



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Diseño de núcleo-cápsidas sintéticas tipo virus con autoensamblaje programado por endonucleasas CRISPR-Cas

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

Carlos Calcines Cruz

TUTOR PRINCIPAL

Dr. Armando Hernández García
[Instituto de Química](#)

MIEMBROS DEL COMITÉ TUTOR
Dra. Laura Alicia Palomares Aguilera
[Instituto de Biotecnología](#)

Dr. Ismael Bustos Jaimes
[Facultad de Medicina](#)

Ciudad de México. Agosto, 2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Jurado revisor

Presidente:

Adela Rodríguez Romero

Vocales:

Alfredo Torres Larios

Miguel Antonio Costas Basín

Adán Oswaldo Guerrero Cárdenas

Secretario:

Daniel Alejandro Fernández Velasco

Agradecimientos

Al Dr. Armando Hernández García (Departamento de Química de Biomacromoléculas, Instituto de Química, UNAM) por supervisar el presente proyecto de Doctorado y la escritura del manuscrito derivado. También al grupo de investigación por su apoyo y los tantos debates.

A los miembros del Comité Tutorial, Dra. Laura Alicia Palomares Aguilera (Instituto de Biotecnología) y Dr. Ismael Bustos Jaimes (Facultad de Medicina), por sus excelentes críticas y discusiones sobre mi trabajo.

A la Dra. Patricia Cano Sánchez (Departamento de Química de Biomacromoléculas, Instituto de Química, UNAM) por su invaluable apoyo durante los experimentos de clonación y purificación de proteínas.

A las y los técnicos de las facilidades del IQ: Eréndira García Ríos, M. en C. Lucía del Carmen Márquez Alonso, M. en C. Lucero Mayra Ríos Ruiz, M. en C. María Teresa Obdulia Ramírez Apan, María de la Paz Orta Pérez y Dra. Annia Rodríguez Hernández. Su apoyo fue esencial durante los experimentos de purificación de proteínas, espectrometría de masas y dicroísmo circular.

Al Dr. Ilya J. Finkelstein (Departamento de Biociencias Moleculares, Universidad de Texas en Austin) por supervisar los experimentos de cortinas de DNA y la escritura del manuscrito derivado de la tesis. También a su grupo de investigación por la importante retroalimentación y apoyo durante mi estancia.

Agradezco al Conacyt, al Programa de Apoyo a los Estudios del Posgrado (PAEP) y a la iniciativa binacional ConTex (Conacyt-Universidad de Texas) por los apoyos económicos para la realización de mi Doctorado y mi estancia en UT-Austin.

Índice

Abreviaturas	1
Lista de figuras	2
Lista de tablas.....	4
Resumen	5
Abstract	6
Introducción	7
Partículas tipo-virus artificiales como frontera en la viromimética.....	7
Antecedentes.....	9
C-S _n -B: proteínas viromiméticas <i>de novo</i>	9
Mecanismo de ensamblaje de C-S ₁₀ -B sobre el DNA.....	11
Señales de empaquetamiento viral	13
Endonucleasas de CRISPR para programar el ensamblaje de C-S ₁₀ -B	15
Diseño de núcleo-cápsidas artificiales que se ensamblan sobre secuencias específicas de DNA	16
Justificación y planteamiento del problema.....	19
Hipótesis.....	20
Objetivos.....	20
Materiales y métodos.....	21
Enzimas, oligos y plásmidos	21
Construcción del gen codificante para dCas12a _(FKBP) para expresión en <i>E. coli</i>	21
Construcción del gen codificante para C-S _{10(FRB)} para expresión en <i>P. pastoris</i>	22
Expresión y purificación de dCas12a y dCas12a _(FKBP)	22
Expresión y purificación de C-S ₁₀ -B y C-S _{10(FRB)}	24
Expresión y purificación de la proteasa SUMO.....	25
Diseño y producción de crRNAs.....	25
Caracterización estructural de dCas12a y dCas12a _(FKBP) por difracción circular	26
Dimerización de dCas12a _(FKBP) y C-S _{10(FRB)} inducida por rapamicina	27
Interacción de dCas12a con el crRNA.....	27
Unión de la ribonucleoproteína (dCas12a-crRNA) a DNA y visualización de las núcleo-cápsidas.....	27
Encapsidación del DNA por C-S ₁₀ -B.....	28
Preparación de sustratos de DNA para los ensayos de cortinas	28

Ensamblaje de las núcleo-cápsidas en tiempo real	30
Obtención de la estructura 3D esperada para el complejo dCas12a-S ₁₀	31
Análisis estadístico.....	31
Resultados y discusión	33
Obtención de los bloques de construcción para las núcleo-cápsidas	33
Ensamblaje de las núcleo-cápsidas artificiales en tiempo real	41
Decoración del DNA con dCas12a y ensamblaje de las núcleo-cápsidas	48
dCas12a-S ₁₀ acelera el ensamblaje de las núcleo-cápsidas	57
Ensamblaje de C-S ₁₀ -B sobre moléculas diana de DNA.....	62
Consideraciones finales	65
Conclusiones	66
Perspectivas	67
Bibliografía.....	68
Anexos	76
Producción científica.....	79

Abreviaturas

AFM: Microscopía de fuerza atómica

AOX1: Alcohol oxidasa 1

BSA: Albúmina de suero bovino

Cas12a: Endonucleasa asociada a CRISPR 12a (antes conocida como Cpf1)

CI: Intervalos de confianza

CRISPR: Repeticiones palindrómicas cortas, agrupadas y regularmente interespaciadas (inglés: *clustered regularly interspaced short palindromic repeats*)

crRNA: CRISPR RNA

DO_{600nm}: Densidad óptica a 600 nm

DTT: Ditioneitol

EMSA: Ensayo de retardo en la movilidad electroforética

FKBP: Proteína de unión a FK506 o rapamicina

FRB: Proteína de unión a FKBP-rapamicina

IPTG: Isopropil- β -D-1-tiogalactopiranosido

Medio LB: Medio de cultivo Luria Bertani

Medio TB: Medio de cultivo Terrific Broth

Mut*: Fenotipo positivo de utilización de metanol

N/P: Razón molar entre cargas positivas (N) del dominio B de C-S₁₀-B y las cargas negativas del DNA (P)

Ni-NTA: Níquel-ácido nitrilotriacético

PAM: Motivo adyacente al espaciador

PBS: Amortiguador de fosfato salino

PCR: Reacción en cadena de la polimerasa

PMSF: Fluoruro de fenilmetilsulfonilo

QD: Punto cuántico

RNP: Ribonucleoproteína o complejo ribonucleoproteico

SDS-PAGE: Electroforesis en gel de poliacrilamida con dodecilsulfato de sodio

SN: Sobrenadante de cultivo

SUMO: Modificador pequeño tipo ubiquitina

t_{half}: Tiempo en el que se alcanza la mitad de la condensación del DNA

UA: Unidades arbitrarias

Urea-PAGE: Electroforesis en gel de poliacrilamida con urea

VLP: Partícula tipo-virus

Lista de figuras

Figura 1. Número de publicaciones sobre virus artificiales en PubMed.	8
Figura 2. Las proteínas C-S _n -B consisten en tres bloques o dominios con función definida.	9
Figura 3. El número de repeticiones en el dominio S ₁₀ determina la fuerza de interacción entre monómeros de C-S _n -B y el grado de compactación del DNA.	10
Figura 4. La proteína C-S ₁₀ -B se ensambla sobre el DNA con mecanismo de nucleación-alargamiento.	12
Figura 5. Múltiples señales de encapsidación promueven el empaquetamiento del genoma viral.	13
Figura 6. dCas12a se une al DNA y forma un lazo R de crRNA-DNA.	16
Figura 7. Proteínas de fusión que forman el complejo dCas12a-S ₁₀ .	17
Figura 8. El complejo dCas12a-S ₁₀ permitiría controlar y dirigir la nucleación de las núcleo-cápsidas artificiales sobre el DNA.	18
Figura 9. Estrategias de clonación para crear los genes codificantes de las proteínas quiméricas dCas12a _(FKBP) y C-S ₁₀ _(FRB) .	34
Figura 10. Expresión de dCas12a y dCas12a _(FKBP) en <i>E. coli</i> .	35
Figura 11. Purificación de dCas12a y dCas12a _(FKBP) .	36
Figura 12. Caracterización estructural de dCas12a y dCas12a _(FKBP) por difracción circular.	37
Figura 13. Expresión en <i>P. pastoris</i> y purificación de las proteínas C-S ₁₀ -B y C-S ₁₀ _(FRB) .	38

Figura 14. dCas12a _(FKBP) y C-S _{10(FRB)} se heterodimerizan en presencia de rapamicina.	39
Figura 15. Síntesis de los CRISPR RNA (crRNA) para la decoración del DNA con dCas12a.	40
Figura 16. dCas12a une el crRNA y forma la ribonucleoproteína (RNP) para la unión específica al DNA.	40
Figura 17. C-S _{10-B} se autoensambla sobre el DNA y lo empaqueta en núcleo-cápsidas.	42
Figura 18. La concentración de C-S _{10-B} regula la velocidad de compactación del DNA.	45
Figura 19. La concentración de C-S _{10-B} regula el tiempo necesario para alcanzar la mitad del ensamblaje y su cooperatividad	46
Figura 20. La concentración de C-S _{10-B} regula la velocidad de su unión al DNA.	47
Figura 21. dCas12a se posiciona en sus sitios diana sobre el DNA vía <i>pools</i> de crRNAs.	48
Figura 22. La decoración del DNA con dCas12a acelera la encapsidación por C-S _{10-B} .	49
Figura 23. dCas12a favorece la nucleación de las núcleo-cápsidas.	51
Figura 24. La decoración con dCas12a restringe la difusión lineal de C-S _{10-B} sobre el DNA.	52
Figura 25. Los nucleosomas se posicionan sobre el DNA sin especificidad de secuencia.	53
Figura 26. La decoración con nucleosomas acelera la encapsidación del DNA por C-S _{10-B} .	54

Figura 27. Los nucleosomas favorecen la nucleación de las núcleo-cápsidas.	55
Figura 28. El complejo dCas12a-S ₁₀ se posiciona en los sitios diana sobre el DNA vía <i>pools</i> de crRNA.	57
Figura 29. La decoración del DNA con dCas12a-S ₁₀ favorece el autoensamblaje de las núcleo-cápsidas	58
Figura 30. La decoración del DNA con dCas12a-S ₁₀ favorece la encapsidación en concentraciones sub-saturantes de C-S ₁₀ -B.	60
Figura 31. C-S ₁₀ -B empaqueta el DNA, y el DNA decorado con dCas12a y dCas12a-S ₁₀ en núcleo-cápsidas artificiales.	61
Figura 32. C-S ₁₀ -B encapsida selectivamente el DNA que ha sido decorado con dCas12a-S ₁₀ .	63

Lista de tablas

Tabla S1. Secuencia de la región variable de los crRNAs usados para posicionar dCas12a y dCas12a-S ₁₀ en el DNA.	76
Tabla S2. Parámetros del ajuste de curvas a la suma de múltiples Gaussianas.	78

Resumen

Las proteínas viromiméticas *de novo* permiten el desarrollo de sistemas de entrega de genes, y de modelos simples para estudiar el ensamblaje viral. Previamente se reportaron partículas tipo-virus (VLPs)—formadas por el ensamblaje del polipéptido C-S₁₀-B sobre el DNA—que resisten el ataque de nucleasas y pueden transfectar células. Sin embargo, el autoensamblaje de proteínas viromiméticas (como C-S₁₀-B) en secuencias específicas de ácidos nucleicos—una propiedad crucial de los virus—ha sido difícil de reproducir. Inspirados en las señales de empaquetamiento viral, hipotetizamos que proteínas que unen DNA con especificidad de secuencia (como CRISPR-dCas12a) y que recluten C-S₁₀-B permitirán la encapsidación de moléculas de DNA diana.

Caracterizamos las cinéticas de compactación del DNA en distintas concentraciones de C-S₁₀-B y seleccionamos aquellas (≤ 50 nM) que limitan el ensamblaje de VLPs. Encontramos que la decoración del DNA con dCas12a nuclea las VLPs y acelera la encapsidación del DNA de dos a tres veces con respecto al DNA no decorado. La decoración del DNA con nucleosomas tuvo el mismo efecto que dCas12a lo que indica que el arresto de la difusión lineal de C-S₁₀-B facilita la nucleación de las núcleo-cápsidas. La decoración del DNA con el complejo dCas12a-S₁₀ (obtenido por dimerización de dCas12a con el dominio de polimerización S₁₀) aceleró la encapsidación dos y cinco veces con respecto al DNA con dCas12a y al DNA no decorado, respectivamente. La decoración con múltiples dCas12a-S₁₀ favoreció la condensación del DNA diana en presencia de DNA no diana.

Estas VLPs programables por CRISPR pueden ser un punto de partida para crear núcleo-cápsidas que empaquetan moléculas específicas de DNA o RNA, y con múltiples usos en bio(nano)tecnología.

Abstract

De novo viromimetic proteins allow the fabrication of safe gene delivery systems and simple models to study viral self-assembly. We have previously reported virus-like particles (VLPs) made of the synthetic triblock polypeptide C-S₁₀-B that protect encapsulated DNA from nucleases and can transfect cells. However, the self-assembly of viromimetic proteins on specific nucleic acids—a crucial property that enables viral replication and success—has remained elusive. Inspired by viral packaging signals, we reasoned that using sequence-specific and high affinity DNA binding proteins (such as CRISPR inactivated endonuclease dCas12a) that recruit C-S₁₀-B onto the DNA template might favor encapsulation of target DNA molecules.

First, we screened C-S₁₀-B concentrations and selected those (≤ 50 nM) that limit VLP self-assembly. At such low concentrations of C-S₁₀-B, we found that positioning up to ten dCas12a on the DNA increased DNA encapsidation rates two- to three-fold with respect to the naked DNA. The same effect was found with nucleosome-decorated DNA, which suggests that arresting the linear diffusion of C-S₁₀-B on the DNA promotes particle nucleation and growth. We coupled the polymerization block (S₁₀) from C-S₁₀-B to dCas12a via the rapamycin-induced heterodimerization of FKBP and FRB domains, so that dCas12a-S₁₀ recruits C-S₁₀-B to the DNA. This strategy increased VLP self-assembly rates two-fold relative to the dCas12a-decorated DNA, and five-fold relative to the nondecorated DNA. Furthermore, selective positioning of dCas12a-S₁₀ allowed discrimination of cognate versus non-cognate DNA in an equimolar DNA mixture.

These CRISPR-guided VLPs could be a starting point for the creation of novel and programmable templated self-assemblies with multiple uses in bio(nano)technology.

