2009]. Aún así, las fuerzas de desplegamiento relativas calculadas mediante las simulaciones correlacionan muy bien con las medidas experimentalmente (Figura 26.A), incluso en los modelos estructurales empleados. Esto indica que a pesar de que los valores de fuerza obtenidos en simulaciones son mayores que los experimentales, el orden relativo se predice bastante bien, como ya se observó en el trabajo pionero con cohesinas [Valbuena et al., 2009]. Estos resultados

también sugieren que el uso de modelos estructurales junto con simulaciones de SMD podría ser una vía bastante fiable y relativamente rápida para estudiar la estabilidad mecánica de cohesinas sin la necesidad de realizar experimentos de SMFS.

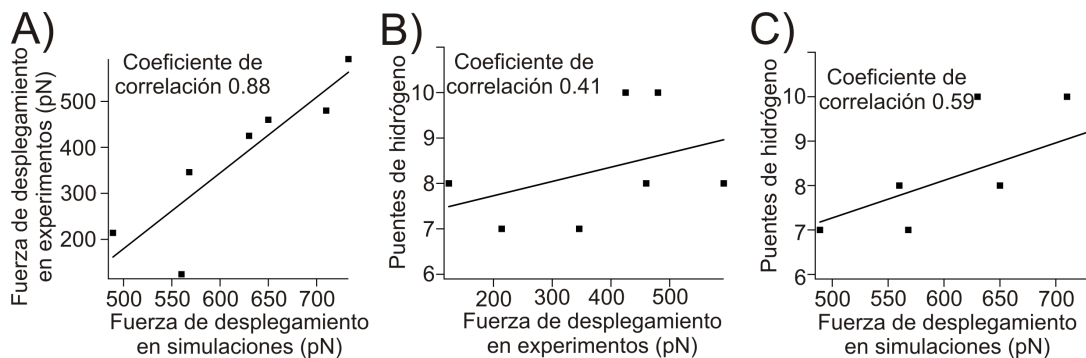


Figura 26: Estudio de la relación de las fuerza de desplegamiento experimentales y calculadas en simulaciones y según el número de puentes de hidrógeno en el parche mecánico. A) Diagrama de dispersión de fuerzas de desplegamiento medidas experimentalmente o en simulaciones. B) Diagrama de dispersión de fuerzas de desplegamiento experimentales respecto al número de puentes de hidrógeno en el parche mecánico vistos en simulaciones. C) Diagrama de dispersión de fuerzas de desplegamiento calculadas en simulaciones respecto al número de puentes de hidrógeno vistos en las mismas. Los cuadros muestran el coeficiente de correlación entre ambas variables.

Anteriormente se había observado que la estabilidad mecánica parecía directamente relacionada con el número de puentes de hidrógeno presentes en el parche mecánico de una proteína [Valbuena et al., 2009]. En la Figura 26.B y C se muestran gráficos de dispersión relacionando las fuerzas de desplegamiento (tanto en experimentos de SMFS como en simulaciones de SMD) y el número de puentes de hidrógeno presentes en el parche mecánico de las cohesinas estudiadas hasta el momento. Se puede observar que el coeficiente de correlación en ambos casos no es muy elevado pero muestra que hay una relación positiva entre ambos parámetros (cuanto mayor es el número de puentes de hidrógeno en el parche mayor es la estabilidad mecánica). Los valores relativamente bajos en los coeficientes de correlación pueden ser indicativos de que otros factores adicionales pueden estar también jugando un papel en la determinación de la estabilidad mecánica de la proteína. Para otras proteínas se han descrito otros factores como el empaquetamiento del núcleo hidrofóbico [Ng et al., 2005] y la movilidad de los bucles [Li et al., 2008b] como factores importantes para determinar la estabilidad mecánica de una proteína.

El desplegamiento mecánico de todos los módulos cohesina estudiados en este trabajo reveló un único pico de fuerza atribuible a cada módulo. Esto indica que en las condiciones experimentales empleadas no se ha detectado ningún intermedio de desplegamiento. La presencia de intermedios en el desplegamiento mecánico de cohesinas según las simulaciones se podría interpretar como otro efecto de la diferencia en la velocidad de estiramiento. El desplegamiento de una proteína suele ser muy cooperativo, por lo que el desplegamiento parcial (del parche mecánico, por ejemplo)

que sufre al empezar a estirarla suele inducir que esta se despliegue completamente. Se asume que este es el motivo por el cual solo se han observado intermedios en el desplegamiento mecánico en pocas proteínas [Marszalek et al., 1999, Oberhauser et al., 2002, Bustamante et al., 2004] y aparecen en casos donde el desplegamiento mecánico provoca una disrupción relativamente pequeña del núcleo de la proteína. Además, si los intermedios presentan bajas estabilidades mecánicas o tiempos de vida muy cortos, estos podrían no observarse en los experimentos de SMFS pero observarse en las simulaciones donde las elevadas velocidades de estiramiento permitirían observarlos.

Es interesante observar que en otro estudio mediante simulaciones [Sikora et al., 2009] en el que se estudió la estabilidad mecánica de un gran número de proteínas con estructura resuelta, incluyendo algunas cohesinas, también se observó la presencia de intermedios que coinciden en posición a los observados en nuestras simulaciones. Además, gracias a las aproximaciones empleadas, en estas simulaciones se emplearon velocidades de estiramiento mucho más bajas ( $5 \cdot 10^5$  nm/s, una velocidad aproximadamente solo 3 órdenes de magnitud mayor que las velocidades empleadas experimentalmente). Esto sugiere que los intermedios podrían ser reales y no artefactos de las simulaciones.

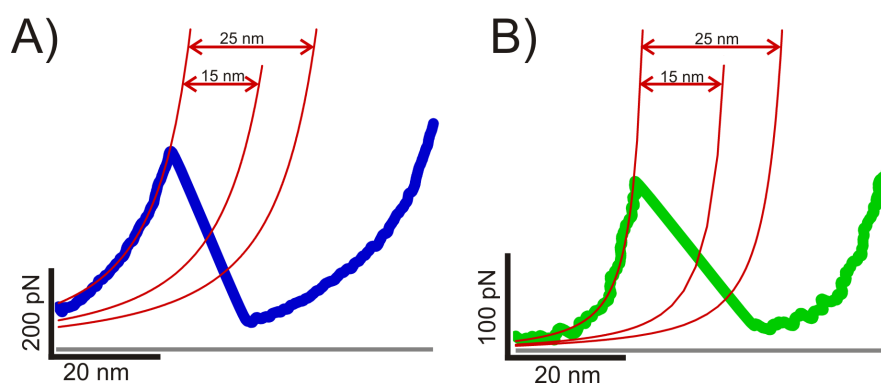


Figura 27: Picos de fuerza de CtA9 (A) y CtA1 (B) donde se superpone un ajuste al modelo WLC empleando como  $\Delta L_C$  la posición aproximada de los intermedios observados en las simulaciones (15 y 25 nm). Los puntos de corte entre estos ajustes y la caída tras el pico de fuerza correspondiente a la rotura del parche mecánico son: CtA9 a 15 nm 250 pN y 170 pN a 25 nm; CtA1 a 15 nm 90 pN y a 25 nm 50 pN.

Dado que la posición de estos intermedios en las simulaciones es relativamente cercana al pico de fuerza inicial, si la estabilidad mecánica de estos es inferior al punto de corte entre el ajuste a WLC y la caída de la fuerza tras ese pico, estos quedarían enmascarados en las curvas de los experimentos de SMFS. Se puede observar en la Figura 27 que estas fuerzas son considerables por lo que no se puede descartar la existencia de estos intermedios en el desplegamiento mecánico.

Si asumimos que estos intermedios vistos en las simulaciones tienen una vida media suficientemente larga para verlos en los experimentos de SMFS y que además su fuerza de desplegamiento tiene la misma dependencia con la velocidad de estiramiento que presenta el parche mecánico, se podría estimar que su magnitud esperada en los experimentos de SMFS sería de unos 170 y 240 pN para los intermedios con  $\Delta L_C$  de 15

y 25 nm para CtA9 y de unos 65 pN para ambos intermedios en CtA1. Teniendo esto en cuenta, es de esperar que la mayoría de estos intermedios quedasen enmascarados en un experimento de SMFS.

### 5.1 NANOMECAÁNICA DE COHESINAS DE *C. thermocellum*

El análisis nanomecánico del celulosoma de *C. thermocellum* se ha centrado en la escafoldina primaria CipA. Esta escafoldina se compone de 9 cohesinas, 2 de ellas en la región externa y las otras 7 en la región conectora. En un primer estudio se caracterizó el comportamiento mecánico de CtA2 y CtA7 [Valbuena et al., 2009] como representantes de la región externa y conectora, respectivamente. Sin embargo, las secuencias de CtA1 y CtA9 son significativamente distintas a las de las otras cohesinas de la misma región (Figura 28) por lo que se decidió estudiar sus propiedades mecánicas.

Los resultados obtenidos en esta tesis indican que la estabilidad mecánica de CtA9 es muy elevada y similar a la medida para CtA7, mientras que la de CtA1 es muy inferior a esta, incluso menor que la de CtA2. Este resultado está de acuerdo con la hipótesis mecánica de trabajo que predice que la estabilidad mecánica de los módulos cohesina en la región conectora debería ser elevada, mientras que la de los módulos de la región externa no precisan este requisito. La estabilidad mecánica de CtA9 se encuentra entre las más elevadas medidas hasta el momento, junto con la obtenida para otras cohesinas o proteínas relacionadas con la adhesión de bacterias a superficies [Valbuena et al., 2009, Gruszka et al., 2015].

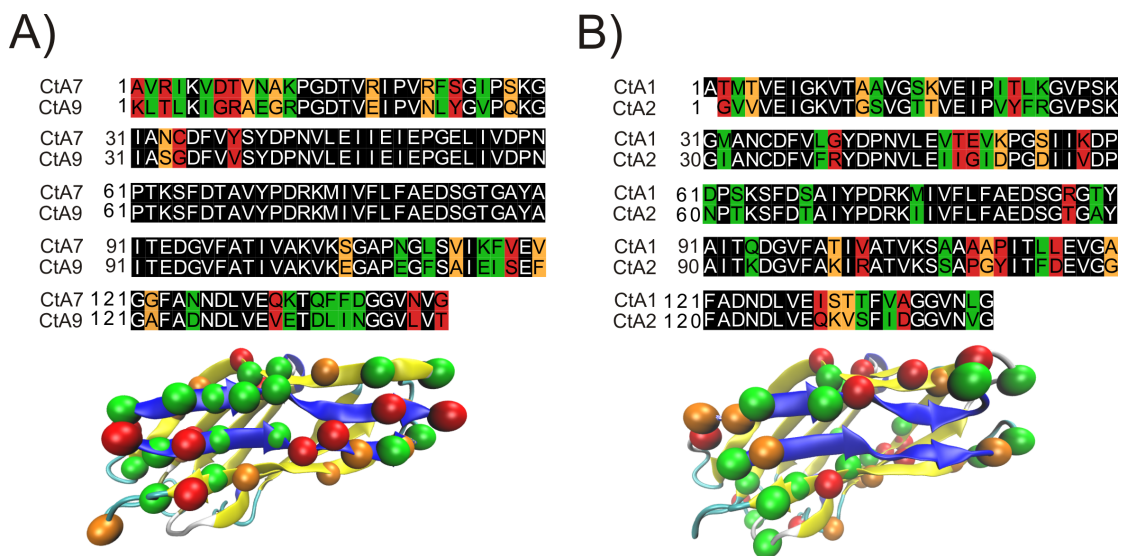


Figura 28: Comparación de módulos cohesina en CipA de *C. thermocellum*. Se muestra el alineamiento (arriba) de la secuencia de los módulos cohesina CtA7 y CtA9 (A) y CtA1 y CtA2 (B), así como la estructura (abajo) de CtA7 (A) y CtA2 (B) resaltando los residuos que son distintos en ambos módulos. Las sustituciones se indican de más a menos conservativas en verde, naranja y rojo. En azul se resaltan la primera y última hebra de la cohesina que forman su parche mecánico para facilitar la visualización.



Es interesante observar que gran parte de las diferencias entre los módulos CtA1 y CtA2 y especialmente entre CtA7 y CtA9, se encuentran en las hebras que forman el parche mecánico y otras cercanas a estas, sin embargo la región responsable de la unión parece muy conservada (Figura 28). En el caso de CtA9, se observan sustituciones que afectan tanto a la carga como tamaño de varios aminoácidos que forman el parche mecánico. A pesar de esto, CtA9 presenta también una muy elevada estabilidad mecánica lo que refuerza la idea de que realmente la estabilidad mecánica es importante en este sistema y que no se trata de un mero epifenómeno.

En proteínas multimodulares eucariotas, se ha observado que los módulos adyacentes no suelen presentar una identidad de secuencia superior al 40% [Han et al., 2007]. Esto se atribuye a una presión evolutiva para evitar que dos módulos consecutivos sean demasiado parecidos como para que si en un momento dado ambos están desplegados puedan formar un plegamiento aberrante por intercambio de dominios [Borgia et al., 2011]. En experimentos de replegamiento mediante SMFS empleando homo-poliproteínas (donde los módulos adyacentes son idénticos) se ha observado que la probabilidad de que esto ocurra es relativamente baja pero detectable [Oberhauser et al., 1999]. Teniendo esto en cuenta, es sorprendente la enorme similitud en la secuencia de cohesinas adyacentes en muchas escafoldinas, que en algunos casos llegan a ser idénticas [Gerngross et al., 1993, Shimon et al., 1997]. La elevada estabilidad mecánica de estas proteínas podría jugar un papel importante al disminuir la probabilidad de que el desplegamiento ocurra. Por otra parte, la presencia de secuencias intermodulares glicosiladas podría jugar un papel importante para evitar este problema.

El análisis de la dependencia de la fuerza de desplegamiento de los módulos con la velocidad de estiramiento junto con el modelado por Monte Carlo ha permitido estudiar los parámetros cinéticos del desplegamiento mecánico de CtA1 y CtA9. Este análisis indicó que CtA9 presenta una mayor dependencia de su fuerza de desplegamiento con la velocidad de estiramiento lo que se traduce en un  $\Delta x_D$  muy pequeño (0.17 nm), lo que indica que se trata de un módulo muy quebradizo (se requiere mucha fuerza para deformarlo pero con una deformación relativamente pequeña, inferior a las dimensiones de una molécula de agua, este ya se despliega).

Para CtA1, en cambio, el valor de  $\Delta x_D$  obtenido fue de 0.31 nm, mucho más parecido al del módulo I27 ( $\Delta x_D = 0.25$  nm) que por su parte es similar al publicado previamente [Carrion-Vazquez et al., 1999b, Best et al., 2002, 2003].

Se puede observar que el  $\Delta x_D$  para la cohesina de alta estabilidad es mucho más corto que el obtenido para la de baja estabilidad mecánica, como en el caso de las cohesinas CcC1, CtA7 y CtA2 ya publicados [Valbuena et al., 2009], lo que sugiere que esta podría ser la base molecular que distingue los módulos cohesina de alta y baja estabilidad mecánica. Sin embargo, los valores obtenidos anteriormente son claramente menores (0.11 nm para CtA7, 0.133 nm para CcC1 y 0.17 nm para CtA2). Hay que tener en cuenta que estos valores se obtuvieron por métodos distintos (a partir de los datos a una única velocidad y mediante experimentos a fuerza controlada)

y además el valor obtenido para I27 en ese estudio es claramente inferior al descrito por otros grupos.

Por otro lado, los valores obtenidos para las tasas de desplegamiento  $\alpha_0$  ( $9 \cdot 10^{-2} \text{s}^{-1}$  para CtA9 y  $3 \cdot 10^{-3} \text{s}^{-1}$  para CtA1) parecen indicar que en ausencia de fuerza CtA1 es más estable cinéticamente que CtA9. Hay que tener en cuenta que los valores de  $\alpha$  están peor determinados ya que dependen exponencialmente de  $\Delta x_D$  [Best et al., 2002, 2003, Bustamante et al., 2004] y por lo tanto son mucho menos fiables. Por otro lado se han descrito otros modelos para interpretar el desplegamiento mecánico de proteínas en los que la dependencia entre la fuerza de desplegamiento y el logaritmo de la tasa de carga puede que no sea lineal [Dudko et al., 2006, Chen et al., 2015].

## 5.2 NANOMECÁNICA DE COHESINAS DE *A. cellulolyticus*

En el caso del celulosoma de *A. cellulolyticus* se ha estudiado la estabilidad mecánica de 4 módulos cohesina provenientes de distintas escafoldinas y regiones escogidos desde el punto de vista mecánico. Para los módulos de ScaA se puede observar que AcA3, de la región externa, presenta una estabilidad mecánica comparable a la de CtA2, también de la región externa de la escafoldina CipA de *C. thermocellum*. Por otro lado, AcA4 presenta una estabilidad mecánica elevada y comparable a la medida para CtA7 o CtA9, ambas de la región conectora de CipA. Estos resultados están de acuerdo con la hipótesis de trabajo.

La estabilidad mecánica de AcB4 sin embargo es comparable a la de AcA3, a pesar de que este módulo se encuentra entre los puntos de anclaje de la célula al sustrato por lo que en principio contradice la hipótesis mecánica del celulosoma. Sin embargo, hay que tener en cuenta que los módulos cohesina de esta escafoldina se unen a la doquiera presente en ScaA no a la de las enzimas directamente. Como se ha comentado, la base de la sinergia del celulosoma radica en la capacidad de la escafoldina de unir las distintas enzimas de forma que estas puedan anclarse al sustrato y además sean más activas en conjunto al acercar actividades complementarias. Por lo tanto, en el caso de que la baja estabilidad mecánica de este módulo fuese insuficiente para prevenir su desplegamiento mecánico el resultado sería la liberación de una escafoldina con sus enzimas unidas que podría seguir siendo activa y promoviendo la sinergia entre las enzimas unidas a ella, resultando en un celulosoma libre.

En este sentido, hay que destacar que se conocen varios microorganismos que forman celulosomas libres (ver en la sección 1.2.1.5) y que incluso entre los que forman celulosoma unido a la superficie celular se ha observado la existencia de escafoldinas que permitirían ensamblar otros celulosomas libres. Por lo tanto no es de esperar que la liberación de estos celulosomas resultase en una pérdida de actividad, asumiendo que la célula pueda seguir unida al sustrato de forma que no se pierda la sinergia CEM y que la integración de varias de estas escafoldinas a un mismo celulosoma no aporte otra fuente de sinergia. Teniendo en cuenta todas estas consideraciones, se puede concluir que estos resultados no contradicen la hipótesis mecánica del celulosoma.

La cohesina AcB4 es la primera cohesina de tipo II estudiada mediante SMFS. El  $\Delta L_C$  medido para este módulo es claramente mayor que el observado en los otros módulos de tipo I estudiados. Esto indica que existe un mayor número de residuos protegidos de la fuerza por su parche mecánico. Teniendo en cuenta que su secuencia es más larga (160 residuos) se estimó que el  $\Delta L_C$  máximo esperado para este módulo (asumiendo que la longitud de la estructura plegada es de 4 nm, similar a la de la cohesina AcB1 de la misma escafoldina [Noach et al., 2010]) es de unos 60 nm. Este valor es similar al hallado experimentalmente ( $57 \pm 0.9$  nm), lo que sugiere que el parche mecánico de esta cohesina está formado por la primera y última hebra de su estructura, como en las otras cohesinas estudiadas. Para las cohesinas de tipo I AcA3, AcA4 y AcC3, los valores de  $\Delta L_C$  medidos son similares a los encontrados en otras cohesinas de tipo I, por lo que se asume que el parche mecánico que las protege de ser desplegadas está formado por los mismos elementos de estructura secundaria.

En un estudio a gran escala de [Sikora et al., 2009] se estiraron mediante simulaciones CG otras cohesinas de tipo II y se observó que los contactos rotos en el primer pico de fuerza corresponden a residuos en los extremos de la molécula y reveló patrones de desplegamiento muy similares a los observados para otras cohesinas de tipo I. Por todo esto, es de esperar que el parche mecánico de este módulo esté formado por los mismos elementos de estructura secundaria que en los módulos de tipo I. En cualquier caso, en el futuro el modelado por homología de esta estructura y la correspondiente simulación SMD respondería a esta pregunta de forma más definitiva.