Introducción

Partículas tipo-virus artificiales como frontera en la viromimética

La diversificación de los virus en cuanto a hábitats, morfología y mecanismos moleculares supera a los de cualquier organismo uni- o multicelular (Wasik & Turner, 2013). Estos patógenos intracelulares obligados son la entidad biológica más abundante en la Tierra con estimados de 1×10^{31} unidades, dos órdenes de magnitud mayor al número de bacterias (1×10^{29}) (Microbiology by Numbers, 2011). Todos los virus se han adaptado por evolución natural para mantener su estructura en climas extremos, proteger y entregar a hospederos específicos su carga genética, y ensamblarse sobre su propio genoma para generar nuevos viriones.

El tamaño homogéneo y la estructura multimodular de los virus los hacen maquinarias con alto potencial bio-ingenieril pero su patogenicidad impide su aplicación. En cambio, las partículas tipo-virus (VLPs)—obtenidas por expresión de las proteínas virales en sistemas heterólogos—son idénticas a los virus, aunque carecen de genoma viral y no son infecciosas. Las VLPs se han usado como sistemas de entrega de genes y fármacos (Liu et al., 2020; Pinnapireddy et al., 2019), bionanomateriales (Dang et al., 2011; Sykora et al., 2015), nano-reactores (Ren et al., 2019) y vacunas para múltiples enfermedades virales incluyendo el SARS-CoV-2 (Xu et al., 2020).

La modificación de las proteínas que forman a las VLPs puede impedir su ensamblaje o resultar en partículas defectuosas, lo que dificulta su diversificación. Una alternativa a este problema es el uso de péptidos y proteínas *de novo* que se autoensamblan en nano-cajas (Bale et al., 2016). Esta estrategia tipo *bottom-up*, basada en el diseño de elementos o bloques de construcción básicos, es relativamente nueva en

el campo de la viromimética artificial y se ha desarrollado en las últimas dos décadas (Figura 1).

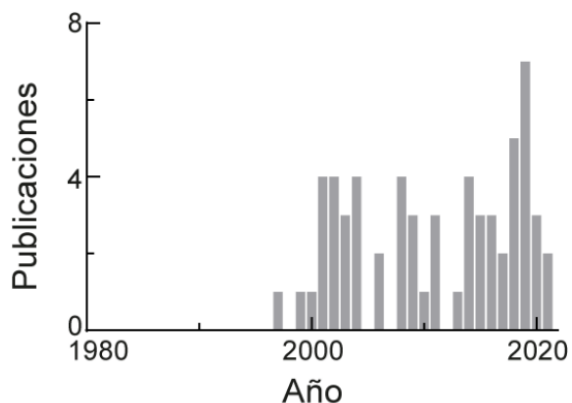


Figura 1. Número de publicaciones sobre virus artificiales en PubMed. La lista fue curada para incluir solo artículos o patentes que incluyen el término “virus artificiales” en el título o en el texto principal. Fuente: NCBI.

Las cápsidas formadas por proteínas o péptidos no virales son más flexibles que las VLPs porque sus propiedades estructurales y funcionales se programan con el diseño de los componentes. Tamaño, mecanismo de ensamblaje y estabilidad son algunas de las propiedades que se pueden manipular fácilmente (Azuma et al., 2018; Subramanian et al., 2018; Zhou et al., 2016).

Un desafío actual para las partículas tipo-virus artificiales es la encapsulación de secuencias específicas de DNA o RNA que codifiquen para las proteínas de la propia cápsida. Este acoplamiento entre genotipo y fenotipo permitiría la evolución de las partículas artificiales y su adaptación a ambientes bioquímicos complejos como la sangre (Butterfield et al., 2017). En este trabajo, usamos una estrategia *bottom-up* para dirigir a moléculas diana de DNA el ensamblaje de VLPs artificiales que encapsulaban DNA sin especificidad de secuencia.

Antecedentes

C-S_n-B: proteínas viromiméticas *de novo*

Inspirados en la proteína del virus del mosaico del tabaco, Hernandez-Garcia et al. (2014) diseñaron y produjeron un grupo de proteínas con nombre “C-S_n-B”. Estas proteínas contienen tres dominios o bloques que imitan las tres funciones principales de la proteína de la cápsida del virus del mosaico del tabaco. (1) El dominio C, que da estabilidad coloidal a la partícula, tiene estructura desordenada tipo colágeno y está compuesto por 400 residuos (35.8% aminoácidos polares en su mayoría no cargados, 33.7% glicinas y 22% prolinas) (Werten et al., 2001). (2) El dominio S_n está inspirado en la proteína de la seda, tiene *n* repeticiones de la secuencia [(AG)₃QG] que forman un solenoide de hebras β (denominado rollo β), y es responsable de la polimerización de C-S_n-B (Krejchi et al., 1994; Smeenk et al., 2005; Zhao et al., 2016). (3) El tercer dominio, B, consiste en doce lisinas que interactúan electrostáticamente con DNA, RNA y otros polianiones (Hernandez-Garcia et al., 2012; Martin & Rice, 2007; Zhou et al., 2021) (**Figura 2**).

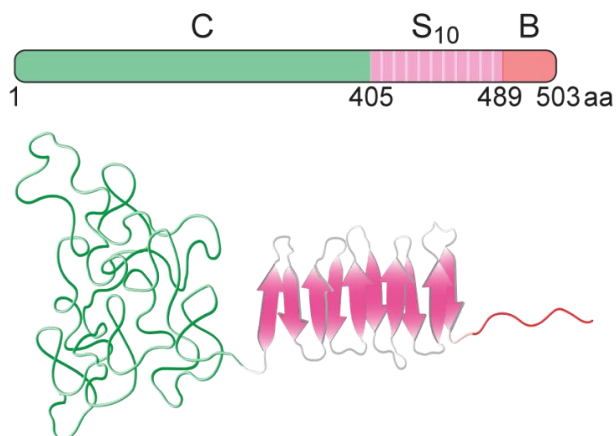


Figura 2. Las proteínas C-S_n-B consisten en tres bloques o dominios con función definida. Dibujo con la estructura trimodular de C-S_n-B, en este caso con diez repeticiones de la secuencia de la proteína de la seda (C-S₁₀-B). C: dominio de solubilidad, S: dominio de polimerización, y B: dominio de unión al DNA.

Los monómeros de C-S_n-B unidos al DNA interactúan entre sí por el dominio S_n y la fuerza de esta interacción depende del número de repeticiones *n* (Razzokov et al., 2018). A diferencia de C-S₀-B, C-S₂-B y C-S₄-B que solo recubren el DNA, las versiones C-S₁₀-B y C-S₁₄-B forman núcleo-cápsidas rígidas tipo rodillo (**Figura 3**).

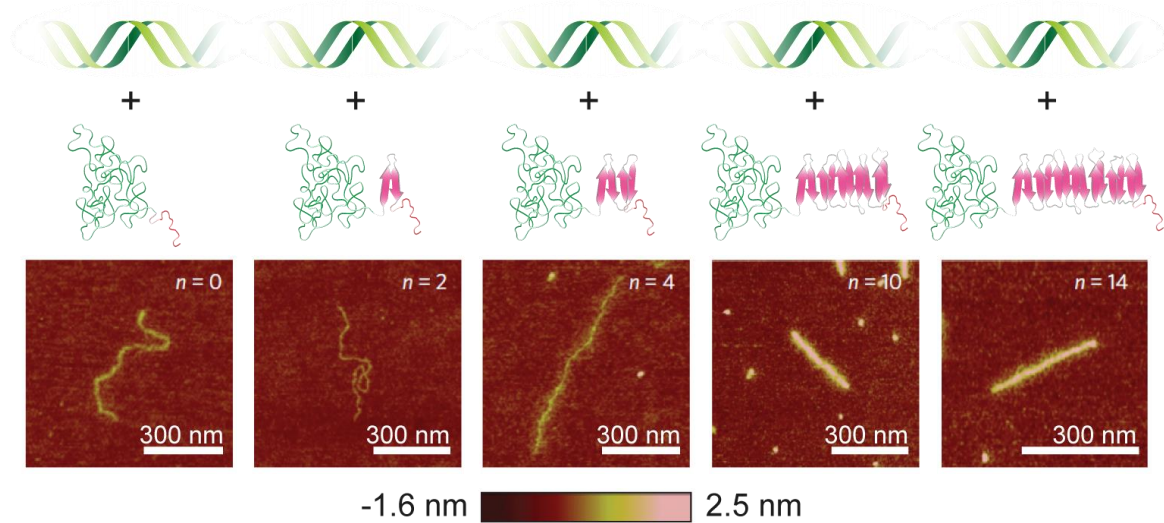


Figura 3. El número de repeticiones en el dominio S₁₀ determina la fuerza de interacción entre monómeros de C-S_n-B y el grado de compactación del DNA. Las proteínas C-S₀-B, C-S₂-B y C-S₄-B solo recubren el DNA mientras que C-S₁₀-B y C-S₁₄-B lo empaquetan en núcleo-cápsidas tipo rodillo. El número de repeticiones (*n*) del dominio S se indica en las micrografías. Las imágenes de microscopía de fuera atómica fueron tomadas de Hernandez-Garcia et al. (2014). La dimensión *z* (altura) se indica en la barra de color.

C-S₁₀-B y C-S₁₄-B se unen al DNA con velocidad similar (tienen energías libres de nucleación similares) y forman VLPs de longitud igual a un tercio de la longitud de contorno del DNA templado (Hernandez-Garcia et al., 2014; Punter et al., 2016). Sin embargo, C-S₁₄-B se agrega fuera del DNA con mayor facilidad que C-S₁₀-B debido al mayor tamaño del dominio central (Vargas et al., 2019). C-S₁₀-B ha sido más estudiada que C-S₁₄-B y es la que usamos en este trabajo.

Las núcleo-cápsidas formadas por C-S₁₀-B protegen el DNA o RNA encapsulado de la degradación por nucleasas, tienen tamaño homogéneo y definido por la longitud del templado, y pueden entregar DNA y RNA a células de mamíferos (Hernandez-Garcia et al., 2014; Jekhmane et al., 2017; Li et al., 2021). Cada uno de los módulos de C-S₁₀-B se pueden modificar fácilmente para favorecer la expresión, obtener responsividad a pH y cambiar la afinidad por DNA (Beun et al., 2014; Hernandez-Garcia et al., 2017; Willems et al., 2019).

Mecanismo de ensamblaje de C-S₁₀-B sobre el DNA

C-S₁₀-B se ensambla en núcleo-cápsidas tipo-rodillo con un mecanismo de nucleación-alargamiento similar al virus del mosaico del tabaco; descrito por un modelo de cremallera (Kraft et al., 2012; Punter et al., 2016) (**Figura 4**).

En experimentos de pinzas de DNA se ha observado que los monómeros u oligómeros de C-S₁₀-B se adsorben en sitios aleatorios sobre el DNA (por atracción electrostática entre los grupos fosfatos del DNA y las doce lisinas del dominio B) y difunden en una dimensión. El núcleo crítico se forma sobre el DNA cuando cinco monómeros de C-S₁₀-B se asocian vía puentes de hidrógeno entre los rollos β del dominio S₁₀. El núcleo es entonces una estructura compacta y rica en hebras β que recluta otros monómeros de C-S₁₀-B; a partir de él crece el rodillo en una o dos direcciones como el movimiento de una cremallera. Que el alargamiento de la partícula ocurra solo a partir de núcleos de formación lenta explica la cooperatividad en el ensamblaje de C-S₁₀-B (Cingil et al., 2017; Kraft et al., 2012; Marchetti et al., 2019; Van Galen et al., 2019; Zhao et al., 2016).

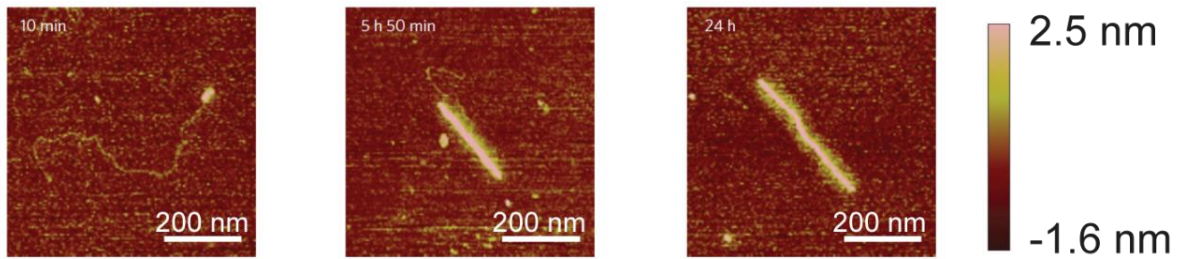


Figura 4. La proteína C-S₁₀-B se ensambla sobre el DNA con mecanismo de nucleación-alargamiento. Micrografías de fuerza atómica que muestran el ensamblaje de C-S₁₀-B en núcleo-cápsidas. Las imágenes se tomaron a 10 min, 5 h 50 min y 24 h de incubación de C-S₁₀-B con el DNA. Tomado de Hernandez-Garcia et al. (2014). La dimensión z (altura) se indica con la barra de color.

En templados de DNA cortos (2.5 kbp) C-S₁₀-B nuclea frecuentemente en uno de los extremos del DNA y la partícula se alarga en una dirección hasta la completitud (Hernandez-Garcia et al., 2014). Como explicación a este fenómeno, Punter et al. (2016) hipotetizan que el efecto estérico del dominio desestructurado C podría impedir la unión de C-S₁₀-B lejos de los extremos del DNA. Sin embargo, en templados largos (50 kbp) los núcleos se forman en sitios aleatorios sobre el DNA sin preferencia por los extremos (Marchetti et al., 2019). Nuestros resultados recientemente publicados indican que el arresto de la difusión de C-S₁₀-B en los extremos del DNA es una explicación más razonable para tal observación (Calcines-Cruz et al., 2021).

Independientemente del tamaño del DNA, la nucleación de C-S₁₀-B ocurre sin especificidad de secuencia; escenario que difiere de las cápsidas virales cuyo ensamblaje se nuclea sobre el genoma propio. Para programar el ensamblaje de C-S₁₀-B nos inspiramos en la forma con que los virus resuelven el problema del “reconocimiento propio” en presencia del exceso de ácidos nucleicos celulares.

Señales de empaquetamiento viral

Las señales de empaquetamiento garantizan el ensamblaje de los virus sobre su genoma propio. Estas son regiones en *cis* (i.e. secuencias del RNA o DNA viral) que reclutan las proteínas de la cápsida y disminuyen la barrera energética para la nucleación del ensamblaje (Borodavka et al., 2012; Chai et al., 1995; D'Souza et al., 2001; De Beer et al., 2002; Kutluay et al., 2014; Patel et al., 2017; Turner et al., 1988) (**Figura 5**).