### 5.3 NANOMECÁNICA DE COHESINAS DE *C. cellulolyticum* Y SECUENCIAS INTERMODULARES

La hipótesis mecánica del celulosoma tiene sentido para esos celulosomas capaces de anclar la célula al sustrato. *C. cellulolyticum* es capaz de unirse a celulosa, además se conoce que al eliminar la escafoldina CipC la unión a celulosa se ve claramente afectada aunque no desaparece [Ferdinand et al., 2013]. Sin embargo, se desconoce el mecanismo de unión y no se han observado elementos en la escafoldina que sugieran como se relaciona esta proteína con la unión de la célula a celulosa. Se ha propuesto que los módulos X puedan jugar un papel al respecto uniendo la escafoldina a la pared celular, como en el caso de *C. cellulovorans* [Kosugi et al., 2004, Ferdinand et al., 2013]. De ser así, dado que CipC presenta 2 módulos X (entre CBM y CcC1 así como entre CcC7 y CcC8), CcC8 se consideraría de la región externa, mientras que las cohesinas 1 a 7 se encontrarían entre 2 módulos X por lo que no estarían sometidas a estrés mecánico y por tanto no es de esperar que sus propiedades mecánicas estuviesen relacionadas de forma simple y directa con la actividad del complejo.

Los resultados obtenidos sugieren que la fuerza de desplegamiento más elevada observada para la construcción CcC1-7 en comparación a CcC1 obtenidos anteriormente [Valbuena et al., 2009] se debe a la presencia de módulos cohesina de mayor estabili-

dad mecánica en la construcción. Este resultado es sorprendente ya que la secuencia de las cohesinas en esta región se encuentra muy conservada.

Para CcC1-7 se observó una distribución unimodal, lo que parece contradecir los resultados obtenidos con CcC123. Sin embargo hay que tener en cuenta que CcC1-7 no presentaba marcadores de monomolecularidad y por tanto CcC1 se encontraba en un extremo de la molécula por lo que es probable que en la mayoría de las curvas obtenidas este módulo no estuviese presente o no se estirase en dirección N-C terminal por lo que su patrón no se reconocería y no se incluiría en el análisis.

Las secuencias intermodulares entre módulos cohesina de las escafoldinas generalmente se han descrito como flexibles, de forma que permiten la adaptación del celulosoma a la complejidad local de su sustrato [Hammel et al., 2005]. Además, en otras proteínas se ha descrito que la inserción de secuencias largas en posiciones cercanas al parche mecánico pueden afectar negativamente a la estabilidad mecánica de la molécula [Li et al., 2008b], por lo que resultaba interesante estudiar si este era el caso para las cohesinas. Los resultados obtenidos muestran que la estabilidad mecánica tanto de CtA7 como de CcC1 no se ve alterada por la presencia de sus secuencias intermodulares. Esto indica que la presencia de estas secuencias no permite desestabilizar el parche mecánico y que no se han formado contactos adicionales entre estas secuencias y los módulos cohesina que les permitan aumentar su estabilidad mecánica.

Solo en el caso de las cohesinas 1 y 2 de ScaB de *A. cellulolyticus* se ha observado la posibilidad de formar este tipo de contactos entre un módulo cohesina y la secuencia intermodular de manera que se establezca también la interacción con el siguiente módulo cohesina [Noach et al., 2010]. Es interesante ver que en esta estructura se forman contactos entre residuos que participan en el parche mecánico de ambas moléculas por lo que sería posible que su estabilidad mecánica se viese alterada. Sin embargo, esto parece un hecho bastante particular y no se espera que sea un patrón general para otras escafoldinas, y ni siquiera para más módulos cohesina dentro de la misma escafoldina ya que los siguientes módulos están separados por secuencias mucho más largas (39 residuos) [Xu et al., 2003].

Por otro lado hay que tener en cuenta que las secuencias intermodulares en las escafoldinas suelen ser ricas en treonina y se cree que la mayoría estarán glicosiladas, como se ha observado para *C. thermocellum* [Gerwig et al., 1989, 1991, 1993]. El efecto que podría tener esto sobre la movilidad o la estabilidad mecánica de los módulos cohesina adyacentes se desconoce.

Como se ha comentado, las diferencias entre las secuencias de CcC1 y CcC3 son escasas, sin embargo su estabilidad mecánica es claramente diferente, siendo la de CcC3 la mayor descrita hasta el momento para proteínas con un plegamiento definido. Al representar los residuos distintos sobre la estructura de CcC1 (código PDB: 1G1K [Spinelli et al., 2000]) se puede observar que solo uno de estos residuos se encuentra en el parche mecánico mientras que el resto se localizan en bucles, especialmente en la zona cercana al extremo N-terminal (ver Figura 29). La cadena lateral de los residuos diferentes está orientada hacia la superficie y las dos sustituciones menos conservativas modifican los aminoácidos por otros más pequeños en el bucle que une la segunda y

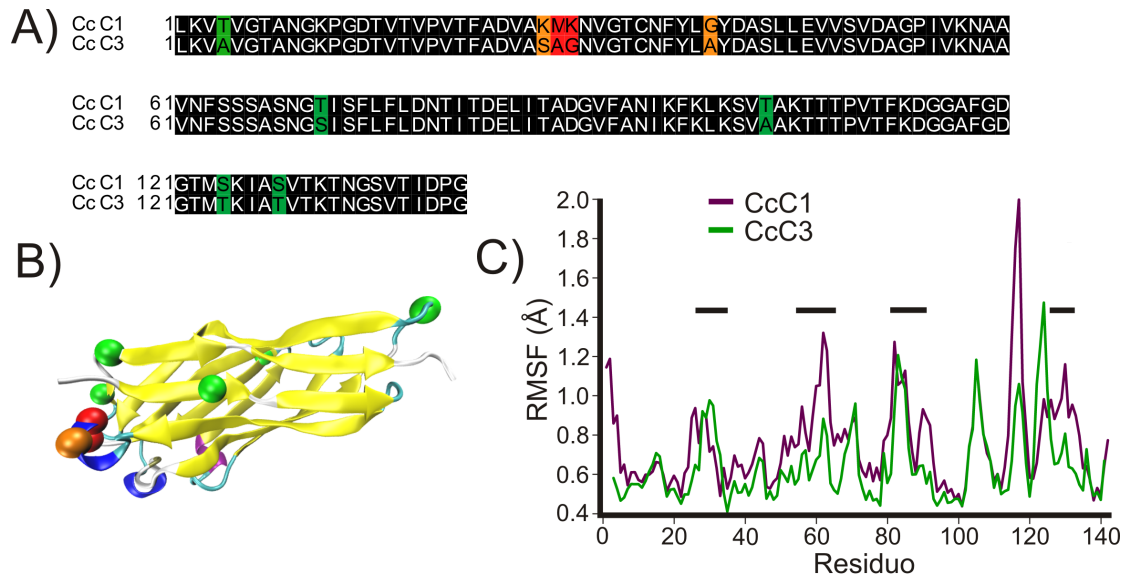


Figura 29: Diferencias entre las secuencias de CcC1 y CcC3. A) alineamiento de las secuencias de CcC1 y CcC3 donde se observa que ambas secuencias son muy parecidas (93.6% identidad). B) estructura de CcC1 (código PDB: 1G1K [Spinelli et al., 2000]) donde se destaca mediante una representación en esferas los  $C_{\alpha}$  de las posiciones en las que se observan diferencias entre ambas estructuras. Las sustituciones se muestran de más a menos conservativas en verde, naranja y rojo respectivamente. C) Comparación del RMSF durante los últimos 5 ns de dinámica libre de CcC1 (código PDB: 1G1K) y CcC3 (modelo por homología). Las líneas negras indican los residuos que forman los bucles en el lado N-terminal de la proteína.

tercera hebra. Se ha observado para otras proteínas que los aminoácidos que forman bucles cerca del parche mecánico pueden influir en la estabilidad mecánica de este [Li et al., 2008b]. En particular se observó que para bucles con mucha movilidad la estabilidad mecánica se reducía. Si se comparan las fluctuaciones de los bucles tanto de CcC1 como de CcC3 en las simulaciones de MD (Figura 29) se observa una mayor fluctuación del extremo N-terminal en CcC1, mientras que los bucles en el lado N-terminal no presentan grandes diferencias. Sin embargo hay que tener en cuenta que dicho extremo presenta un aminoácido adicional en la estructura de CcC1 respecto a la de CcC3, lo que podría influir en este resultado.

#### 5.4 NANOMECÁNICA DE COHESINAS INTERACCIONANDO CON DOQUERINA

Se ha observado que la unión a doquerina conduce a la estabilización termodinámica tanto de la cohesina como de la doquerina [Salama-Alber et al., 2013]. Sin embargo, según los resultados obtenidos en esta tesis, la interacción de ambas proteínas no conduce a la estabilización mecánica de la cohesina. Este resultado no es sorprendente teniendo en cuenta la estructura del complejo formado por ambas proteínas, donde se puede observar claramente que la región de interacción en la cohesina se encuentra alejada del parche mecánico, que es la región que le confiere principalmente la estabilidad mecánica (Figura 30). Además, mediante el uso de simulaciones de SMD,

se ha observado que la interacción cohesina-doquerina se puede mantener cuando la cohesina se encuentra parcialmente desplegada, aunque se desconoce si este proceso se puede dar *in vivo* o si puede mantenerse este estado durante un periodo de tiempo suficientemente largo.

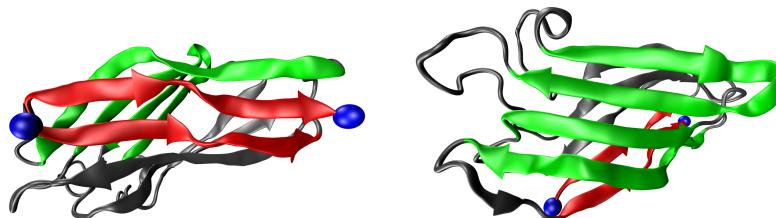


Figura 30: Parche mecánico y región de unión a doquerina en CtA7. Se muestran dos vistas de la cohesina CtA7 (código PDB: 1AOH [Tavares et al., 1997]) en las que se destaca el parche mecánico (en rojo) y la cara de interacción con la doquerina (en verde). En azul se muestran los  $C_{\alpha}$  desde los que se estira en las simulaciones de SMD.