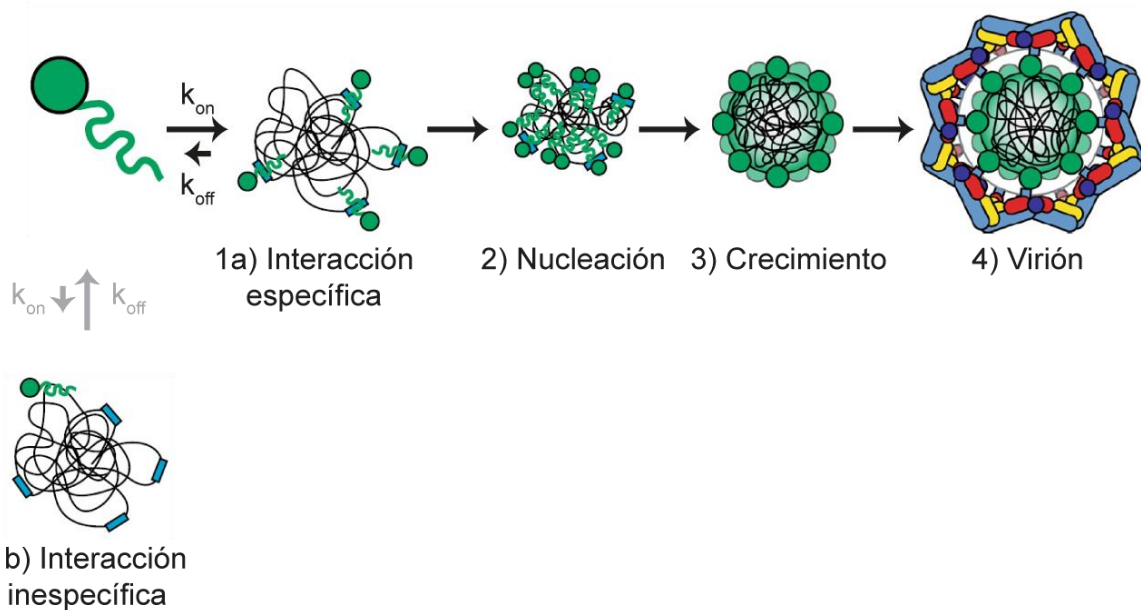


Figura 5. Múltiples señales de encapsidación promueven el empaquetamiento del genoma viral. 1a) La proteína de la cápsida (verde) interactúa establemente con los sitios de alta afinidad en el genoma viral. **1b)** La proteína de la cápsida puede interactuar con otros sitios en el genoma viral y los ácidos nucleicos de la célula hospedera. Estas interacciones son de baja afinidad y corta vida. **2)** La proteína de la cápsida se oligomeriza alrededor de los sitios de empaquetamiento y compactan el genoma. **3)** La partícula viral culmina el ensamblaje y **4)** forma viriones maduros. Tomado de Brown et al. (2020).

De una a múltiples señales de empaquetamiento se han reportado para virus de bacterias, plantas y animales; y su rol en el ensamblaje viral se ha demostrado con ensayos *in silico*, *in vitro* e *in vivo* (Chandler-Bostock et al., 2020; Comas-Garcia et al., 2017; Morton et al., 2010; Stewart et al., 2016; Umunnakwe et al., 2021). Por ejemplo, el dominio de unión a RNA de la poliproteína Gag del VIH-1 se une vía dedos de zinc a 28 ± 1 señales de empaquetamiento sobre el genoma viral, con afinidad de 40 nM a 1.4 μ M. De hecho, la molécula NSC260594—que inhibe el empaquetamiento del VIH-1—entorpece las interacciones de mayor afinidad entre Gag y el RNA (Ding et al., 2020; Ingemarsdotter et al., 2018).

Las señales de empaquetamiento son quizás el “talón de Aquiles” de los virus y su descubrimiento cambió el paradigma de que solo las interacciones electrostáticas son suficientes para un ensamblaje viral competente (Stockley et al., 2013; Twarock et al., 2018).

El papel de las señales de encapsidación se demostró recientemente para núcleo-cápsidas artificiales. Terasaka et al. (2018) demostraron que la fusión del péptido λ N+ (con afinidad $K_D = 5$ nM para una secuencia específica de RNA) a la proteína autoensamblante lumazina sintasa favoreció la encapsulación del RNA con dos sitios diana para λ N+. En trabajos posteriores con el mismo sistema, Tetter et al. (2021) describieron que la optimización de las interacciones entre el RNA y la proteína de fusión (lumazina sintasa- λ N+) incrementó la especificidad y la eficiencia de la encapsulación.

A partir de lo mencionado anteriormente, el ensamblaje eficiente y dirigido de las VLPs requiere interacciones específicas entre las proteínas de la cápsida y el ácido nucleico templado. Para las VLPs formadas por C-S₁₀-B, esto implica el uso de una proteína que una DNA de manera afín y discriminada, e idealmente con especificidad programable.

Endonucleasas de CRISPR para programar el ensamblaje de C-S₁₀-B

Las endonucleasas asociadas a CRISPR (Cas) han revolucionado el campo de la ingeniería genética. Las características más importantes que explican su éxito son la selectividad con que reconocen su DNA (o RNA) diana, y su alta programabilidad. A diferencia de otras proteínas que reconocen al DNA vía interacciones aminoácidos-bases nitrogenadas (como las nucleasas tipo activadores de transcripción o de dedos de zinc), las Cas usan un RNA guía (CRISPR RNA, crRNA) complementario a la secuencia diana (Gaj et al., 2013). La especificidad de secuencia de las proteínas Cas se modifica de manera sencilla y costo-eficiente con el diseño del crRNA, sin necesidad de mutar dominios proteicos (Jiang et al., 2013; Zetsche et al., 2017).

Las endonucleasas Cas forman parte del sistema bacteriano de inmunidad adaptativa contra bacteriófagos y se organizan en dos clases. Los sistemas CRISPR de clase 1 consisten en sistemas multiproteicos que unen y escinden el DNA patogénico. En cambio, en los sistemas de clase 2 una sola proteína une el DNA (o RNA diana para Cas13, de tipo VI) y realiza la interferencia. Las proteínas Cas más usadas para ingeniería genética (Cas9, Cas12, Cas13) pertenecen a la clase II (Jinek et al., 2012; Zetsche et al., 2015; Zhang & You, 2020).

Para unir el DNA de manera afín y programada elegimos la versión catalíticamente inactiva de Cas12a (dCas12a, D908A) que se une al DNA, pero no lo corta. dCas12a une DNA con afinidad de 54 fM y es más selectiva que la más usada dCas9 (Strohkendl et al., 2018). El crRNA de dCas12a tiene 20 nt complementarios a la secuencia diana del DNA, la cual tiene que estar flanqueada por el motivo adyacente al espaciador (PAM) con secuencia TTTV donde V es A, G o C (Dong et al., 2016) (**Figura 6**).

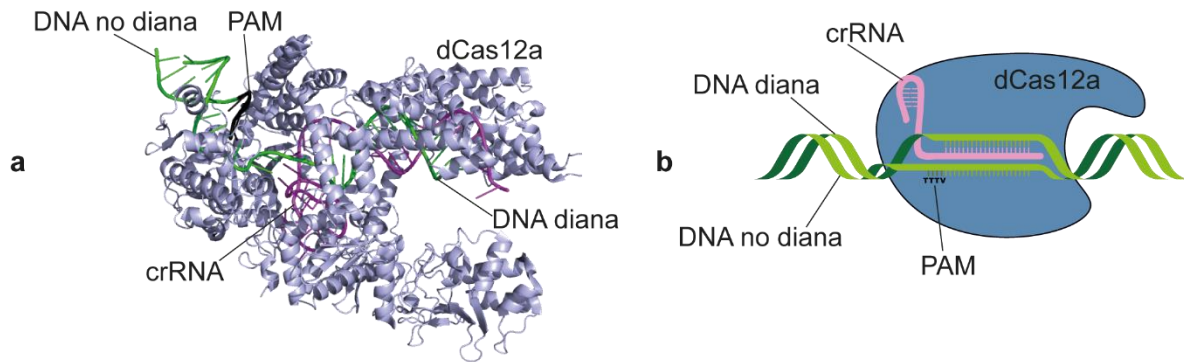


Figura 6. dCas12a se une al DNA y forma un lazo R de crRNA-DNA. (a) Estructura cristalográfica reportada para dCas12a de *Acidaminococcus* sp. (PDB: 5B43). (b) Modelo simplificado. La unión de dCas12a al DNA está mediada por un corto crRNA de 44 nt. Después de reconocer el motivo adyacente al espaciador (PAM, con secuencia TTTV) dCas12a separa la doble hélice y establece el lazo R entre el crRNA y la hebra diana del DNA.

Diseño de núcleo-cápsidas artificiales que se ensamblan sobre secuencias específicas de DNA

A partir de dCas12a y C-S₁₀-B creamos dos componentes para reclutar C-S₁₀-B a sitios diana sobre el DNA: dCas12a_(FKBP) y C-S₁₀_(FRB). Ambas proteínas tienen una fusión C-terminal con FKBP (dominio de unión a rapamicina) o FRB (dominio de unión a FKBP-rapamicina) y en presencia de rapamicina forman el heterodímero dCas12a-S₁₀ (Banaszynski et al., 2005; Inobe & Nukina, 2016) (**Figura 7**).

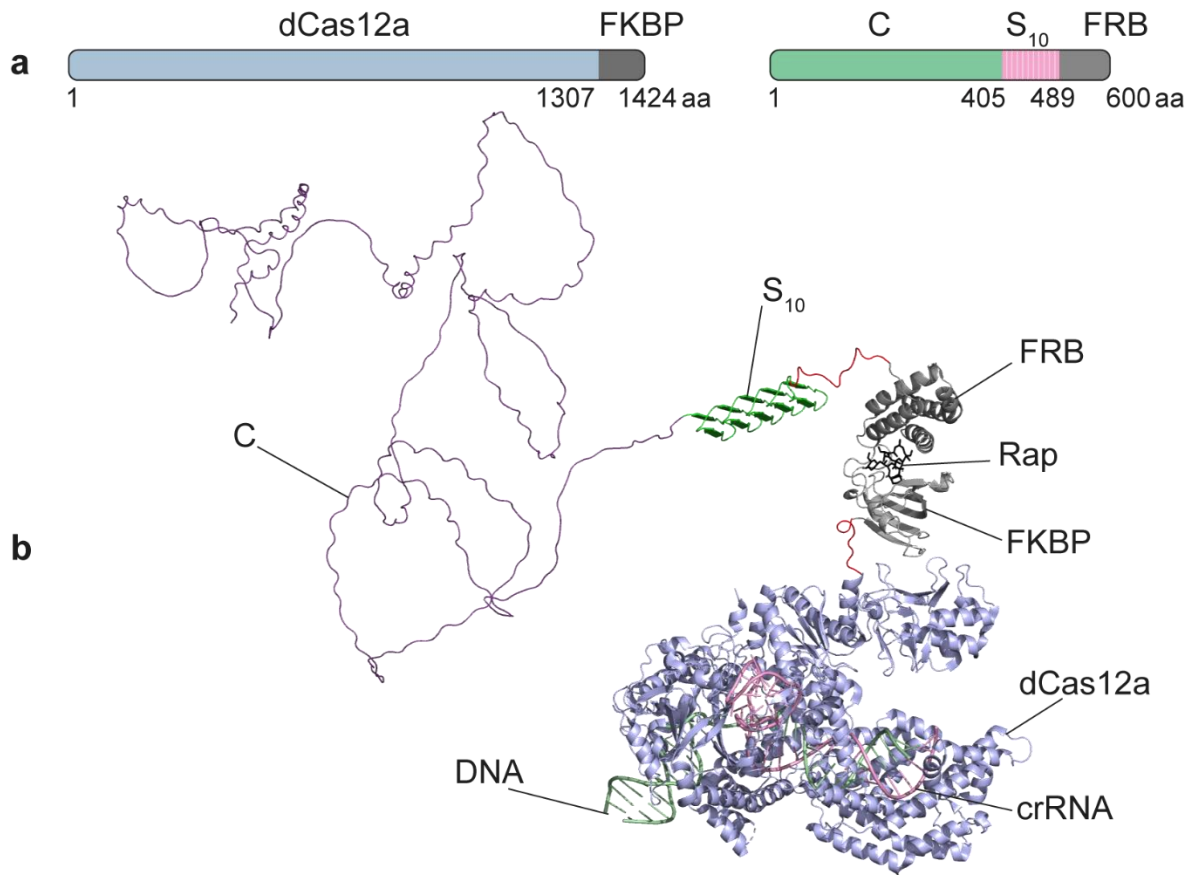


Figura 7. Proteínas de fusión que forman el complejo dCas12a-S₁₀. (a) Esquema con la estructura primaria de las proteínas de fusión dCas12a_(FKBP) y C-S₁₀_(FRB). (b) Estructura tridimensional esperada para el complejo dCas12a-S₁₀ formado por rapamicina (Rap) y unido al DNA diana.

Utilizamos la estrategia de heterodimerización con rapamicina por dos razones:

- 1) dCas12a se expresa en el citosol de *E. coli* y C-S₁₀ se expresa en *P. pastoris* por vía secretoria. La secreción de dCas12a en *P. pastoris* podría ocasionar modificaciones postraduccionales (como oxidación de puentes disulfuro y glicosilación) que afecten su plegamiento. Asimismo, la expresión de C-S₁₀ en el citosol de *E. coli* ocasiona su profunda hidrólisis por proteasas

citoplasmáticas (Willems et al., 2019). La conjugación irreversible de las proteínas resuelve el problema de la incompatibilidad de expresión.

2) La heterodimerización inducida de las proteínas permite observar directamente el efecto de acoplar las funciones de unión a DNA y reclutamiento de C-S₁₀-B. De esta forma se eliminan efectos secundarios debidos a la sola presencia de ambos bloques en el ensayo.

El heterodímero irreversible dCas12a-S₁₀ debe tener dos funciones: 1) unir el DNA con alta afinidad y programablemente por el diseño del crRNA y 2) reclutar C-S₁₀-B vía el dominio de polimerización S₁₀ (**Figura 8**).