Una ventaja de las simulaciones de MD respecto a los experimentos es que se puede controlar muy bien el sistema estudiado. Para este caso concreto, esto nos permite conocer el modo de unión entre la cohesina y la doquerina (mientras que en los experimentos no podemos asegurar en qué modo se une o si ambos se encuentran presentes en distintas moléculas) y descartar que esto tenga efecto sobre la estabilidad mecánica de la cohesina a la que se une.

## 5.5 GENERALIDAD DE LA HIPÓTESIS

Los resultados presentados son relativos a las propiedades mecánicas de las cohesinas, sin embargo la estabilidad mecánica de las interacciones no covalentes que mantienen el celulosoma son también muy importantes. Mediante SMFS se han estudiado las propiedades mecánicas de la interacción cohesina-doquerina que es la que dirige la formación del complejo. Sorprendentemente, a pesar de la superficie relativamente pequeña de interacción entre la cohesina y la doquerina, las fuerzas observadas son bastante altas entre la doquerina de Cel48S y la CtA2 (>120 pN a una tasa de carga similar) [Stahl et al., 2012]. Mientras que para la interacción de tipo III entre el módulo XDoquerinade CtA y la cohesina E de *R. flavofaciens*, que se espera que sea el principal anclaje de la célula al sustrato, la estabilidad mecánica observada es excepcionalmente alta (600-750 pN a una tasa de carga similar y con poca dependencia de este parámetro) [Schoeler et al., 2014], superando incluso la estabilidad de algunas cohesinas. La presencia de uniones tan fuertes permitiría que fuerzas relativamente altas se transmitiesen a través de la escafoldina.

La respuesta mecánica de las interacciones cohesina-doquerina también se ha estudiado mediante modelos CG. En este caso se confirmó que la respuesta de estos complejos depende fuertemente de la geometría de estiramiento (según el extremo

de cada molécula empleado para aplicar fuerza) [Wojciechowski et al., 2014]. En las estructuras estudiadas se observó que al estirar en la geometría en la que se podría esperar que actúe la fuerza en el sistema natural, estas solían presentar mayor estabilidad mecánica y patrones de desplegamiento mejor definidos.

Una cuestión importante es el origen y la escala de las fuerzas que pueden afectar al celulosoma en su ambiente natural. Al plantear la hipótesis mecánica en la introducción se ha comentado que los microorganismos que los producen pueden estar sujetos a fuerzas en su ambiente lo que podría ser una fuente de estrés mecánico al moverse el microorganismo respecto al sustrato. Es difícil estimar la escala de estas fuerzas dado que estos ambientes son poco definidos. Sin embargo, ya sea en el tracto digestivo de rumiantes o en el suelo, se pueden establecer flujos hidrodinámicos relativamente fuertes [Marcos, 2011]. Además, muchos de los microorganismos productores de celulosoma también producen flagelos y otros elementos de movilidad [Raman et al., 2011, Wei et al., 2014] que podrían aumentar el estrés mecánico en el sistema al promover la movilidad de la bacteria, aunque estos elementos suelen expresarse en etapas avanzadas de la degradación de un sustrato. Es interesante observar que en un estudio de la organización del celulosomas mediante microscopía electrónica de transmisión se observó que al incubar una muestra de celulosoma y secarla por capilaridad empleando papel de filtro (lo que crea un flujo unidireccional), estos aparecían extendidos. Además en estos experimentos se observó la presencia de moléculas grandes separadas del celulosoma lo que se interpretó como enzimas que habían sido desancladas de la escafoldina [Madkour & Mayer, 2003].

Finalmente, se ha descrito para bacterias monoflageladas que al nadar cerca de una superficie se producen fuertes cambios en su posición vertical como resultado de una amplificación del efecto Browniano cerca de la superficie [Li et al., 2008a]. Esta podría ser una situación similar al caso de *C. thermocellum* al acercarse al sustrato.

En los experimentos de SMFS las fuerzas de desplegamiento observadas son relativamente altas, sin embargo en las situaciones descritas anteriormente difícilmente se podrían alcanzar fuerzas tan grandes. Hay que tener en cuenta que la fuerza necesaria para desplegar una proteína depende de la tasa de carga, es decir de la velocidad en la que se aplica la fuerza. En los experimentos de AFM las tasas de carga accesibles son relativamente elevadas lo que resulta en fuerzas de desplegamiento muy altas. Sin embargo, en el sistema natural las fuerzas presentes pueden ser muy pequeñas, pero estas fuerzas actuarán sobre el sistema durante tiempos muy prolongados ya que la degradación de estos sustratos es lenta. En estas condiciones, las cohesinas de alta estabilidad serán capaces de mantenerse plegadas durante más tiempo que las de baja estabilidad mecánica en presencia de estas fuerzas por lo que a pesar de que no se presentasen fuerzas tan elevadas, la estabilidad mecánica de las cohesinas seguiría siendo un parámetro relevante.

Parte VI

CONCLUSIONES



## CONCLUSIONES

---

A continuación se enumeran las conclusiones más relevantes obtenidas del trabajo de esta tesis:

1. En las escafoldinas primarias, aquellas encargadas de unir las unidades catalíticas, la estabilidad mecánica de sus módulos cohesina de la región conectora es mayor que la de los de la región externa. La elevada estabilidad mecánica de los módulos cohesina parece estar originada por presentar una distancia al estado de transición muy corta.
2. Las simulaciones por SMD permiten estudiar la estabilidad mecánica de módulos cohesina, incluso en el caso de utilizar modelos estructurales generados por homología. Las estabilidades mecánicas calculadas por este método presentan una buena correlación con las medidas experimentalmente.
3. La interacción con doquerina no altera la estabilidad mecánica ni la posición del parche mecánico de los módulos cohesina, tanto de la región conectora como externa y para los dos modos de unión de la doquerina. Las simulaciones por SMD refuerzan esta conclusión.
4. La presencia de secuencias intermodulares naturales de las cohesinas no produce efectos sobre la estabilidad mecánica ni en la posición del parche mecánico de los módulos cohesina estudiados.

## BIBLIOGRAFÍA

---

- Adams J.J., Webb B.A., Spencer H.L. & Smith S.P. (2005) Structural characterization of type II dockerin module from the cellulosome of *Clostridium thermocellum*: calcium-induced effects on conformation and target recognition. *Biochemistry* 44(6):2173–2182.
- Adams J.J., Pal G., Jia Z. & Smith S.P. (2006) Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. *Proc Natl Acad Sci U S A* 103(2):305–310.
- Adams J.J., Currie M.A., Ali S., Bayer E.A., Jia Z. & Smith S.P. (2010) Insights into higher-order organization of the cellulosome revealed by a dissect-and-build approach: crystal structure of interacting *Clostridium thermocellum* multimodular components. *J Mol Biol* 396(4):833–839.
- Ainavarapu S.R.K., Brujic J., Huang H.H., Wiita A.P., Lu H., Li L., Walther K.A., Carrion-Vazquez M., Li H. & Fernandez J.M. (2007) Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys J* 92(1):225–233.
- Alber O., Noach I., Rincon M.T., Flint H.J., Shimon L.J.W., Lamed R., Frolow F. & Bayer E.A. (2009) Cohesin diversity revealed by the crystal structure of the anchoring cohesin from *Ruminococcus flavefaciens*. *Proteins* 77(3):699–709.
- Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
- Amber (2012) *Amber 12*.
- Arfi Y., Shamshoum M., Rogachev I., Peleg Y. & Bayer E.A. (2014) Integration of bacterial lytic polysaccharide monoxygenases into designer cellulosomes promotes enhanced cellulose degradation. *Proc Natl Acad Sci U S A* 111(25):9109–9114.
- Artzi L., Dassa B., Borovok I., Shamshoum M., Lamed R. & Bayer E.A. (2014) Cellulosomics of the cellulolytic thermophile *Clostridium clariflavum*. *Biotechnology for Biofuels* 7.
- Aubin-Tam M.E., Olivares A.O., Sauer R.T., Baker T.A. & Lang M.J. (2011) Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell* 145(2):257–267.
- Basché T., Nie S. & Fernandez J.M. (2001) Single molecules. *Proc Natl Acad Sci U S A* 98(19):10527–10528.

- Bayer E.A. & Lamed R. (1986) Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. *J Bacteriol* 167(3):828–836.
- Bayer E.A., Kenig R. & Lamed R. (1983) Adherence of *Clostridium thermocellum* to cellulose. *J Bacteriol* 156(2):818–827.
- Bayer E.A., Morag E. & Lamed R. (1994) The cellulosome—a treasure-trove for biotechnology. *Trends Biotechnol* 12(9):379–386.
- Bayer E.A., Shimon L.J., Shoham Y. & Lamed R. (1998) Cellulosomes-structure and ultrastructure. *J Struct Biol* 124(2-3):221–234.
- Bayer E.A., Smith S.P., Noach I., Alber O., Adams J.J., Lamed R., Shimon L.W. & Frolow F. (2009) *Can we crystallize a cellulosome?* Ito Print Publishing Division.
- Bayer E.A., Belaich J.P., Shoham Y. & Lamed R. (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol* 58: 521–554.
- Bayer E.A., Lamed R. & Himmel M.E. (2007) The potential of cellulases and cellulosomes for cellulosic waste management. *Curr Opin Biotechnol* 18(3):237–245.
- Bayer E.A., Lamed R., White B.A. & Flint H.J. (2008) From cellulosomes to cellulosomes. *Chem Rec* 8(6):364–377.
- Bayer E.A., Shoham Y. & Lamed R. (2013) *Lignocellulose-decomposing bacteria and their enzyme systems*. The Prokaryotes – Prokaryotic Physiology and Biochemistry. Springer-Verlag Berlin Heidelberg.
- Béguin P. & Aubert J.P. (1994) The biological degradation of cellulose. *FEMS Microbiol Rev* 13(1):25–58.
- Bell G.I. (1978) Models for the specific adhesion of cells to cells. *Science* 200(4342): 618–627.
- Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. & Bourne P.E. (2000) The protein data bank. *Nucleic Acids Res* 28(1):235–242.
- Best R.B., Li B., Steward A., Daggett V. & Clarke J. (2001) Can non-mechanical proteins withstand force? stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys J* 81(4):2344–2356.
- Best R.B., Fowler S.B., Toca-Herrera J.L. & Clarke J. (2002) A simple method for probing the mechanical unfolding pathway of proteins in detail. *Proc Natl Acad Sci U S A* 99(19):12143–12148.
- Best R.B., Brockwell D.J., Toca-Herrera J.L., Blake A.W., Smith D., Radford S.E. & Clarke J. (2003) Force mode atomic force microscopy as a tool for protein folding studies. *Analytica Chimica Acta* 479:87–105.

- Binnig, Quate & Gerber (1986) Atomic force microscope. *Phys Rev Lett* 56(9):930–933.
- Bomble Y.J., Beckham G.T., Matthews J.F., Nimlos M.R., Himmel M.E. & Crowley M.F. (2011) Modeling the self-assembly of the cellulosome enzyme complex. *J Biol Chem* 286(7):5614–5623.
- Boraston A.B., Bolam D.N., Gilbert H.J. & Davies G.J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* 382(Pt 3):769–781.
- Borgia A., Williams P.M. & Clarke J. (2008) Single-molecule studies of protein folding. *Annu Rev Biochem* 77:101–125.
- Borgia M.B., Borgia A., Best R.B., Steward A., Nettels D., Wunderlich B., Schuler B. & Clarke J. (2011) Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* 474(7353):662–665.
- Borrion A.L., McManus M.C. & Hammond G.P. (2012) Environmental life cycle assessment of bioethanol production from wheat straw. *Biomass and Bioenergy* 47:9–19.
- Brockwell D.J., Beddard G.S., Clarkson J., Zinober R.C., Blake A.W., Trinick J., Olmsted P.D., Smith D.A. & Radford S.E. (2002) The effect of core destabilization on the mechanical resistance of I27. *Biophys J* 83(1):458–472.
- Brockwell D.J., Paci E., Zinober R.C., Beddard G.S., Olmsted P.D., Smith D.A., Perham R.N. & Radford S.E. (2003) Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nat Struct Biol* 10(9):731–737.
- Burnham N. & Colton R. (1989) Measuring the nanomechanical properties and surface forces of materials using an atomic force microscope. *J. Vacuum Sci. Technol* 7:2906–2913.
- Bustamante C., Marko J.F., Siggia E.D. & Smith S. (1994) Entropic elasticity of lambda-phage DNA. *Science* 265(5178):1599–1600.
- Bustamante C., Chemla Y.R., Forde N.R. & Izhaky D. (2004) Mechanical processes in biochemistry. *Annu Rev Biochem* 73:705–748.
- Cao Y., Balamurali M.M., Sharma D. & Li H. (2007) A functional single-molecule binding assay via force spectroscopy. *Proc Natl Acad Sci U S A* 104(40):15677–15681.
- Cao Y., Yoo T., Zhuang S. & Li H. (2008) Protein-protein interaction regulates proteins' mechanical stability. *J Mol Biol* 378(5):1132–1141.
- Carrion-Vazquez M., Marszalek P.E., Oberhauser A.F. & Fernandez J.M. (1999a) Atomic force microscopy captures length phenotypes in single proteins. *Proc Natl Acad Sci U S A* 96(20):11288–11292.
- Carrion-Vazquez M., Oberhauser A.F., Fowler S.B., Marszalek P.E., Broedel S.E., Clarke J. & Fernandez J.M. (1999b) Mechanical and chemical unfolding of a single protein: a comparison. *Proc Natl Acad Sci U S A* 96(7):3694–3699.

- Carrion-Vazquez M., Oberhauser A.F., Fisher T.E., Marszalek P.E., Li H. & Fernandez J.M. (2000) Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog Biophys Mol Biol* 74(1-2):63–91.
- Carrión-Vázquez M., Oberhauser A., Díez H., Hervás R., Oroz J., Fernández J. & Martínez-Martín D. (2006) *Protein nanomechanics, as studied by AFM Single-Molecule Force Spectroscopy*. Advanced Techniques in Biophysics. Springer-Verlag.
- Carrion-Vazquez M., Li H., Lu H., Marszalek P.E., Oberhauser A.F. & Fernandez J.M. (2003) The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Biol* 10 (9):738–743.
- Carvalho A.L., Dias F.M.V., Prates J.A.M., Nagy T., Gilbert H.J., Davies G.J., Ferreira L.M.A., ao M.J.R. & Fontes C.M.G.A. (2003) Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proc Natl Acad Sci U S A* 100(24): 13809–13814.
- Carvalho A.L., Dias F.M.V., Nagy T., Prates J.A.M., Proctor M.R., Smith N., Bayer E.A., Davies G.J., Ferreira L.M.A., ao M.J.R., Fontes C.M.G.A. & Gilbert H.J. (2007) Evidence for a dual binding mode of dockerin modules to cohesins. *Proc Natl Acad Sci U S A* 104(9):3089–3094.
- Case D.A., Darden T.A., Cheatham T.E., Simmerling C.L., Duke J.W.R.E., Luo R., Crowley M., Walker R.C., Merz W.Z.K.M., Wang B., Hayik S., and G. Seabra A.R., Kolosvary I., Wong K.F., Paesani F., Vanicek J., Wu X., Brozell S.R., Steinbrecher T., Gohlke H., Yang L., Tan C., Mongan J., Hornak V., Cui G., Mathews D.H., Seetin M.G., Sagui C., Babin V. & Kollman P.A. (2008) *Amber10*. University of California, San Francisco.
- Caspi J., Irwin D., Lamed R., Li Y., Fierobe H.P., Wilson D.B. & Bayer E.A. (2008) Conversion of *Thermobifida fusca* free exoglucanases into cellulosomal components: comparative impact on cellulose-degrading activity. *J Biotechnol* 135(4):351–357.
- Caspi J., Barak Y., Haimovitz R., Irwin D., Lamed R., Wilson D.B. & Bayer E.A. (2009) Effect of linker length and dockerin position on conversion of a *Thermobifida fusca* endoglucanase to the cellulosomal mode. *Appl Environ Microbiol* 75(23):7335–7342.
- Caspi J., Barak Y., Haimovitz R., Gilary H., Irwin D.C., Lamed R., Wilson D.B. & Bayer E.A. (2010) *Thermobifida fusca* exoglucanase Cel6B is incompatible with the cellulosomal mode in contrast to endoglucanase Cel6A. *Syst Synth Biol* 4(3):193–201.
- CAZY. Carbohydrate active enzymes database. (1998).
- Chen Y., Radford S.E. & Brockwell D.J. (2015) Force-induced remodelling of proteins and their complexes. *Curr Opin Struct Biol* 30:89–99.
- Chundawat S.P.S., Beckham G.T., Himmel M.E. & Dale B.E. (2011) Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu Rev Chem Biomol Eng* 2:121–145.

- Chwastyk M., Galera-Prat A., Sikora M., Gómez-Sicilia A., Carrión-Vázquez M. & Cieplak M. (2014) Theoretical tests of the mechanical protection strategy in protein nanomechanics. *Proteins* 82(5):717–726.
- Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W. & Kollman P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.
- Currie M.A., Adams J.J., Faucher F., Bayer E.A., Jia Z. & Smith S.P. (2012) Scaffoldin conformation and dynamics revealed by a ternary complex from the *Clostridium thermocellum* cellulosome. *J Biol Chem* 287(32):26953–26961.
- Currie M.A., Cameron K., Dias F.M.V., Spencer H.L., Bayer E.A., Fontes C.M.G.A., Smith S.P. & Jia Z. (2013) Small angle x-ray scattering analysis of clostridium thermocellum cellulosome n-terminal complexes reveals a highly dynamic structure. *J Biol Chem* 288(11):7978–7985.
- Czjzek M., Fierobe H.P. & Receveur-Bréchet V. (2012) Small-angle x-ray scattering and crystallography: a winning combination for exploring the multimodular organization of cellulolytic macromolecular complexes. *Methods Enzymol* 510:183–210.
- Dassa B., Borovok I., Lamed R., Henrissat B., Coutinho P., Hemme C.L., Huang Y., Zhou J. & Bayer E.A. (2012) Genome-wide analysis of acetivibrio cellulolyticus provides a blueprint of an elaborate cellulosome system. *BMC Genomics* 13:210.
- Dassa B., Borovok I., Ruimy-Israeli V., Lamed R., Flint H.J., Duncan S.H., Henrissat B., Coutinho P., Morrison M., Mosoni P., Yeoman C.J., White B.A. & Bayer E.A. (2014) Rumen cellulosomes: divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six ruminococcal strains. *PLoS One* 9(7):e99221.
- Demain A.L., Newcomb M. & Wu J.H.D. (2005) Cellulase, clostridia, and ethanol. *Microbiol Mol Biol Rev* 69(1):124–154.
- Dietz H., Bertz M., Schlierf M., Berkemeier F., Bornschlögl T., Junker J.P. & Rief M. (2006) Cysteine engineering of polyproteins for single-molecule force spectroscopy. *Nat Protoc* 1(1):80–84.
- Ding S.Y., Bayer E.A., Steiner D., Shoham Y. & Lamed R. (1999) A novel cellulosomal scaffoldin from *Acetivibrio cellulolyticus* that contains a family 9 glycosyl hydrolase. *J Bacteriol* 181(21):6720–6729.
- Ding S.Y., Bayer E.A., Steiner D., Shoham Y. & Lamed R. (2000) A scaffoldin of the *Bacteroides cellulosolvens* cellulosome that contains 11 type II cohesins. *J Bacteriol* 182(17):4915–4925.
- Dudko O.K., Hummer G. & Szabo A. (2006) Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys Rev Lett* 96(10):108101.