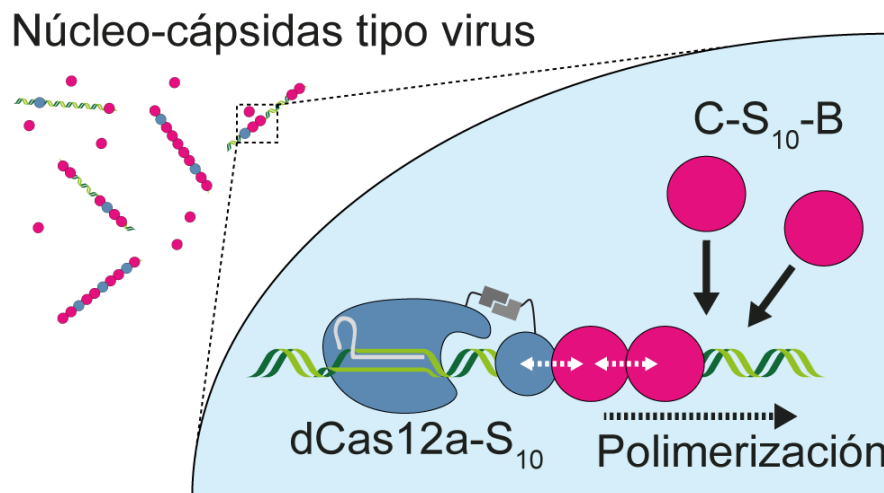


Figura 8. El complejo dCas12a-S₁₀ permitiría controlar y dirigir la nucleación de las núcleo-cápsidas artificiales sobre el DNA. dCas12a se une al DNA y el dominio de polimerización S₁₀ recluta C-S₁₀-B.

Justificación y planteamiento del problema

El desarrollo de partículas viromiméticas promete resolver problemas científicos actuales en los campos de nanomedicina, nanomateriales y vacunas, y complementar el conocimiento que tenemos de los virus. La protección y liberación responsiva de la carga, los mecanismos de ensamblaje y desensamblaje, y la adaptación son propiedades virales que se han recapitulado con partículas diseñadas *de novo*. Sin embargo, el empaquetamiento de moléculas específicas de DNA o RNA continúa siendo desafiante.

Para los virus, la encapsidación específica del genoma propio durante el ensamblaje se debe a regiones del genoma viral, denominadas señales de empaquetamiento, que interactúan fuertemente con las proteínas de la cápsida. Inspirados en las señales de empaquetamiento, diseñamos una estrategia para favorecer la interacción de la proteína autoensamblante C-S₁₀-B con secuencias específicas del DNA. Para ello, necesitamos una proteína que interactúe simultáneamente con determinadas secuencias de DNA y con C-S₁₀-B; de ahí el complejo dCas12a-S₁₀. Las partículas obtenidas de esta manera, con afinidad programable por el DNA, podrían ser el punto de partida para la generación de núcleo-cápsidas inteligentes.

Hipótesis

El complejo dCas12a-S₁₀—que acopla las funciones de unión específica al DNA y reclutamiento de C-S₁₀-B—puede programar el autoensamblaje de las núcleo-cápsidas artificiales formadas por C-S₁₀-B y DNA.

Objetivos

Objetivo general

Obtener núcleo-cápsidas tipo-virus artificiales que se ensamblan de manera programable sobre secuencias diana de DNA.

Objetivos específicos

1. Implementar una estrategia para acoplar dCas12a al dominio S₁₀.
2. Producir los componentes para el ensamblaje: dCas12a, complejo dCas12a-S₁₀, C-S₁₀-B y crRNAs.
3. Estandarizar un método de molécula única para monitorear el ensamblaje de las núcleo-cápsidas en tiempo real.
4. Caracterizar el ensamblaje de las núcleo-cápsidas a partir de DNA decorado con complejos proteicos que no interactúan con C-S₁₀-B (dCas12a y nucleosomas).
5. Caracterizar el ensamblaje de las núcleo-cápsidas a partir de DNA decorado con el complejo dCas12a-S₁₀.
6. Evaluar el potencial de dCas12a-S₁₀ para permitir la discriminación entre sustratos de DNA durante el ensamblaje de las núcleo-cápsidas.

Materiales y métodos

Enzimas, oligos y plásmidos

Las reacciones en cadena de la polimerasa (PCRs) para la clonación se hicieron con la DNA polimerasa de alta fidelidad Q5 (New England Biolabs, USA), mientras que las PCRs de colonia se hicieron con la DNA polimerasa Phire Green Hot Start (ThermoFisher, USA). Los oligonucleótidos fueron sintetizados por Integrated DNA Technologies (IDT, USA) y las enzimas de restricción fueron compradas de New England Biolabs (NEB, USA). Las reacciones de secuenciación se hicieron por el método de Sanger en Laragen (USA). Los plásmidos codificantes para los dominios FKBP y FRB fueron donados por Tom Wells (Universidad de Stanford) y el plásmido para dCas12a (D908A) fue donado por Ilya Finkelstein (Universidad de Texas en Austin).

Construcción del gen codificante para dCas12a_(FKBP) para expresión en *E. coli*

La construcción de la proteína de fusión dCas12a_(FKBP) se hizo por métodos convencionales de biología molecular: amplificación por PCR, restricción de los extremos de los amplicones y ligación isotérmica. Primeramente, se adicionaron por PCR 21 nucleótidos codificantes para el conector AAGGSGR en el extremo 3' del gen para dCas12a, seguido de los sitios de restricción para NotI y XhoI. El gen extendido de dCas12a se religó al plásmido de origen. La secuencia para el dominio FKBP, flanqueada por los sitios de restricción NotI y XhoI en ambos extremos y con 6 nucleótidos en el extremo 5' codificando para GG, se clonó en el extremo 3' del gen extendido para dCas12a. Las colonias de *E. coli* DH10B transformadas con las construcciones correctas fueron identificadas por PCR de colonia con cebadores complementarios a las regiones 5' y 3' del gen de fusión, respectivamente. Posteriormente se hizo secuenciación de plásmidos para descartar mutaciones. La amplificación y el mantenimiento de los plásmidos se hizo con *E.*

coli DH10B en medio Luria Bertani (LB, triptona 10 g/L, extracto de levadura 5 g/L y NaCl 10 g/L) suplementado con ampicilina 100 µg/mL (GoldBiotechnology, USA).

Construcción del gen codificante para C-S_{10(FRB)} para expresión en *P. pastoris*

El plásmido pPIC9 con la secuencia para la proteína tribloque (C-S₁₀-B) entre los sitios de restricción XhoI y EcoRI fue idéntico al usado por Hernandez-Garcia *et al.* (2014). Para reemplazar el dominio de unión B por FRB se amplificó el fragmento C-S₁₀ y se agregó el sitio de restricción NotI en su extremo 3'. La secuencia para FRB se amplificó para agregar los sitios de restricción NotI y EcoRI a los extremos 5' y 3' respectivamente. Los fragmentos de DNA para C-S₁₀, FRB, y el plásmido pPIC9 se ligaron simultáneamente y el producto de la ligación se usó para transformar células DH10B. La presencia del gen de fusión se verificó por PCR de colonia con los cebadores 5' AOX y 3' AOX y se hizo secuenciación para descartar mutaciones.

Los plásmidos para C-S₁₀-B y C-S_{10(FRB)} se linearizaron con Sall, se purificaron a partir de gel y se introdujeron por electroporación en *P. pastoris* GS115 electrocompetente. Las colonias recombinantes se seleccionaron por auxotrofia de histidina y se identificaron por PCR de colonia con los cebadores 5' AOX y 3' AOX. Una colonia recombinante con fenotipo Mut⁺ se creció en medio mínimo de glicerol (MGY: YNB 1.34%, glicerol 1%, biotina 4×10⁻⁵%) hasta DO_{600nm} ~4 UA en un matraz convencional de 250 mL. Se hicieron alícuotas del cultivo que después de sumergirse en nitrógeno líquido se almacenaron a -70°C hasta la expresión.

Expresión y purificación de dCas12a y dCas12a_(FKBP)

Para la expresión de dCas12a y dCas12a_(FKBP) se usó *E. coli* BL21 (DE3) y el medio Terrific Broth (TB, triptona 12 g/L, extracto de levadura 24 g/L, glicerol 0.4%, KH₂PO₄ 17 mM, K₂HPO₄ 72 mM). El crecimiento inicial se hizo en matraces Fernbach de 1L hasta

DO_{600nm}=0.8 UA a 37°C y posteriormente los matraces se pasaron a 12°C. La expresión de las proteínas se indujo con 1mM IPTG (GoldBiotechnology, USA) 1 h después del cambio de temperatura y durante 24 h. La selección se hizo con 100 µg/mL ampicilina.

El pellet de células recuperado por centrifugación se resuspendió en buffer de lisis (Tris-HCl 20 mM pH 8.0, NaCl 1 M, EDTA 1 mM, 10 mM imidazol y PMSF 1 mM) y se sonicó en hielo con ciclos ON/OFF de 5s/25s por 6 min totales de sonicación. El sobrenadante obtenido por centrifugación (30 min, 30 000 g) y filtración (0.22 µm) se inyectó a una columna de Ni-NTA (Sigma, USA). La columna se lavó con buffer de lavado (Tris-HCl 20 mM pH 8.0, NaCl 1 M, EDTA 1 mM y 25 mM imidazol) y se eluyó la proteína con buffer de elución (Tris-HCl 50 mM pH 8.0, NaCl 1 M, EDTA 1 mM, 250 mM imidazol). Se añadió la proteasa SUMO para remover la etiqueta 6xHis y la etiqueta de solubilidad SUMO. La mezcla se dializó toda la noche a 4°C contra buffer de diálisis (dCas12a: buffer de Fosfatos 50 mM pH 6, 150 mM KCl, 5 mM MgCl₂, 10% glicerol, DTT 2 mM; dCas12a_(FKBP): 20mM HEPES-KOH pH 7.2, 150 mM KCl, 5mM MgCl₂, 10% glicerol, 2 mM DTT) en una membrana con tamaño de corte de 50 kDa. La muestra filtrada se inyectó a una columna HiTrap (Sigma, USA) de intercambio catiónico y se eluyó con un gradiente de 150-2000 mM KCl. La proteína se concentró en unidades de filtración por centrifuga (Merck, USA) hasta ~200 µL y se inyectó en una columna de exclusión molecular HiLoad (Sigma, USA) para la separación por tamaño y la eliminación de agregados. Las fracciones con dCas12a o dCas12a_(FKBP) puras (>90%) se mezclaron y concentraron en unidades de filtración hasta 12 µM. Las proteínas se congelaron en nitrógeno líquido y se guardaron a -70°C. La pureza e identidad de las proteínas se confirmó por SDS-PAGE y Western-blot con anticuerpo anti-6xHis.

Expresión y purificación de C-S₁₀-B y C-S₁₀(FRB)

Previamente reportamos que el bloque C sufre proteólisis cuando se expresa en el citosol de *E. coli* (Willems et al., 2019). Por lo tanto, expresamos C-S₁₀-B y C-S₁₀(FRB) en *P. pastoris* con un péptido señal N-terminal para su secreción al medio. Se plaquearon 25 μ L del banco en una placa con medio mínimo selectivo para auxotrofia de histidina (MGY) y se dejó crecer durante 72 h a 30°C. Una colonia se usó para inocular un matraz Erlenmeyer bafleado de 1 L con 300 mL de MGY y se dejó crecer toda la noche a 30°C. Las células se recuperaron mediante centrifugación a 20°C con 5 000 g y se usaron para inocular matraces Fernbach bafleados de 2.8 L con 1 L de medio mínimo de metanol (MMY: YNB 1.34%, biotina $4 \times 10^{-5}\%$). El cultivo se dejó crecer por 72 h a 30°C y se añadió 1% de metanol cada 24 h. El sobrenadante de cultivo se obtuvo por centrifugación (4°C, 10 000 g) y filtración por 0.22 μ m.

El medio libre de células se alcalinizó hasta pH 8.0 con 1 M NaOH para precipitar el exceso de sales y se suplementó con 1 mM EDTA y 1 mM PMSF para inhibir la proteólisis. Para precipitar proteínas por sales se añadió (NH₄)₂SO₄ hasta 50% de saturación y se incubó toda la noche a 4°C. El precipitado se obtuvo por centrifugación a 12 000 g por 30 min y se resuspendió en agua milli-Q a 65°C. Se repitió la precipitación con sales y el nuevo precipitado se resuspendió en agua milli-Q a 65°C y se añadió NaCl 50 mM y acetona hasta 40%. La solución se centrifugó a 12 000 g por 30 min y se descartó el precipitado. La concentración de acetona en el sobrenadante se elevó hasta 80% para precipitar selectivamente a C-S₁₀-B. El precipitado se recuperó por centrifugación, se resuspendió en agua milli-Q a 65°C y se dializó toda la noche contra agua milli-Q antes de ser liofilizado.

La purificación de C-S₁₀(FRB) se hizo de manera similar a C-S₁₀-B con algunas variaciones. El medio de cultivo libre de células se ultrafiltró con un dispositivo Amicon (Merck, USA) con membrana de 30 kDa. La solución retenida se precipitó dos veces con

(NH₄)₂SO₄ hasta 80% de saturación durante toda la noche. La precipitación con acetona a 40% y 80% se hizo con PBS al igual que la diálisis. Las alícuotas con C-S_{10(FRB)} se sumergieron en nitrógeno líquido y se guardaron a -70°C. La pureza de las proteínas se confirmó por SDS-PAGE.

C-S₁₀-B (45 kDa) y C-S_{10(FRB)} (58 kDa) se identificaron por espectrometría de masas (MALDI-TOF) con un espectrómetro Microflex (Bruker). Se usó una matriz de ácido sinapínico saturante y ácido trifluoroacético 1% (v/v). Para la calibración se usaron las proteínas citocromo c (12 kDa) y albúmina de suero bovino (66.5 kDa) en la misma matriz.

Expresión y purificación de la proteasa SUMO

El plásmido para la expresión de la proteasa SUMO fue proporcionado por el Dr. Ilya Finkelstein (Universidad de Texas en Austin). El plásmido se transformó en células de *E. coli* BL21 (DE3) y una colonia se usó para inocular un cultivo de 1 L de LB y se creció a 37°C hasta DO_{600nm} = 0.6 UA. Seguidamente se indujo la expresión con IPTG 0.5 mM y se dejó crecer por 4 h a 37°C.

Las células se colectaron por centrifugación y se sonicaron sobre hielo en buffer de lisis (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 20 mM imidazol y 10% glicerol). El lisado se clarificó por centrifugación a 30 000 g por 45 min y se pasó por una columna de Ni-NTA (Sigma, USA). La proteína se eluyó con buffer de elución (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 300 mM imidazol y 25% glicerol) y se dializó toda la noche en PBS, pH 7.5 con 10% glicerol. Las alícuotas se congelaron en nitrógeno líquido y se almacenaron a -70°C.

Diseño y producción de crRNAs

Para los ensayos de fluorescencia y AFM se usaron crRNAs transcritos a partir de templados de DNA obtenidos por PCR. Dos oligonucleótidos (44 nt) complementarios entre

sí en una región de 20 nt se usaron en una PCR sin templado para generar fragmentos de doble hebra de 44 bp. El amplicón se purificó por columna (Zymo) y se usó como templado para transcripción *in vitro* a 37°C durante la noche (HiScribe™ T7, NEB). Los crRNAs se purificaron por kit (Monarch® RNA Cleanup Kit, NEB).

Los crRNAs para los ensayos de cortinas de DNA se transcribieron *in vitro* a partir de templados de DNA parcialmente de doble hebra (obtenidos por hibridación de dos oligos de diferente tamaño). El primer oligo (*forward*) es de 24 nt y contiene el promotor de la polimerasa T7. El segundo oligo (*reverse*) de 68 nt es complementario al oligo *forward* y además contiene las regiones constante y variable del crRNA. Los dos oligos se mezclaron en razón 1.5:1 en buffer de hibridación (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA) a 75°C por 5 min y luego se enfrió hasta 25°C gradualmente durante 25 min. La reacción de transcripción se hizo con el kit HiScribe™ T7 (NEB, USA) y los crRNA se purificaron con TRizol (Ambion, USA). De cinco a veinticinco oligos *reverse* se usaron simultáneamente en la reacción de hibridación para crear *pools* de crRNAs. El tamaño y la pureza de los crRNAs se evaluó por electroforesis en gel de poliacrilamida con urea 8 M (Urea-PAGE). Para la tinción se usó SYBR Gold (ThermoFisher Scientific, USA)

Caracterización estructural de dCas12a y dCas12a_(FKBP) por dicroísmo circular

El contenido de estructura secundaria en las proteínas dCas12a y dCas12a_(FKBP) se analizó por dicroísmo circular en UV-lejano (190-260 nm). Primeramente, los iones cloruro—que absorben fuertemente en longitudes de onda debajo de 200 nm, donde está el pico de absorbancia de hélices alfa—se remplazaron por iones fluoruro. El intercambio de buffer se hizo con unidades de filtración por centrifuga de 50 kDa MWCO (Merck, USA) contra buffer de fluoruro (50 mM buffer de fosfatos, pH 7.4, 100 mM KF y 1 mM DTT). Las proteínas se usaron en concentración final 0.1 mg/mL.

Dimerización de dCas12a_(FKBP) y C-S_{10(FRB)} inducida por rapamicina

La dimerización de dCas12a_(FKBP) y C-S_{10(FRB)} en presencia de rapamicina se evaluó por dispersión dinámica de la luz con un Zetasizer μ V a 173° (Malvern Instruments). El experimento se hizo a temperatura ambiente con una celda de vidrio de 20 μ L. Ambas proteínas se mezclaron en razón equimolar (2 μ M) con 2 μ M rapamicina en buffer de ensayo (20 mM HEPES-KOH, pH 7.5, 100 mM KCl, 1 mM DTT). Cada muestra se midió en cinco ciclos de diez repeticiones cada uno. Los resultados se reportaron en número de partículas.

Interacción de dCas12a con el crRNA

La unión a crRNA se monitoreó por el apagamiento de la fluorescencia intrínseca de los triptófanos en dCas12a. La fluorescencia a diferentes concentraciones de crRNA se caracterizó por fluorimetría en un lector para placas (Cytation, Biotek) con $\lambda_{ex} = 290$ nm y $\lambda_{em} = 350$ nm. dCas12a se usó a concentración 2 μ M. Los valores de fluorescencia se graficaron según la ecuación Stern-Volmer que relaciona el apagamiento de la fluorescencia con la concentración del *quencher* (crRNA).

Unión de la ribonucleoproteína (dCas12a-crRNA) a DNA y visualización de las núcleo-cápsidas

Para evaluar la unión sitio-específica de dCas12a al DNA, la ribonucleoproteína (o complejo ribonucleoproteico, RNP) se obtuvo al incubar dCas12a con el crRNA en razón 1:10 en buffer de AFM (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 15 mM MgCl₂, 1 mM DTT) a 37°C por 30 min. El fragmento lineal de DNA (1.5 kbp) con cuatro sitios de unión idénticos se mezcló con la RNP en razón 1:10 (sitio de unión-RNP) y se incubó toda la noche a temperatura ambiente antes de la visualización.

Para confirmar la morfología de las núcleo-cápsidas, las RNPs (dCas12a o dCas12a_(FKBP) con crRNA) se mezclaron con un fragmento de DNA lineal de 2.5 kbp (con diez sitios de unión) a razón 1:10 (sitio de unión-RNP) por 1 h a temperatura ambiente. C-S₁₀-B se añadió a N/P = 3 (razón entre [C-S₁₀-B] y [sitios de unión de 6 bp]) y se incubó toda la noche.

Las muestras se diluyeron a una concentración final de 5 nM DNA y se depositaron en una mica fresca durante 10 min. La mica se lavó con 500 µL de agua milli-Q y se secó bajo una corriente de aire filtrado. Las imágenes se adquirieron en un microscopio de fuerza atómica (NanoScope V, Bruker) en modo *ScanAsyst* en aire con resolución 1024 muestras/línea. Las imágenes se procesaron con el software NanoScope Analysis v1.89.

Encapsidación del DNA por C-S₁₀-B

La encapsidación del DNA por C-S₁₀-B se estudió mediante un ensayo de retardo de la movilidad electroforética (EMSA, del inglés *electrophoretic mobility shift assay*). dCas12a cargada con el crRNA (1:10) se incubó con 100 ng de un DNA lineal (2.5 kbp, con diez sitios diana) o un plásmido circular (9.5 kbp, con cuatro sitios diana idénticos) en estequiometría 1:10 (sitio de unión-RNP) por 1 h a temperatura ambiente. C-S₁₀-B se añadió a N/P de 0 a 3 (razón entre [C-S₁₀-B] y [sitios de unión de 6 bp]) y se incubó por 3 h (DNA lineal) o toda la noche (plásmido). Las muestras se cargaron en geles de agarosa y se corrieron a 70 V a 4°C por 2 h. La tinción se hizo con SYBR Green. Las imágenes se procesaron en FIJI. Para cada carril se cuantificó el porcentaje del DNA libre y el retenido en el gel; el último corresponde al DNA completamente encapsidado.

Preparación de sustratos de DNA para los ensayos de cortinas

El DNA del bacteriófago λ (λDNA, NEB) se hibridó con un oligo biotilado complementario a uno de sus extremos cohesivos a 70°C por 15 min y enfriando

gradualmente hasta 15°C durante 2 h. Se añadió la ligasa T4 (NEB) y se incubó toda la noche a temperatura ambiente seguido por inactivación de la enzima con 2 M NaCl. El λ DNA biotinilado se purificó por exclusión molecular en una columna Sephacryl S-1000 (GE Healthcare, USA).

Para los experimentos con DNA nucleosomal se usaron histonas humanas (3xHA H2A and H2B, H3, H4; Histone Source). El λ DNA biotinilado y purificado se precipitó con acetato de sodio (pH 5.5) y 0.3 M isopropanol a razón 1:1 (v/v) en hielo y centrifugación a 15 000 g por 30 min. El DNA precipitado se resuspendió a concentración 0.8 nM en buffer 2 M TE (10 mM Tris-HCl pH 8, 1 mM EDTA, 2 M NaCl) y se añadieron los octámeros de histona a razón molar 50:1 y 100:1 (octámero-DNA). Las mezclas de 100 μ L se depositaron en botones de diálisis en buffer de diálisis (10 mM Tris-HCl pH 7.6, 1 mM EDTA, 1 mM DTT) con concentraciones decrecientes de NaCl a 4°C. El primer paso fue contra 1.5 M NaCl por 1 h, seguido de cuatro pasos de 2 h cada uno con 1, 0.8, 0.6 y 0.4 M NaCl. Finalmente, la diálisis se dejó toda la noche contra 0.2 M NaCl.

Para los ensayos de cortinas con dos hebras de DNA, el λ DNA se ligó con oligos biotinilados en ambos extremos cohesivos. La hibridación y la ligación se hicieron como se describió anteriormente, salvo que se usó el buffer CutSmart (NEB) suplementado con 1 mM rATP en lugar del buffer para la T4 ligasa. La mezcla se calentó a 65°C por 10 min para inactivar la ligasa y se añadió NgoMIV para la restricción a 37°C por 3 h. Finalmente, se añadió 2 M NaCl y ambos fragmentos de DNA se purificaron en conjunto por exclusión molecular. La completa restricción por NgoMIV se confirmó por electroforesis en 0.3% agarosa. Todos los sustratos de DNA (λ DNA, DNA nucleosomal y las dos hebras biotiniladas de DNA) se almacenaron a 4°C, por no más de cuatro semanas, hasta su uso.

Ensamblaje de las núcleo-cápsidas en tiempo real

dCas12a y dCas12a_(FKBP) se mezclaron con los *pools* de crRNAs en razón 1:10 en buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM MgCl₂, 2% glicerol, 2 mM DTT) a 37°C por 30 min. La RNP a 10 nM se inyectó en la celda durante 30 min a temperatura ambiente para la unión al DNA. El marcaje de dCas12a se hizo vía anticuerpos anti-FLAG biotinilados contra la etiqueta FLAG en el extremo C-terminal de dCas12a. A su vez, los anticuerpos se conjugaron con puntos cuánticos decorados con estreptavidina (QD₇₀₅, Sigma). C-S₁₀-B y C-S₁₀(FRB) se marcaron vía maleimido con Alexa488 y Atto647N, respectivamente.

El ensamblaje de las núcleo-cápsidas individuales se monitoreó con un ensayo de cortinas de DNA. Las celdas de microfluídica se prepararon a partir de una placa con nanoimpresión de barreras de cromo de grosor 1-2 μm y un portaobjeto de sílica; el canal entre ambas superficies se construyó con cinta doble cara. El interior de la celda se recubrió con una bicapa lipídica al incubar con una solución de lípidos (1.954% DOPC, 0.04% DOPE-mPEG2k y 0.006% DOPE-biotina, de Avanti Polar Lipids) en buffer (10 mM Tris-HCL pH 8, 100 mM NaCl) por 30 min a temperatura ambiente. La celda se lavó con buffer de BSA (40 mM Tris-HCl pH 8, 2 mM MgCl₂, 0.2 mg/mL BSA) durante 10 min y luego con 0.1 mg/mL estreptavidina en buffer de BSA por 10 min. El DNA biotinilado diluido en buffer BSA se inyectó en la celda y se incubó por 30 min para permitir la fijación; el DNA no pegado a la celda se lavó con buffer de BSA. El buffer de ensayo fue buffer de BSA suplementado con 100 mM NaCl, 5 mM MgCl₂, 2 mM DTT.

En los experimentos en que el DNA se tiñó con el reactivo fluorescente YOYO-1 se añadió solución gloxy (500 U catalasa, 70 U glucosa oxidasa y 1% glucosa w/v) al buffer de ensayo. Durante la iluminación, el YOYO-1 en singlete activado puede cruzar a triplete y ceder electrones al oxígeno para formar radicales libres que atacan el enlace fosfodiéster

del DNA. La glucosa oxidasa y la catalasa reducen el oxígeno en peróxido y luego en agua, desoxigenando el buffer y protegiendo al DNA del fotodaño.

Las imágenes se adquirieron con aumento 60X en un microscopio invertido Nikon Ti-E para reflexión total interna de la fluorescencia. La excitación de los fluoróforos se hizo con un láser de 488 nm y la emisión se separó en canales rojo y verde con un espejo dicróico a 638 nm (Chroma). La captura de las imágenes se hizo con dos cámaras EM-CCD (Andor iXon DU897) y tamaño de píxel de 266 nm. Las imágenes se procesaron en FIJI.

Obtención de la estructura 3D esperada para el complejo dCas12a-S₁₀

Los modelos tridimensionales para las proteínas de fusión dCas12a_(FKBP) y C-S_{10(FRB)} se obtuvieron en PyMOL 2.4.1. dCas12a_(FKBP) se construyó a partir de las estructuras reportadas para dCas12a de *Acidaminococcus* sp. (PDB: 5B43) y FKBP (PDB: 1FAP). Para representar el dominio tipo colágeno (C) de C-S_{10(FRB)} se usó la estructura predicha para el colágeno α -1 tipo II de ratón (AlphaFold, P28481). Similar al colágeno, el dominio C consiste en repeticiones del triplete Gly-Xaa-Yaa, donde Xaa y Yaa son frecuentemente prolina (Werten et al., 2001). El dominio S₁₀ se obtuvo a partir de la estructura de la proteína de la seda (PDB: 2SLK) con secuencia GAGAGA. Se reorganizó la estructura para obtener un rollo- β con dos hojas- β paralelas conectadas entre sí por giros de Gly-Gln-Gly. La estructura de FRB se obtuvo del PDB (PDB: 1FAP). Los péptidos conectores entre los bloques de cada proteína de fusión se construyeron en PyMOL.

Análisis estadístico

Todos los resultados se reportan como media \pm desviación estándar o media \pm intervalos de confianza de la media al 95% según se indica en el texto. Para detectar diferencias entre medias se usaron las pruebas no paramétricas U de Mann-Whitney (para dos medias) y el ANOVA Kruskal-Wallis (para múltiples medias) con comparación de

medias de Dunn, todas disponibles en GraphPad Prism 8. El ajuste de histogramas a la suma de cinco o diez Gaussianas se hizo en OriginPro 8.5. Las barras de error para los histogramas se generaron por Bootstrap.

La curva de velocidad de condensación del DNA contra la concentración de C-S₁₀-B se ajustó a la ecuación de Hill, comúnmente usada para ensamblajes cooperativos (**Ecuación 1**). Aquí, *velocidad₀* y *velocidad_f* son las velocidades de compactación obtenidas para 10 y 300 nM, respectivamente. $[C-S_{10}-B]_{1/2}$ es la concentración de C-S₁₀-B a la que se alcanza el 50% de la velocidad máxima y *Hill* es la pendiente de la sigmoide.

$$\text{velocidad} = \text{velocidad}_0 + \left(\frac{\text{velocidad}_f - \text{velocidad}_0}{1 + ([C-S_{10}-B]_{1/2}/[C-S_{10}-B])^{\text{Hill}}} \right) \quad (1)$$

Las curvas de condensación del DNA en el tiempo se ajustaron a la ecuación de Hill (**Ecuación 2**). Aquí, *longitud del DNA₀* y *longitud del DNA_f* son la longitud del DNA completamente extendido por el flujo del buffer (100%) y del DNA completamente compactado (33%), respectivamente. El parámetro *t_{half}* es el tiempo en el que se alcanza la mitad de la compactación máxima (66.5%) y *Hill* es la pendiente de la sigmoide.

$$\text{longitud del DNA} = \text{longitud del DNA}_f + \left(\frac{\text{longitud del DNA}_0 - \text{longitud del DNA}_f}{1 + (t_{\text{half}}/t)^{\text{Hill}}} \right) \quad (2)$$

Resultados y discusión

Obtención de los bloques de construcción para las núcleo-cápsidas

Los genes codificantes para las nuevas proteínas quiméricas (dCas12a_(FKBP) y C-S_{10(FRB)}) se obtuvieron mediante pasos de PCR, restricción y ligación (**Figura 9a-b**). Ambas proteínas en cada fusión se conectaron por un espaciador flexible y polar (AAGGSGRGG y SGRGGR para dCas12a_(FKBP) y C-S_{10(FRB)}, respectivamente), previamente usado para una proteína de fusión con dCas9 (Guilinger et al., 2014). La correcta fusión de los genes se confirmó por PCR de colonia en *E. coli* o *P. pastoris* (**Figura 9c, d**).