- Eswar N., Webb B., Marti-Renom M.A., Madhusudhan M.S., Eramian D., Shen M.Y., Pieper U. & Sali A. (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* Chapter 5:Unit 5.6.
- Evans E. & Ritchie K. (1997) Dynamic strength of molecular adhesion bonds. *Biophys J* 72(4):1541–1555.
- Evans E. & Ritchie K. (1999) Strength of a weak bond connecting flexible polymer chains. *Biophys J* 76(5):2439–2447.
- Ferdinand P.H., Borne R., Trotter V., Pagés S., Tardif C., Fierobe H.P. & Perret S. (2013) Are cellulosome scaffolding protein CipC and CBM3-containing protein HycP, involved in adherence of *Clostridium cellulolyticum* to cellulose? *PLoS One* 8(7):e69360.
- Fierobe H.P., Mechaly A., Tardif C., Belaich A., Lamed R., Shoham Y., Belaich J.P. & Bayer E.A. (2001) Design and production of active cellulosome chimeras. Selective incorporation of dockerin-containing enzymes into defined functional complexes. *J Biol Chem* 276(24):21257–21261.
- Fierobe H.P., Bayer E.A., Tardif C., Czjzek M., Mechaly A., Bélaïch A., Lamed R., Shoham Y. & Bélaïch J.P. (2002) Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components. *J Biol Chem* 277(51):49621–49630.
- Fierobe H.P., Mingardon F., Mechaly A., Bélaïch A., Rincon M.T., Pagés S., Lamed R., Tardif C., Bélaïch J.P. & Bayer E.A. (2005) Action of designer cellulosomes on homogeneous versus complex substrates: controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin. *J Biol Chem* 280(16):16325–16334.
- Fontes C.M.G.A. & Gilbert H.J. (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu Rev Biochem* 79:655–681.
- Fowler S.B., Best R.B., Herrera J.L.T., Rutherford T.J., Steward A., Paci E., Karplus M. & Clarke J. (2002) Mechanical unfolding of a titin ig domain: structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering. *J Mol Biol* 322(4):841–849.
- Francisco J.A., Earhart C.F. & Georgiou G. (1992) Transport and anchoring of beta-lactamase to the external surface of *Escherichia coli*. *Proc Natl Acad Sci U S A* 89(7):2713–2717.
- Francisco J.A., Stathopoulos C., Warren R.A., Kilburn D.G. & Georgiou G. (1993) Specific adhesion and hydrolysis of cellulose by intact *Escherichia coli* expressing surface anchored cellulase or cellulose binding domains. *Biotechnology (N Y)* 11(4):491–495.
- Galera-Prat A., Hermans R., Hervás R., Gómez-Sicilia A. & Carrión-Vázquez M. (2011) *Single-molecule force spectroscopy in: Atomic Force Microscopy in Liquids*. Wiley.

- Gefen G., Anbar M., Morag E., Lamed R. & Bayer E.A. (2012) Enhanced cellulose degradation by targeted integration of a cohesin-fused  $\beta$ -glucosidase into the *Clostridium thermocellum* cellulosome. *Proc Natl Acad Sci U S A* 109(26):10298–10303.
- Gerngross U.T., Romaniec M.P., Kobayashi T., Huskisson N.S. & Demain A.L. (1993) Sequencing of a *Clostridium thermocellum* gene (*cipA*) encoding the cellulosomal SL-protein reveals an unusual degree of internal homology. *Mol Microbiol* 8(2):325–334.
- Gerwig G.J., de Waard P., Kamerling J.P., Vliegthart J.F., Morgenstern E., Lamed R. & Bayer E.A. (1989) Novel O-linked carbohydrate chains in the cellulase complex (cellulosome) of *Clostridium thermocellum*. 3-O-Methyl-N-acetylglucosamine as a constituent of a glycoprotein. *J Biol Chem* 264(2):1027–1035.
- Gerwig G.J., Kamerling J.P., Vliegthart J.F., Morag E., Lamed R. & Bayer E.A. (1991) Primary structure of o-linked carbohydrate chains in the cellulosome of different *Clostridium thermocellum* strains. *Eur J Biochem* 196(1):115–122.
- Gerwig G.J., Kamerling J.P., Vliegthart J.F., Morag E., Lamed R. & Bayer E.A. (1993) The nature of the carbohydrate-peptide linkage region in glycoproteins from the cellulosomes of *Clostridium thermocellum* and *Bacteroides cellulosolvens*. *J Biol Chem* 268(36):26956–26960.
- Ghrayeb J. & Inouye M. (1984) Nine amino acid residues at the NH<sub>2</sub>-terminal of lipoprotein are sufficient for its modification, processing, and localization in the outer membrane of *Escherichia coli*. *J Biol Chem* 259(1):463–467.
- Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. & Warren R.A. (1991) Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. *Microbiol Rev* 55(2):303–315.
- Gilmore S.P., Henske J.K. & O'Malley M.A. (2015) Driving biomass breakdown through engineered cellulosomes. *Bioengineered* 6(4):204–208.
- Good M.C., Zalatan J.G. & Lim W.A. (2011) Scaffold proteins: hubs for controlling the flow of cellular information. *Science* 332(6030):680–686.
- Gruszka D.T., Whelan F., Farrance O.E., Fung H.K.H., Paci E., Jeffries C.M., Svergun D.I., Baldock C., Baumann C.G., Brockwell D.J., Potts J.R. & Clarke J. (2015) Cooperative folding of intrinsically disordered domains drives assembly of a strong elongated protein. *Nat Commun* 6:7271.
- Haimovitz R., Barak Y., Morag E., Voronov-Goldman M., Shoham Y., Lamed R. & Bayer E.A. (2008) Cohesin-dockerin microarray: Diverse specificities between two complementary families of interacting protein modules. *Proteomics* 8(5):968–979.
- Hamberg Y., Ruimy-Israeli V., Dassa B., Barak Y., Lamed R., Cameron K., Fontes C.M.G.A., Bayer E.A. & Fried D.B. (2014) Elaborate cellulosome architecture of *Acetivibrio cellulolyticus* revealed by selective screening of cohesin-dockerin interactions. *PeerJ* 2:e636.



- Hammel M., Fierobe H.P., Czjzek M., Kurkal V., Smith J.C., Bayer E.A., Finet S. & Receveur-Bréchet V. (2005) Structural basis of cellulosome efficiency explored by small angle x-ray scattering. *J Biol Chem* 280(46):38562–38568.
- Han J.H., Batey S., Nickson A.A., Teichmann S.A. & Clarke J. (2007) The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol* 8(4):319–330.
- Hasunuma T., Okazaki F., Okai N., Hara K.Y., Ishii J. & Kondo A. (2013) A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. *Bioresour Technol* 135:513–522.
- Hawkins G.D., Cramer C.J. & Truhlar D.G. (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *The Journal of Physical Chemistry* 100(51):19824–19839.
- Hervás R., Oroz J., Galera-Prat A., Goñi O., Valbuena A., Vera A.M., Gómez-Sicilia A., Losada-Urzáiz F., Uversky V.N., Menéndez M., Laurents D.V., Bruix M. & Carrión-Vázquez M. (2012) Common features at the start of the neurodegeneration cascade. *PLoS Biol* 10(5):e1001335.
- Himmel M.E., Ding S.Y., Johnson D.K., Adney W.S., Nimlos M.R., Brady J.W. & Foust T.D. (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315(5813):804–807.
- Himmel M.E., Xu Q., Luo Y., Ding S.Y., Lamed R. & Bayer E.A. (2010) Microbial enzyme systems for biomass conversion: emerging paradigms. *Future Science* 1:323–341.
- Howard J. (2001) *Mechanics of motor proteins and the cytoskeleton*. Sinauer.
- Huang G.L., Anderson T.D. & Clubb R.T. (2014) Engineering microbial surfaces to degrade lignocellulosic biomass. *Bioengineered* 5(2):96–106.
- Huffer S., Roche C.M., Blanch H.W. & Clark D.S. (2012) *Escherichia coli* for biofuel production: bridging the gap from promise to practice. *Trends Biotechnol* 30(10):538–545.
- Humphrey W., Dalke A. & Schulten K. (1996) VMD: visual molecular dynamics. *J Mol Graph* 14(1):33–8, 27–8.
- Hyeon J.E., Jeon W.J., Whang S.Y. & Han S.O. (2011) Production of minicellulosomes for the enhanced hydrolysis of cellulosic substrates by recombinant *Corynebacterium glutamicum*. *Enzyme Microb Technol* 48(4-5):371–377.
- Javadi Y., Fernandez J.M. & Perez-Jimenez R. (2013) Protein folding under mechanical forces: a physiological view. *Physiology (Bethesda)* 28(1):9–17.
- Johnson E.A. & Demain A.L. (1984) Probable involvement of sulfhydryl groups and a metal as essential components of the cellulase of *Clostridium thermocellum*. *Arch Microbiol* 137:135–138.

- Junker J.P., Ziegler F. & Rief M. (2009) Ligand-dependent equilibrium fluctuations of single calmodulin molecules. *Science* 323(5914):633–637.
- Karami A., Latifi A.M. & Khodi S. (2014) Comparison of the organophosphorus hydrolase surface display using InaVN and Lpp-OmpA systems in *Escherichia coli*. *J Microbiol Biotechnol* 24(3):379–385.
- Kern J., Wilton R., Zhang R., Binkowski T.A., Joachimiak A. & Schneewind O. (2011) Structure of surface layer homology (SLH) domains from *Bacillus anthracis* surface array protein. *J Biol Chem* 286(29):26042–26049.
- King A. (2013) *Removal of excess cellulose and associated polysaccharides in fruit and vegetable by-products – Implication for use in feed for monogastric farm animals*. En *Cellulose – fundamental aspects*. InTech.
- Kosugi A., Amano Y., Murashima K. & Doi R.H. (2004) Hydrophilic domains of scaffolding protein CbpA promote glycosyl hydrolase activity and localization of cellulosomes to the cell surface of *Clostridium cellulovorans*. *J Bacteriol* 186(19):6351–6359.
- Lamed R., Setter E. & Bayer E.A. (1983) Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *J Bacteriol* 156(2):828–836.
- Lamed R., Kenig R., Setter E. & Bayer E.A. (1985) Major characteristics of the cellulolytic system of *Clostridium thermocellum* coincide with those of the purified cellulosome. *Enzyme and Microbial Technology* 7(1):37–41.
- Laskowski R., MacArthur M., Moss D. & Thornton J. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26:283–291.
- Lee E.H., Hsin J., Sotomayor M., Comellas G. & Schulten K. (2009) Discovery through the computational microscope. *Structure* 17(10):1295–1306.
- Li G., Tam L.K. & Tang J.X. (2008a) Amplified effect of brownian motion in bacterial near-surface swimming. *Proc Natl Acad Sci U S A* 105(47):18355–18359.
- Li H., Oberhauser A.F., Redick S.D., Carrion-Vazquez M., Erickson H.P. & Fernandez J.M. (2001) Multiple conformations of PEVK proteins detected by single-molecule techniques. *Proc Natl Acad Sci U S A* 98(19):10682–10686.
- Li H., Wang H.C., Cao Y., Sharma D. & Wang M. (2008b) Configurational entropy modulates the mechanical stability of protein GB1. *J Mol Biol* 379(4):871–880.
- Li L., Kang D.G. & Cha H.J. (2004) Functional display of foreign protein on surface of *Escherichia coli* using N-terminal domain of ice nucleation protein. *Biotechnol Bioeng* 85(2):214–221.
- Linke W.A., Rudy D.E., Centner T., Gautel M., Witt C., Labeit S. & Gregorio C.C. (1999) I-band titin in cardiac muscle is a three-element molecular spring and is critical for maintaining thin filament structure. *J Cell Biol* 146(3):631–644.