Las proteínas dCas12a y dCas12a_(FKBP) se aislaron de la fracción soluble del lisado de *E. coli* BL21 (DE3) y sus pesos moleculares se confirmaron por SDS-PAGE y por Western-blot (**Figura 10**). Las dos bandas que se observan en el Western-blot no aparecen en los geles de SDS-PAGE por lo que hipotetizamos que éstas son un artefacto de la transferencia. dCas12a y dCas12a_(FKBP) se transfieren lentamente a la membrana (debido a su tamaño) y el sándwich se desensambló repetidamente para evitar la desecación; esto posiblemente causó el deslizamiento del gel sobre la membrana.

La purificación consistió en tres pasos cromatográficos: afinidad por níquel, intercambio catiónico y exclusión molecular; las proteínas se obtuvieron con pureza >90% (**Figura 11**).

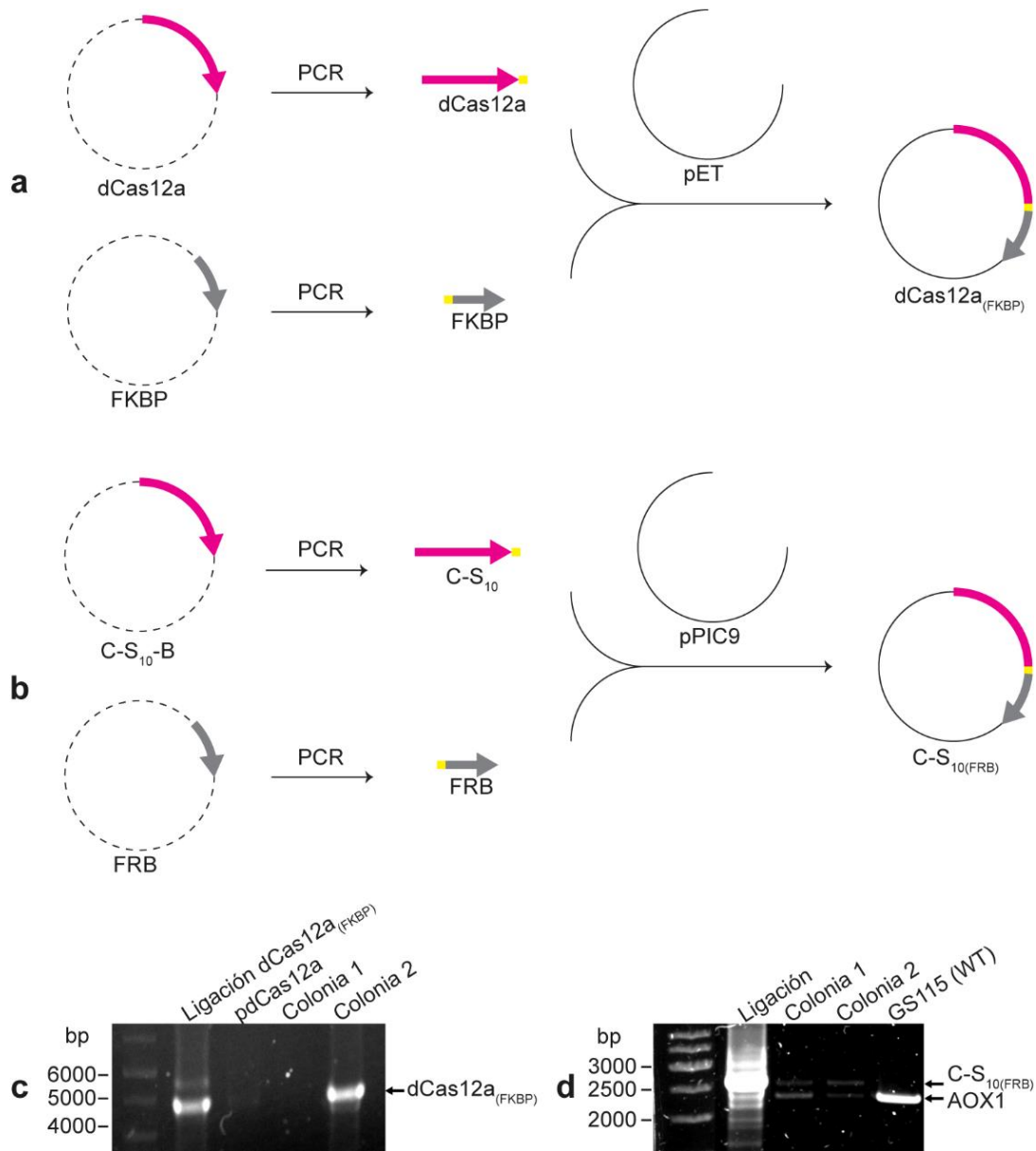


Figura 9. Estrategias de clonación para crear los genes codificantes de las proteínas quiméricas dCas12a_(FKBP) y C-S₁₀(FRB). (a) Las secuencias codificadoras para dCas12a y FKBP se amplificaron por PCR y se añadieron extremos complementarios entre el 3' y 5' de las secuencias para dCas12a y FKBP, respectivamente. Ambas secuencias se insertaron secuencialmente en el plásmido pET19 (para inducción por IPTG en *E. coli*). Un espaciador de nueve residuos (AAGGSGRGG, en amarillo) se incluyó entre ambas secuencias codificantes. (b) Las secuencias codificantes para el dibloque C-S₁₀ y para FRB se amplificaron por PCR y se añadieron extremos complementarios entre ellas. La ligación de ambas secuencias y el plásmido pPIC9 (para inducción por metanol y secreción de la proteína en *P. pastoris*) se hizo en una sola reacción de ligación. Un espaciador de ocho

residuos (SGRGGR, en amarillo) se incluyó entre ambas secuencias. **(c)** PCR para identificar las colonias de *E. coli* DH10B con el plásmido para dCas12a_(FKBP). El producto de la reacción de ligación y el plásmido para dCas12a se usaron como controles positivo y negativo, respectivamente. Solo la colonia 2 contiene el plásmido con dCas12a_(FKBP). **(d)** PCR para identificar la colonia de *P. pastoris* GS115 con el DNA integrado para C-S₁₀(FRB). El producto de la reacción de ligación y la cepa *wild-type* se usaron como controles positivo y negativo, respectivamente. Ambas colonias (1 y 2) contienen el gen de interés, y son Mut⁺ porque se observa la banda correspondiente al gen de la enzima alcohol oxidasa 1 (AOX1).

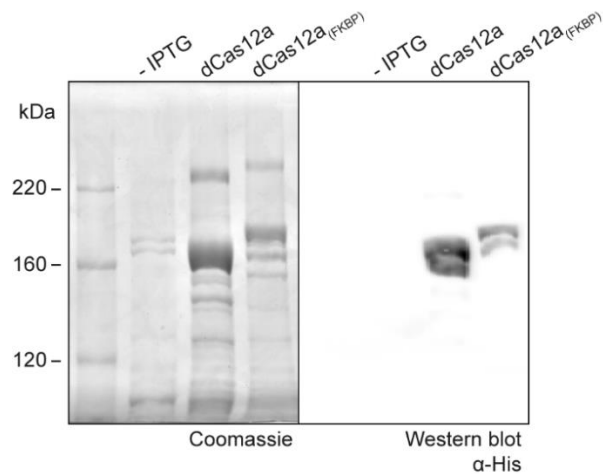


Figura 10. Expresión de dCas12a y dCas12a_(FKBP) en *E. coli*. SDS-PAGE que muestra la inducción con IPTG de la expresión de dCas12a (171 kDa) y dCas12a_(FKBP) (180 kDa). Ambas proteínas se identificaron vía Western-blot contra la etiqueta de 6xHis en el extremo N-terminal. En el gel se muestra la fracción de proteína soluble al final del cultivo.

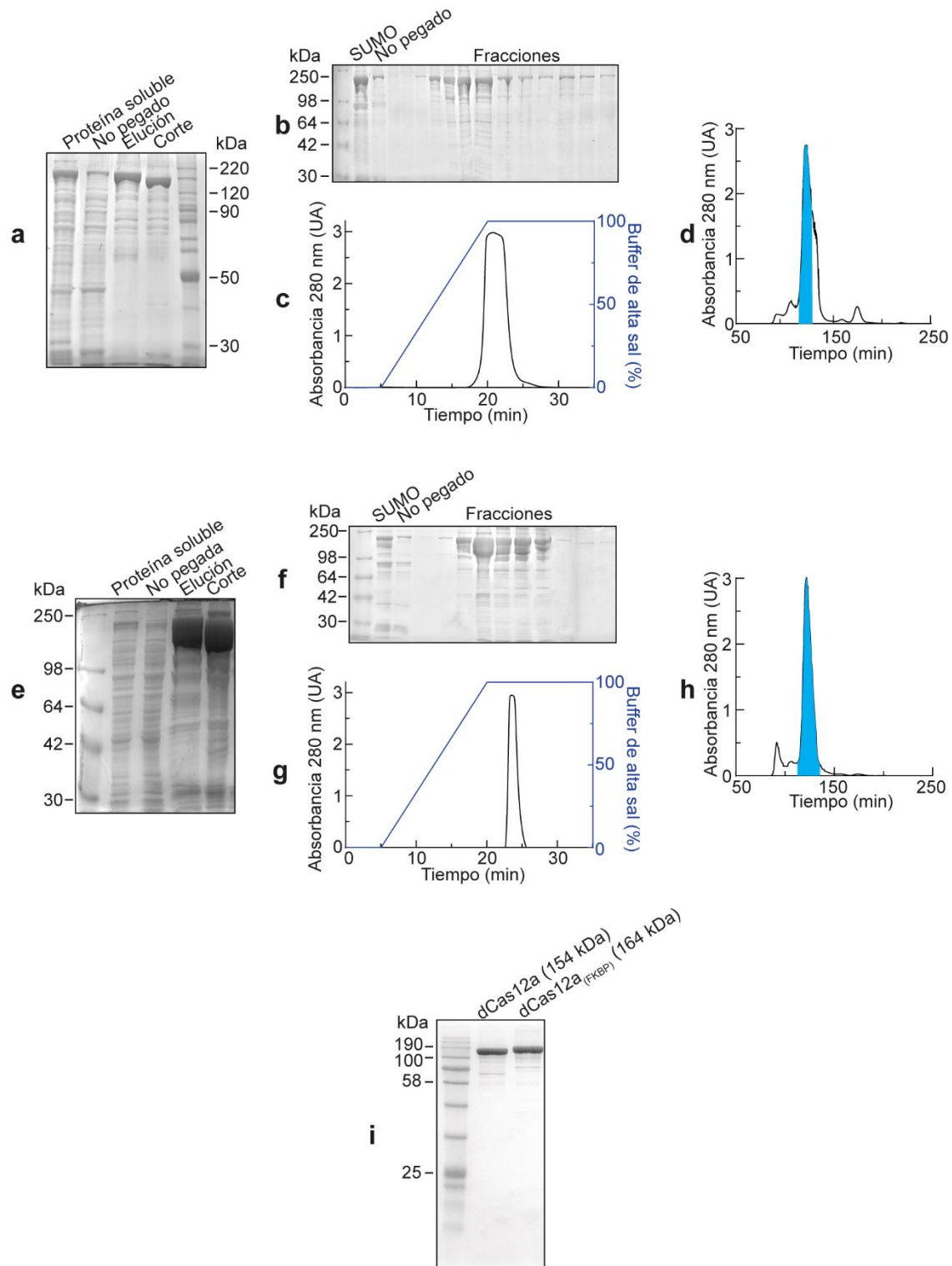


Figura 11. Purificación de dCas12a y dCas12a_(FKBP). Ambas proteínas se purificaron con tres pasos de cromatografía: afinidad por níquel, intercambio catiónico y exclusión molecular. **(a-d):** dCas12a y **(e-h):** dCas12a_(FKBP). La etiqueta de solubilidad SUMO, junto con la etiqueta 6xHis, se eliminó durante la diálisis entre la columna de afinidad y el

intercambio iónico. **(a,e)** SDS-PAGE para la cromatografía de afinidad con níquel. **(b,f)** SDS-PAGE para el intercambio iónico. **(c,g)** Cromatograma del intercambio iónico. **(d,h)** Cromatograma de la exclusión molecular. Aquí se aisló el pico de mayor absorbancia (azul). **(i)** SDS-PAGE con las proteínas dCas12a y Cas12a_(FKBP) purificadas.

En cuanto a estructura secundaria, dCas12a y dCas12a_(FKBP) mostraron un contenido predominante de hélices alfa y, en menor parte, de hojas beta (**Figura 12a**). Lo anterior coincide con lo esperado según las estructuras cristalográficas de dCas12a y FKBP (Wilson et al., 1995; Yamano et al., 2016) (**Figura 12b**).

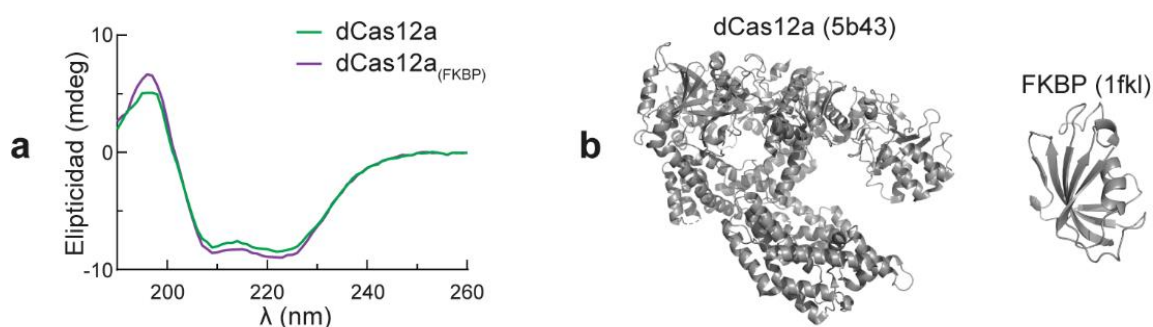


Figura 12. Caracterización estructural de dCas12a y dCas12a_(FKBP) por dicroísmo circular. **(a)** dCas12a y dCas12a_(FKBP) mostraron el espectro distintivo de hélices alfa y hojas beta esperado para ambas proteínas. **(b)** Se muestran las estructuras cristalográficas para dCas12a y FKBP como referencia.

C-S₁₀-B y C-S_{10(FRB)} son altamente hidrofílicas debido al bloque C tipo colágeno. Por lo tanto, la purificación de ambas proteínas se basó en pasos de precipitación con sales y acetona (**Figura 13a, d**). Las prolinas en el bloque C causan que C-S₁₀-B (45 kDa) y C-S_{10(FRB)} (58 kDa) migren aberrantemente en SDS-PAGE con tamaño aparente de ~100 kDa, por lo que la identificación de las proteínas se hizo por MALDI-TOF. El peso observado para C-S₁₀-B y C-S_{10(FRB)} fue 44 666.78 Da y 56 781.09 Da, respectivamente, lo que implica una

desviación del valor teórico de 0.2% y 2.3%. (**Figura 13b, e**). El marcaje de ambas proteínas con fluoróforos orgánicos (Alexa488 y Atto647N) no afectó la migración en el gel (**Figura 13c, f**) ni su actividad, como se demostró posteriormente.

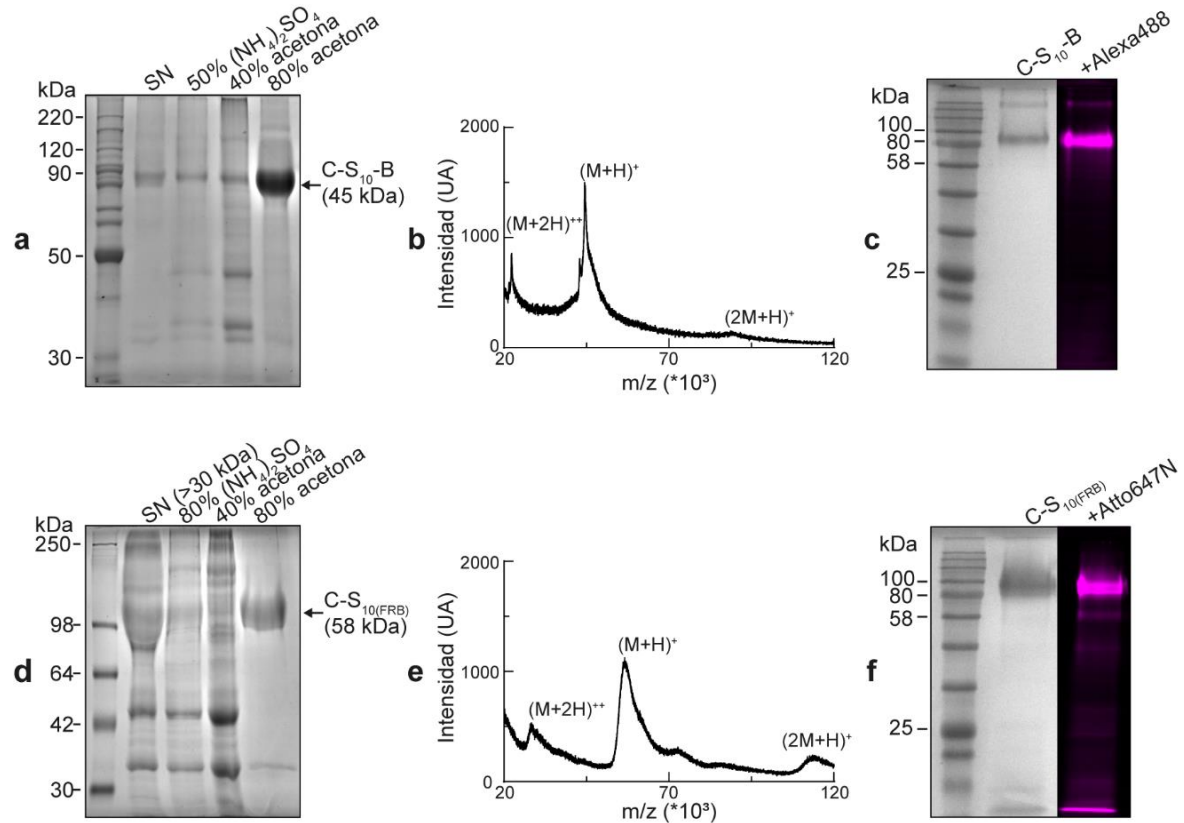


Figura 13. Expresión en *P. pastoris* y purificación de las proteínas C-S₁₀-B y C-S₁₀(FRB). **(a)** SDS-PAGE con la purificación de C-S₁₀-B a partir del sobrenadante de cultivo (SN) con pasos de precipitación con sulfato de amonio y acetona. C-S₁₀-B migra de manera aberrante en el gel debido al alto contenido de prolina (22%) en el dominio C. **(b)** Identificación de C-S₁₀-B por MALDI-TOF donde se observan las variantes monomérica (M+H)⁺, monomérica doblemente cargada (M+2H)⁺⁺ y dimérica (2M+H)⁺. **(c)** SDS-PAGE con C-S₁₀-B antes y después del marcaje con Alexa488-maleimido en el único residuo de cisteína N-terminal. **(d)** SDS-PAGE con la purificación de C-S₁₀(FRB) a partir del sobrenadante de cultivo (SN) filtrado por 30 kDa y con pasos de precipitación con sulfato de amonio y acetona. C-S₁₀(FRB) migra de manera aberrante en el gel debido al alto contenido de prolina (22%) en el dominio C. **(e)** Identificación de C-S₁₀(FRB) por MALDI-TOF donde se observan las variantes monomérica (M+H)⁺, monomérica doblemente cargada (M+2H)⁺⁺ y dimérica (2M+H)⁺. **(f)**

SDS-PAGE con C-S_{10(FRB)} antes y después del marcaje con Atto647N-maleimido en dos posibles residuos de cisteína: en el extremo N-terminal o C65 en FRB.

La heterodimerización de dCas12a_(FKBP) y C-S_{10(FRB)} (**Figura 14a**) ocurrió inmediatamente después de la adición de rapamicina. La formación del dímero se observó por dispersión dinámica de la luz como un cambio del tamaño de las partículas de 10 nm a 18 nm (**Figura 14b**).

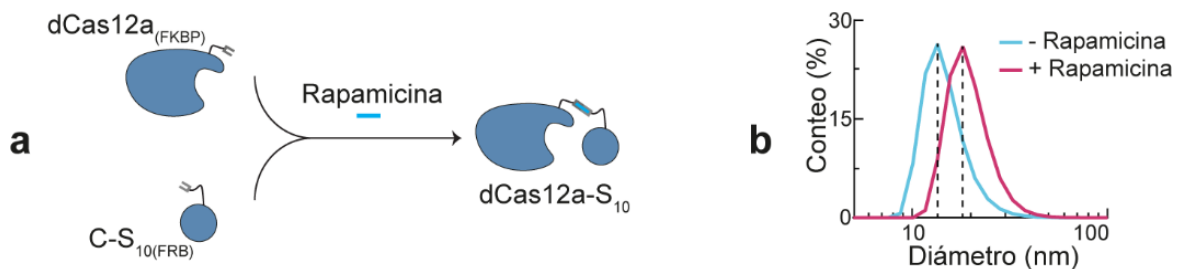


Figura 14. dCas12a_(FKBP) y C-S_{10(FRB)} se heterodimerizan en presencia de rapamicina. (a) Ilustración de la estrategia de heterodimerización usada para acoplar la unión específica al DNA (dCas12a) con el reclutamiento de C-S₁₀-B (C-S₁₀). (b) Ensayo de dispersión dinámica de la luz que muestra la heterodimerización de dCas12a_(FKBP) y C-S_{10(FRB)} inmediatamente después de añadir rapamicina.

Los crRNAs usados para la unión sitio-específica de dCas12a y dCas12a_(FKBP) al DNA se sintetizaron por transcripción *in vitro* a partir de sustratos de DNA total o parcialmente de doble hebra (**Figura 15**). La unión del crRNA a dCas12a ocasionó el apagamiento de la fluorescencia proveniente de los triptófanos de la proteína, indicando el cambio conformacional de la forma *apo* a *holo* (**Figura 16a, b**). Este cambio conformacional también se ha reportado para Cas12a de la bacteria de *Lachnospiraceae* y Cas9 de *Streptococcus pyogenes* (Dong et al., 2016; Shibata et al., 2017). Asimismo, la unión sitio-

específica del complejo dCas12a-crRNA al DNA se confirmó por microscopía de fuerza atómica (**Figura 16c**).

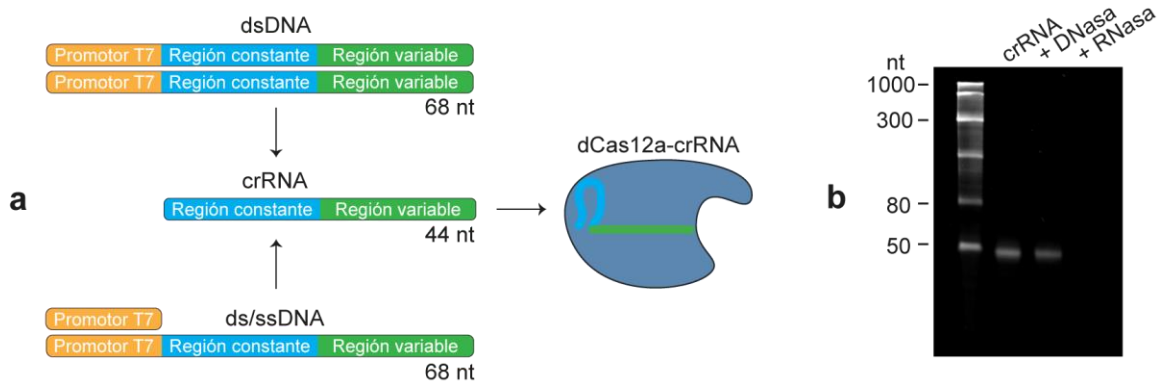


Figura 15. Síntesis de los CRISPR RNA (crRNA) para la decoración del DNA con dCas12a. (a) Fragmentos de DNA total o parcialmente de doble hebra se usaron como plantillas para la reacción de transcripción *in vitro*. Para los *pools* de crRNA, múltiples plantillas de DNA se mezclaron y se usaron en una única reacción de transcripción. (b) Urea-PAGE con uno de los crRNAs (46 nt) usados en este estudio. Tratamientos con DNasa y RNasa se usaron para descartar contaminación con el DNA sustrato.

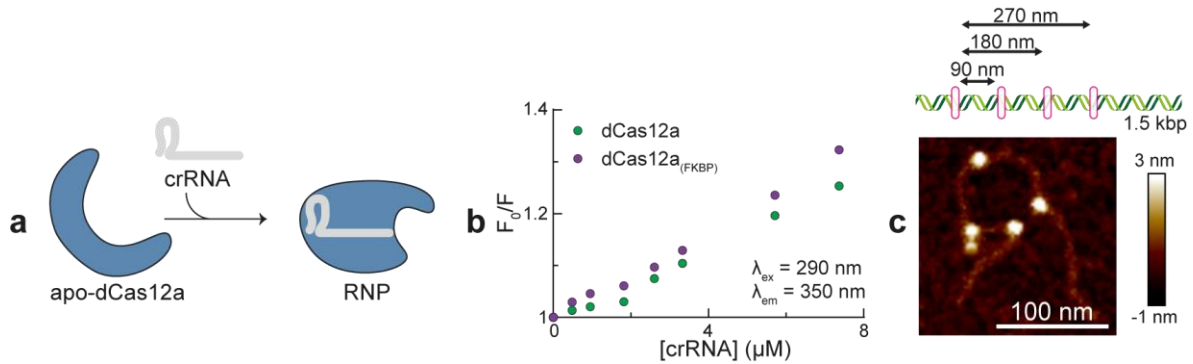


Figura 16. dCas12a une el crRNA y forma la ribonucleoproteína (RNP) para la unión específica al DNA. (a) Ilustración de la formación del complejo dCas12a-crRNA o RNP. (b) La formación de la RNP se confirmó por el apagamiento de la fluorescencia intrínseca de los triptófanos de dCas12a. (c) Ensayo de microscopía de fuerza atómica que muestra las RNPs en sus sitios diana sobre el DNA. La altura (dimensión z) se muestra en la barra de color.

Ensamblaje de las núcleo-cápsidas artificiales en tiempo real

La cinética de autoensamblaje de C-S₁₀-B sobre el DNA se caracterizó con ensayos de cortinas de DNA (Gallardo et al., 2015; Soniat et al., 2017). En este ensayo, cientos de moléculas de DNA lineares (DNA del genoma del bacteriófago λ , de tamaño 48.5 kbp y longitud de contorno 16.5 μm) se fijan por un extremo a una superficie lipídica dentro de una celda de microfluídica (**Figura 17a**). Las hebras de DNA se organizan en barreras horizontales y se extienden por flujo lento del buffer (**Figura 17b**). C-S₁₀-B se ensambla sobre el DNA y lo empaqueta en partículas de tipo rodillo de un tercio de la longitud original del DNA (Hernandez-Garcia et al., 2014) (**Figura 17c**).

El ensamblaje en tiempo real de C-S₁₀-B sobre moléculas únicas de DNA se caracterizó previamente con pinzas ópticas sujetando ambos extremos del DNA templado (Marchetti et al., 2019). No obstante, la tensión ocasionada por las pinzas impidió el crecimiento de la núcleo-cápsida, y la cobertura del DNA por C-S₁₀-B fue menos del 2%. Las cortinas de DNA permiten el ensamblaje total de la VLP manteniendo la resolución de molécula única. Asimismo, el presente ensayo permite monitorear el ensamblaje de múltiples partículas simultáneamente, garantizando la robustez estadística de los resultados. El número de réplicas y el procesamiento de las muestras son los principales desafíos de las pinzas ópticas y otros métodos usados previamente como AFM y crío-TEM (Hernandez-Garcia et al., 2014; Punter et al., 2016).