- Lu H., Isralewitz B., Krammer A., Vogel V. & Schulten K. (1998) Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys J* 75(2):662–671.
- Lu Y., Zhang Y.H.P. & Lynd L.R. (2006) Enzyme-microbe synergy during cellulose hydrolysis by *Clostridium thermocellum*. *Proc Natl Acad Sci U S A* 103(44):16165–16169.
- Lynd L.R., Weimer P.J., van Zyl W.H. & Pretorius I.S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66(3):506–77.
- Lynd L.R., van Zyl W.H., McBride J.E. & Laser M. (2005) Consolidated bioprocessing of cellulosic biomass: an update. *Curr Opin Biotechnol* 16(5):577–583.
- Lytle B.L., Volkman B.F., Westler W.M. & Wu J.H. (2000) Secondary structure and calcium-induced folding of the *Clostridium thermocellum* dockerin domain determined by NMR spectroscopy. *Arch Biochem Biophys* 379(2):237–244.
- Lytle B.L., Volkman B.F., Westler W.M., Heckman M.P. & Wu J.H. (2001) Solution structure of a type I dockerin domain, a novel prokaryotic, extracellular calcium-binding domain. *J Mol Biol* 307(3):745–753.
- Madkour M. & Mayer F. (2003) Structural organization of the intact bacterial cellulosome as revealed by electron microscopy. *Cell Biol Int* 27(10):831–836.
- Maillard R.A., Chistol G., Sen M., Righini M., Tan J., Kaiser C.M., Hodges C., Martin A. & Bustamante C. (2011) ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* 145(3):459–469.
- Marcos. *Bacteria in Shear Flow*. PhD thesis MIT (2011).
- Marko J.F. & Siggia E.D. (1995) Stretching DNA. *Macromolecules* 28:8759–8770.
- Marszalek P.E., Lu H., Li H., Carrion-Vazquez M., Oberhauser A.F., Schulten K. & Fernandez J.M. (1999) Mechanical unfolding intermediates in titin modules. *Nature* 402(6757):100–103.
- Mayer F., Coughlan M.P., Mori Y. & Ljungdahl L.G. (1987) Macromolecular organization of the cellulolytic enzyme complex of *Clostridium thermocellum* as revealed by electron microscopy. *Appl Environ Microbiol* 53(12):2785–2792.
- McBee R.H. (1954) The characteristics of *Clostridium thermocellum*. *J Bacteriol* 67(4):505–506.
- Merkel R., Nassoy P., Leung A., Ritchie K. & Evans E. (1999) Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* 397(6714):50–53.
- Mingardon F., Chanal A., López-Contreras A.M., Dray C., Bayer E.A. & Fierobe H.P. (2007a) Incorporation of fungal cellulases in bacterial minicellulosomes yields viable, synergistically acting cellulolytic complexes. *Appl Environ Microbiol* 73(12):3822–3832.

- Mingardon F., Chanal A., Tardif C., Bayer E.A. & Fierobe H.P. (2007b) Exploration of new geometries in cellulosome-like chimeras. *Appl Environ Microbiol* 73(22):7138–7149.
- Miroux B. & Walker J.E. (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* 260(3):289–298.
- Morag E., Lapidot A., Govorko D., Lamed R., Wilchek M., Bayer E.A. & Shoham Y. (1995) Expression, purification, and characterization of the cellulose-binding domain of the scaffoldin subunit from the cellulosome of *Clostridium thermocellum*. *Appl Environ Microbiol* 61(5):1980–1986.
- Morag E., Yaron S., Lamed R., Kenig R., Shoham Y. & Bayer E.A. (1996) Dissociation of the cellulosome of *Clostridium thermocellum* under nondenaturing conditions. *Journal of Biotechnology* 51:235–242.
- Moraïs S., Barak Y., Caspi J., Hadar Y., Lamed R., Shoham Y., Wilson D.B. & Bayer E.A. (2010) Cellulase-xylanase synergy in designer cellulosomes for enhanced degradation of a complex cellulosic substrate. *MBio* 1(5).
- Moraïs S., Morag E., Barak Y., Goldman D., Hadar Y., Lamed R., Shoham Y., Wilson D.B. & Bayer E.A. (2012) Deconstruction of lignocellulose into soluble sugars by native and designer cellulosomes. *MBio* 3(6).
- Muñoz-Gutiérrez I., Oropeza R., Gosset G. & Martínez A. (2012) Cell surface display of a  $\beta$ -glucosidase employing the type V secretion system on ethanologenic *Escherichia coli* for the fermentation of cellobiose to ethanol. *J Ind Microbiol Biotechnol* 39(8):1141–1152.
- Neuman K.C. & Nagy A. (2008) Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Methods* 5(6):491–505.
- Ng S.P., Rounsevell R.W.S., Steward A., Geierhaas C.D., Williams P.M., Paci E. & Clarke J. (2005) Mechanical unfolding of TNfn3: the unfolding pathway of a fniii domain probed by protein engineering, AFM and MD simulation. *J Mol Biol* 350(4):776–789.
- Ng S.P., Randles L.G. & Clarke J. (2007) Single molecule studies of protein folding using atomic force microscopy. *Methods Mol Biol* 350:139–167.
- Noach I., Frolow F., Jakoby H., Rosenheck S., Shimon L.W., Lamed R. & Bayer E.A. (2005) Crystal structure of a type-II cohesin module from the *Bacteroides cellulosolvens* cellulosome reveals novel and distinctive secondary structural elements. *J Mol Biol* 348(1):1–12.
- Noach I., Levy-Assaraf M., Lamed R., Shimon L.J.W., Frolow F. & Bayer E.A. (2010) Modular arrangement of a cellulosomal scaffoldin subunit revealed from the crystal structure of a cohesin dyad. *J Mol Biol* 399(2):294–305.

- Oberhauser A.F., Marszalek P.E., Erickson H.P. & Fernandez J.M. (1998) The molecular elasticity of the extracellular matrix protein tenascin. *Nature* 393(6681):181–185.
- Oberhauser A.F., Marszalek P.E., Carrion-Vazquez M. & Fernandez J.M. (1999) Single protein misfolding events captured by atomic force microscopy. *Nat Struct Biol* 6(11):1025–1028.
- Oberhauser A.F., Badilla-Fernandez C., Carrion-Vazquez M. & Fernandez J.M. (2002) The mechanical hierarchies of fibronectin observed with single-molecule AFM. *J Mol Biol* 319(2):433–447.
- Olson D.G., McBride J.E., Shaw A.J. & Lynd L.R. (2012) Recent progress in consolidated bioprocessing. *Curr Opin Biotechnol* 23(3):396–405.
- Oroz J., Valbuena A., Vera A.M., Mendieta J., Gómez-Puertas P. & Carrión-Vázquez M. (2011) Nanomechanics of the cadherin ectodomain: “canalization” by Ca<sup>2+</sup> binding results in a new mechanical element. *J Biol Chem* 286(11):9405–9418.
- Oroz J., Hervás R. & Carrión-Vázquez M. (2012) Unequivocal single-molecule force spectroscopy of proteins by AFM using pFS vectors. *Biophys J* 102(3):682–690.
- Pagés S., Bélaïch A., Bélaïch J.P., Morag E., Lamed R., Shoham Y. & Bayer E.A. (1997) Species-specificity of the cohesin-dockerin interaction between *Clostridium thermocellum* and *Clostridium cellulolyticum*: prediction of specificity determinants of the dockerin domain. *Proteins* 29(4):517–527.
- Park M., Sun Q., Liu F., DeLisa M.P. & Chen W. (2014) Positional assembly of enzymes on bacterial outer membrane vesicles for cascade reactions. *PLoS One* 9(5):e97103.
- Peer A., Smith S.P., Bayer E.A., Lamed R. & Borovok I. (2009) Noncellulosomal cohesin- and dockerin-like modules in the three domains of life. *FEMS Microbiol Lett* 291(1):1–16.
- Pinheiro B.A., Proctor M.R., Martinez-Fleites C., Prates J.A.M., Money V.A., Davies G.J., Bayer E.A., Fontesm C.M.G.A., Fierobe H.P. & Gilbert H.J. (2008) The *Clostridium cellulolyticum* dockerin displays a dual binding mode for its cohesin partner. *J Biol Chem* 283(26):18422–18430.
- Qu W., Xue Y. & Ding Q. (2015) Display of fungi xylanase on *Escherichia coli* cell surface and use of the enzyme in xylan biodegradation. *Curr Microbiol* 70(6):779–785.
- Quiroz-Castañeda R.E. & Folch-Mallol J.L. (2013) *Hydrolysis of biomass mediated by cellulases for the production of sugars en: Sustainable degradation of lignocellulosic biomass - techniques, applications and commercialization*. InTech.
- Raman B., McKeown C.K., Rodriguez M., Brown S.D. & Mielenz J.R. (2011) Transcriptomic analysis of *Clostridium thermocellum* ATCC 27405 cellulose fermentation. *BMC Microbiol* 11:134.

- Reese E.T., Siu R.G.H. & Levinson H.S. (1950) The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. *J Bacteriol* 59(4):485–497.
- Rief M., Gautel M., Oesterhelt F., Fernandez J.M. & Gaub H.E. (1997) Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* 276(5315):1109–1112.
- Rief M., Pascual J., Saraste M. & Gaub H.E. (1999) Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J Mol Biol* 286(2):553–561.
- Rief M., Fernandez J.M. & Gaub H.E. (1998) Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys Rev Lett* 81(21):4764–4767.
- Ritort F. (2006) Single-molecule experiments in biological physics: methods and applications. *J Phys Condens Matter* 18(32):R531–R583.
- Salama-Alber O., Jobby M.K., Chitayat S., Smith S.P., White B.A., Shimon L.J.W., Lamed R., Frolow F. & Bayer E.A. (2013) Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *J Biol Chem* 288(23):16827–16838.
- Sali A. & Blundell T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779–815.
- Sambrook J. & Russell D.W. (2001) *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press.
- Schoeler C., Malinowska K.H., Bernardi R.C., Milles L.F., Jobst M.A., Durner E., Ott W., Fried D.B., Bayer E.A., Schulten K., Gaub H.E. & Nash M.A. (2014) Ultrastable cellulosome-adhesion complex tightens under load. *Nat Commun* 5:5635.
- Shimon L.J., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y. & Frolow F. (1997) A cohesin domain from *Clostridium thermocellum*: the crystal structure provides new insights into cellulosome assembly. *Structure* 5(3):381–390.
- Sievers F. & Higgins D.G. (2014) Clustal omega. *Curr Protoc Bioinformatics* 48:3.13.1–3.13.16.
- Sikora M., Sułkowska J.I. & Cieplak M. (2009) Mechanical strength of 17,134 model proteins and cysteine slipknots. *PLoS Comput Biol* 5(10):e1000547.
- Sippl M.J. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins* 17(4):355–362.
- Slutzki M., Barak Y., Reshef D., Schueler-Furman O., Lamed R. & Bayer E.A. (2012) Indirect ELISA-based approach for comparative measurement of high-affinity cohesin-dockerin interactions. *J Mol Recognit* 25(11):616–622.

- Smith S.P. & Bayer E.A. (2013) Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Curr Opin Struct Biol* 23(5):686–694.
- Spinelli S., Fiérobe H.P., Belaïch A., Belaïch J.P., Henrissat B. & Cambillau C. (2000) Crystal structure of a cohesin module from *Clostridium cellulolyticum*: implications for dockerin recognition. *J Mol Biol* 304(2):189–200.
- Stahl S.W., Nash M.A., Fried D.B., Slutzki M., Barak Y., Bayer E.A. & Gaub H.E. (2012) Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proc Natl Acad Sci U S A* 109(50):20431–20436.
- Stern J., Kahn A., Vazana Y., Shamshoum M., Moraïs S., Lamed R. & Bayer E.A. (2015) Significance of relative position of cellulases in designer cellulosomes for optimized cellulolysis. *PLoS One* 10(5):e0127326.
- Steward A., Toca-Herrera J.L. & Clarke J. (2002) Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy. *Protein Sci* 11(9):2179–2183.
- Still W.C., Tempczyk A., Hawley R.C. & Hendrickson T. (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–6129.
- Sułkowska J.I. & Cieplak M. (2008) Stretching to understand proteins - a survey of the protein data bank. *Biophys J* 94(1):6–13.
- Tanaka T. & Kondo A. (2015) Cell surface engineering of industrial microorganisms for biorefining applications. *Biotechnol Adv.*
- Tavares G.A., Béguin P. & Alzari P.M. (1997) The crystal structure of a type I cohesin domain at 1.7 Å resolution. *J Mol Biol* 273(3):701–713.
- Tilbeurgh H.V., Tomme P., Claeysens M., Bhikhabhai R. & Pettersson G. (1986) Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. separation of functional domains. *FEBS* 204(2):223–227.
- Tomme P., Tilbeurgh H.V., Pettersson G., Damme J.V., Vandekerckhove J., Knowles J., Teeri T. & Claeysens M. (1988) Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur J Biochem* 170(3):575–581.
- Tsui V. & Case D.A. (2000) Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* 56(4):275–291.
- Valbuena A., Oroz J., Vera A.M., Gimeno A., Gómez-Herrero J. & Carrión-Vázquez M. (2007) Quasi-simultaneous imaging/pulling analysis of single polyprotein molecules by atomic force microscopy. *Rev Sci Instrum* 78(11):113707.

- Valbuena A., Oroz J., Hervás R., Vera A.M., Rodríguez D., Menéndez M., Sułkowska J.I., Cieplak M. & Carrión-Vázquez M. (2009) On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proc Natl Acad Sci U S A* 106(33): 13791–13796.
- Vazana Y., Barak Y., Unger T., Peleg Y., Shamshoum M., Ben-Yehezkel T., Mazor Y., Shapiro E., Lamed R. & Bayer E.A. (2013) A synthetic biology approach for evaluating the functional contribution of designer cellulosome components to deconstruction of cellulosic substrates. *Biotechnol Biofuels* 6(1):182.
- Voronov-Goldman M., Levy-Assaraf M., Yaniv O., Wisserman G., Jindou S., Borovok I., Bayer E.A., Lamed R., Shimon L.J.W. & Frolow F. (2014) Structural characterization of a novel autonomous cohesin from *Ruminococcus flavefaciens*. *Acta Crystallogr F Struct Biol Commun* 70(Pt 4):450–456.
- Watanabe K., Nair P., Labeit D., Kellermayer M.S.Z., Greaser M., Labeit S. & Granzier H. (2002) Molecular mechanics of cardiac titin's PEVK and N2B spring elements. *J Biol Chem* 277(13):11549–11558.
- Wei H., Xu Q., Taylor L.E., Baker J.O., Tucker M.P. & Ding S.Y. (2009) Natural paradigms of plant cell wall degradation. *Curr Opin Biotechnol* 20(3):330–338.
- Wei H., Fu Y., Magnusson L., Baker J.O., Maness P.C., Xu Q., Yang S., Bowersox A., Bogorad I., Wang W., Tucker M.P., Himmel M.E. & Ding S.Y. (2014) Comparison of transcriptional profiles of *Clostridium thermocellum* grown on cellobiose and pretreated yellow poplar using RNA-Seq. *Front Microbiol* 5:142.
- White B.A., Lamed R., Bayer E.A. & Flint H.J. (2014) Biomass utilization by gut microbiomes. *Annu Rev Microbiol* 68:279–296.
- Wiederstein M. & Sippl M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35(Web Server issue):W407–W410.
- Wilson C.A. & Wood T.M. (1992) The anaerobic fungus *Neocallimastix frontalis*: isolation and properties of a cellulosome-type enzyme fraction with the capacity to solubilize hydrogen-bond-ordered cellulose. *Appl Microbiol Biotechnol* 37:125–129.
- Wojciechowski M., Thompson D. & Cieplak M. (2014) Mechanostability of cohesin-dockerin complexes in a structure-based model: Anisotropy and lack of universality in the force profiles. *J Chem Phys* 141(24):245103.
- Xu J. & Smith J.C. (2010) Probing the mechanism of cellulosome attachment to the *Clostridium thermocellum* cell surface: computer simulation of the type II cohesin-dockerin complex and its variants. *Protein Eng Des Sel* 23(10):759–768.
- Xu Q., Gao W., Ding S.Y., Kenig R., Shoham Y., Bayer E.A. & Lamed R. (2003) The cellulosome system of *Acetivibrio cellulolyticus* includes a novel type of adaptor protein and a cell surface anchoring protein. *J Bacteriol* 185(15):4548–4557.



- Xu Q., Barak Y., Kenig R., Shoham Y., Bayer E.A. & Lamed R. (2004) A novel *Acetivibrio cellulolyticus* anchoring scaffoldin that bears divergent cohesins. *J Bacteriol* 186(17): 5782–5789.
- Yaron S., Morag E., Bayer E.A., Lamed R. & Shoham Y. (1995) Expression, purification and subunit-binding properties of cohesins 2 and 3 of the *Clostridium thermocellum* cellulosome. *FEBS Lett* 360(2):121–124.
- Zhang Y.H.P. & Lynd L.R. (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnol Bioeng* 88(7):797–824.

Parte VII

APÉNDICE

## RESUMEN DE LAS FUERZAS DE DESPLEGAMIENTO

Módulo	Región	$F_D$ SMFS	$F_D$ SMD 300 K	$F_D$ SMD 340 K
CtA1	Externa	$124 \pm 25$	$560 \pm 80$	$375 \pm 48$
CtA2	Externa	$214 \pm 41^*$	$470 \pm 80^*$	ND
CtA7	Conectora	$480 \pm 77^*$	$710 \pm 120^*$	ND
CtA9	Conectora	$460 \pm 110$	$650 \pm 52$	$497 \pm 87$
CtA2	Externa	$179 \pm 133$	$470 \pm 80^*$	ND
CtA2: Cel8A	Externa	$195 \pm 42$	$489 \pm 51$	$375 \pm 72$
CtA2: Cel8Arev	Externa	ND**	$476 \pm 91$	$456 \pm 63$
CtA7	Conectora	$515 \pm 59$	$710 \pm 120^*$	ND
CtA7: Cel8A	Conectora	$506 \pm 87$	$621 \pm 60$	$567 \pm 105$
CtA7l	Conectora	$529 \pm 82$	ND	ND

Cuadro 1: Resumen de las fuerzas de desplegamiento obtenidas para cohesinas de *C. thermocellum*. Todos los valores se dan en pN. \* indica los resultados que provienen de [Valbuena et al., 2009]. \*\* en los experimentos de AFM no se puede distinguir el modo de unión de Cel8A.

Módulo	$F_D$ SMFS	$F_D$ SMD 300 K
CcC1	$425 \pm 53$	$630 \pm 70^*$
CcC3	$592 \pm 125$	$733 \pm 133$
CcC8	$346 \pm 58$	$568 \pm 86$
CcC1l	$398 \pm 47$	ND
CcC123	$404 \pm 64$	ND
	$627 \pm 79$	ND

Cuadro 2: Resumen de las fuerzas de desplegamiento obtenidas para cohesinas de *C. cellulolyticum*. Todos los valores se dan en pN. \* indica los resultados que provienen de [Valbuena et al., 2009].

Módulo	Región	$F_D$ SMFS
AcA <sub>3</sub>	Externa	209 ± 54
AcA <sub>4</sub>	Conectora	442 ± 49
AcB <sub>4</sub>	Conectora*	259 ± 39
AcC <sub>3</sub>	Conectora*	455 ± 51

Cuadro 3: Resumen de las fuerzas de desplegamiento obtenidas para cohesinas de *A. cellulolyticus*. Todos los valores se dan en pN. \* indica los resultados las cohesinas que se encuentran entre los dos puntos de anclado aunque no son de la escafoldina primaria.