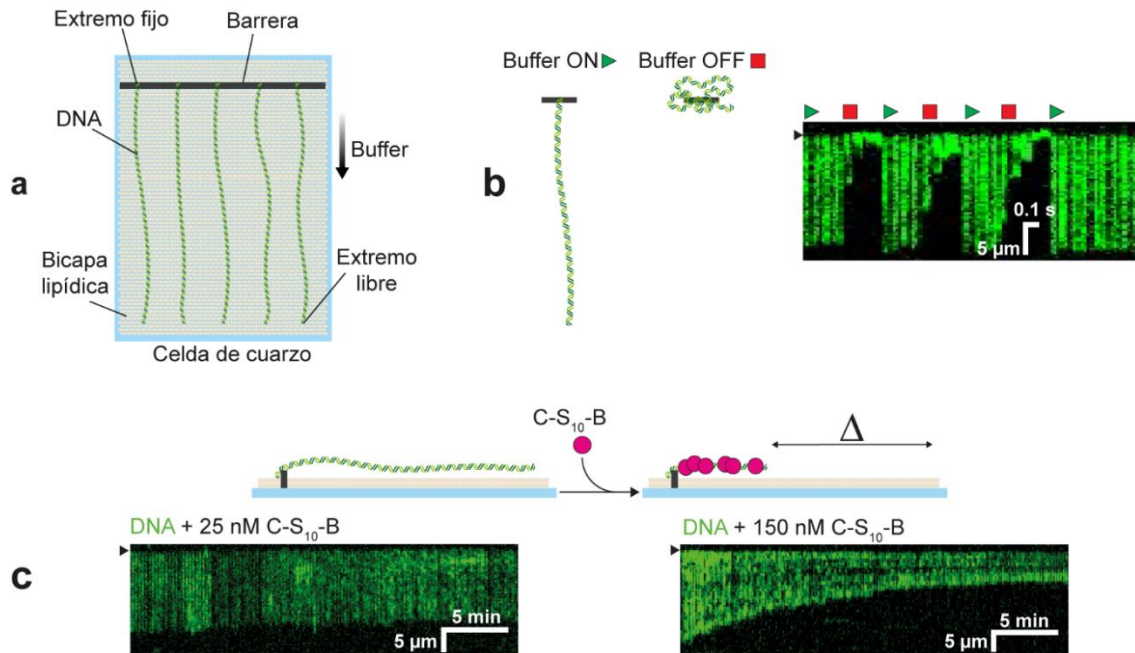


Figura 17. C-S₁₀-B se autoensambla sobre el DNA y lo empaqueta en núcleo-cápsidas. (a) Ilustración del ensayo de cortinas de DNA. Moléculas individuales de DNA se fijan a una superficie lipídica dentro de la celda de microfluídica y se extienden por flujo del buffer. (b) El flujo del buffer extiende las hebras del DNA que están ancladas a la barrera. En ausencia de flujo el DNA se retrae y se aleja del campo de vista. La posición de la barrera se indica con una flecha negra. (c) Quimogramas representativos que muestran la condensación del DNA (verde, YOYO-1) en dos concentraciones de C-S₁₀-B. La posición de la barrera se indica con una flecha negra.

La polimerización de C-S₁₀-B sobre el DNA ocurre mediante un mecanismo de nucleación-alargamiento, similar a microtúbulos y fibras amiloides. Este tipo de ensamblaje tiene varias características entre las cuales están: 1) una fase *lag* antes de la formación de las fibras y 2) una fuerte dependencia de la velocidad de polimerización con respecto a la concentración de la proteína (Oosawa & Asakura, 1975). Los perfiles de compactación de hebras individuales de DNA mostraron la fase *lag* durante el ensamblaje de C-S₁₀-B, más evidente en las concentraciones 100 y 150 nM (**Figura 18a**). Asimismo, la velocidad de

condensación del DNA aumentó con [C-S₁₀-B] y solo para el rango 100-300 nM se obtuvieron partículas completas al final del experimento (**Figura 18b**).

Todas las hebras de DNA se condensaron monotónicamente (*i.e.* en un solo sentido) sin eventos de disociación notables, independientemente de la concentración de C-S₁₀-B. Además, no se observó desensamblaje de las núcleo-cápsidas en ausencia de C-S₁₀-B, al menos en los minutos siguientes al agotamiento.

La condensación monotónica del DNA y la resistencia de las núcleo-cápsidas al desensamblaje sugieren una histéresis en el proceso de asociación-disociación. Este comportamiento, observado en el VIH-1, HVB y el bacteriófago MS2, se considera general para las cápsidas virales y es responsable de su estabilidad fuera de la célula hospedera (Garmann et al., 2019; Singh & Zlotnick, 2003; Valbuena et al., 2020; Valbuena & Mateu, 2017). La histéresis en el ensamblaje (o resistencia a la disociación) de las núcleo-cápsidas formadas por C-S₁₀-B se reporta por primera vez en este trabajo. Este resultado contribuye al conocimiento de la estabilidad y los mecanismos de autoensamblaje de las VLPs, y al estudio teórico de las mismas (es más fácil modelar irreversibilidad que reversibilidad).

Los datos de velocidad inicial contra [C-S₁₀-B] se ajustaron a la ecuación logística de cuatro parámetros o 4PL (**Figura 18c, Ecuación 1, Materiales y Métodos**), idéntica a la ecuación de Hill (Prinz, 2010). Aquí, los cuatro parámetros de la 4PL son el menor y el mayor valor de velocidad de condensación, [C-S₁₀-B]_{1/2} (concentración de C-S₁₀-B a la que se obtiene la mitad de la velocidad máxima) y el coeficiente de Hill.

[C-S₁₀-B]_{1/2} se interpreta como una constante de afinidad que es dependiente de las condiciones del experimento y resultó de 95 ± 12 nM. Este valor difiere en casi dos órdenes de magnitud del reportado (1.4 nM) por Marchetti et al. (2019). Lo anterior puede explicarse por la fuerza iónica del buffer usado en ambos experimentos: Marchetti et al. usaron un

buffer de fosfatos con fuerza iónica 25 mM y nosotros usamos un buffer con fuerza iónica de 150 mM, similar al medio fisiológico. El aumento en la concentración de sales puede explicar la menor afinidad de C-S₁₀-B, cuya interacción con el DNA es completamente entrópica y dependiente del desplazamiento de contraiones (Jantz & Berg, 2010; Lin & Riggs, 1975; Privalov et al., 2011).

Cuando el coeficiente de Hill es mayor a 1 existe cooperatividad, la cual es evidente en el comportamiento sigmoideal de la curva (**Figura 18c**), y el valor del coeficiente se usa para estimar el tamaño del núcleo (Hunter & Anderson, 2009; Van der heijden et al., 2007). En nuestro análisis, el coeficiente de Hill fue 4 ± 1 . A favor de este resultado, Marchetti et al., (2019) encontraron que las menores unidades que pueden nuclear el ensamblaje son pentámeros de C-S₁₀-B.

Los ensayos de cortinas permiten entonces recapitular las propiedades del ensamblaje de C-S₁₀-B sobre el DNA, y describir en tiempo real la formación de las núcleo-cápsidas artificiales. Aquí describimos por primera vez la aplicabilidad de este ensayo para el estudio de VLPs naturales o artificiales.

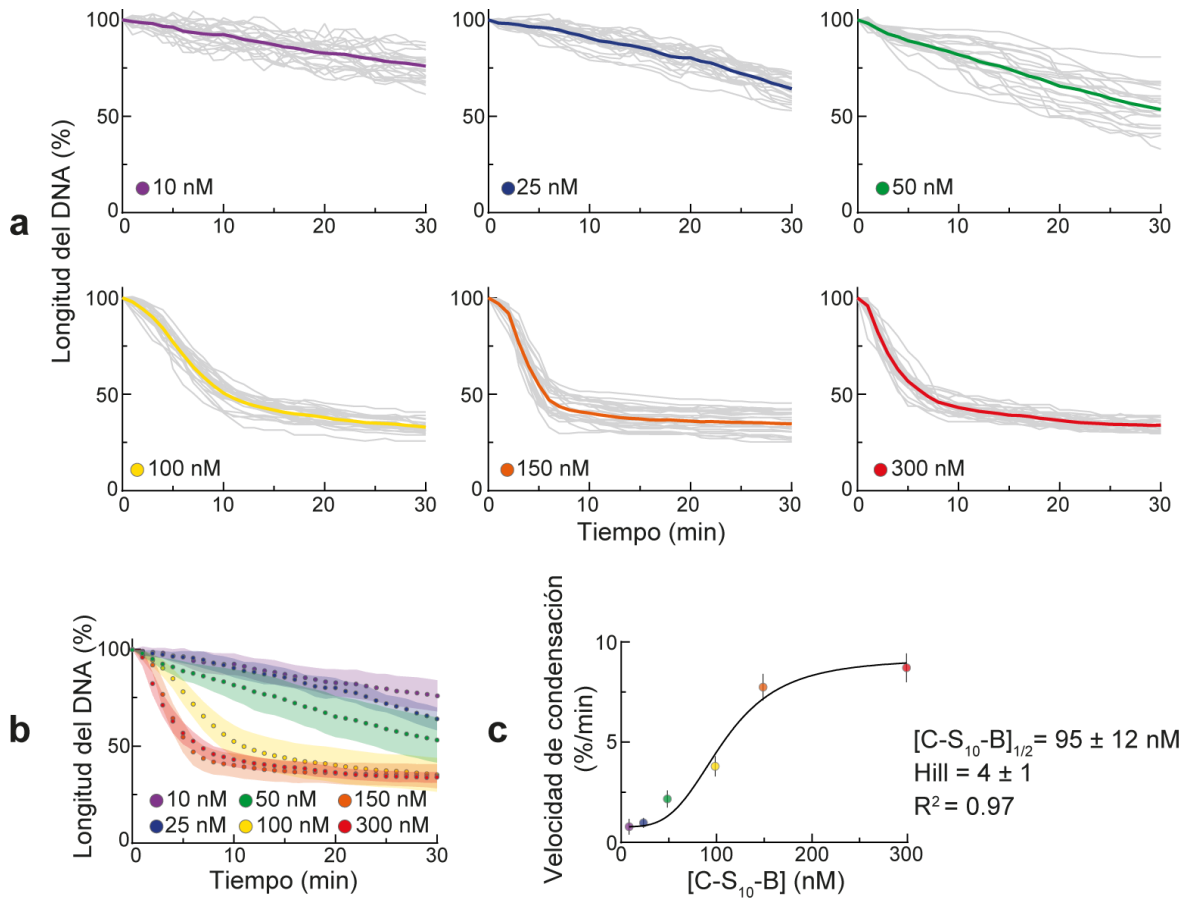


Figura 18. La concentración de C-S₁₀-B regula la velocidad de compactación del DNA.

(a) Condensación de moléculas individuales de DNA en distintas concentraciones de C-S₁₀-B. **(b)** Promedio de los perfiles de condensación obtenidos en las distintas concentraciones de C-S₁₀-B. Los círculos y áreas sombreadas indican la media \pm desviación estándar para las veinticinco moléculas de DNA en **(a)**. **(c)** Velocidad de condensación del DNA en los tres primeros minutos de incubación con C-S₁₀-B. Los valores de las pendientes se ajustaron a la ecuación de Hill (línea sólida). Las barras de error indican la media \pm intervalos de confianza 95% para veinticinco moléculas por condición.

Los perfiles temporales de compactación de DNA tuvieron un comportamiento sigmoideal, típico del ensamblaje de cápsidas virales (Casini et al., 2004; Hagan & Elrad, 2010; Prevelige et al., 1993). Una forma sencilla y comúnmente usada para describir este tipo de cinéticas es el tiempo al que se obtiene la mitad del ensamblaje (Powers & Powers, 2006). En estos experimentos de cortinas de DNA, t_{half} es el tiempo en el que el DNA está

compactado hasta el 66.5% ($=0.5 \cdot [100\% + 33\%]$) de su tamaño original, y se obtiene después de ajustar la curva a la ecuación de Hill (**Ecuación 2, Materiales y Métodos**).

La velocidad del ensamblaje, expresada como t_{half} , disminuyó de 46 ± 13 min a 8 ± 4 min entre 10-100 nM C-S₁₀-B y se saturó en el rango 100-300 nM, con un mínimo de 4 ± 1 min para 300 nM (**Figura 19a**). El coeficiente de Hill aumentó dos veces (de 1.5 ± 0.6 a 3.1 ± 0.7) entre 10 y 150 nM C-S₁₀-B, lo que indica que el incremento de la concentración de C-S₁₀-B favorece la transición de la fase de nucleación al alargamiento de la partícula. Sin embargo, este parámetro disminuyó hasta 1.8 ± 0.3 para 300 nM C-S₁₀-B (**Figura 19b**). Razonamos que con 300 nM de C-S₁₀-B se rebasa la concentración supercrítica—donde el monómero está a tan alta concentración que es éste, y no el núcleo, la especie de mayor energía. En otras palabras, a 300 nM C-S₁₀-B ocurre una polimerización irreversible de las VLPs por lo que no existe una transición entre nucleación y alargamiento (Powers & Powers, 2006).

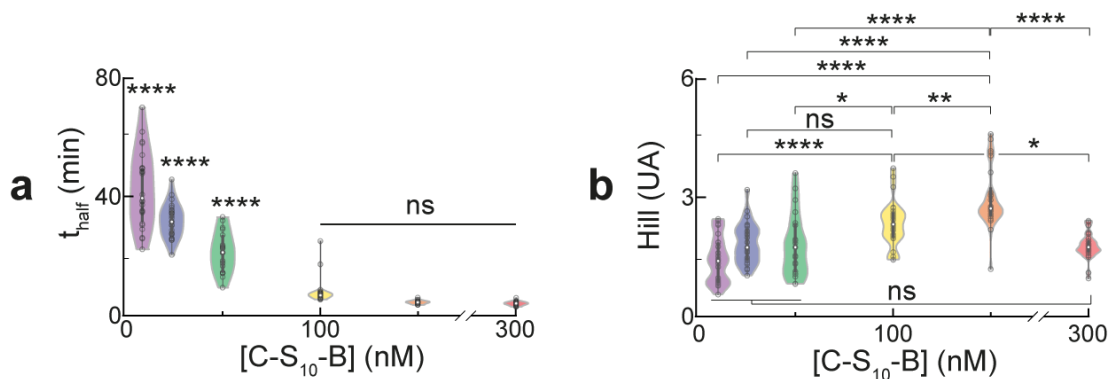


Figura 19. La concentración de C-S₁₀-B regula el tiempo necesario para alcanzar la mitad del ensamblaje y su cooperatividad. (a) El tiempo (t_{half}) requerido para alcanzar la mitad de la condensación máxima disminuye con la concentración de C-S₁₀-B. **(b)** El factor de cooperatividad (Hill) aumenta con la concentración de C-S₁₀-B y disminuye para 300 nM. Los círculos blancos y las barras gruesas en los violines representan la media y los intervalos de confianza 95% para veinticinco moléculas de DNA. ns: $p > 0.05$, * : $p < 0.05$, **: $p < 0.01$, ****: $p < 0.0001$.

La concentración de C-S₁₀-B determinó la velocidad de unión del polipéptido al DNA, confirmando los resultados de condensación (**Figura 20a**). Al final del experimento, solo el $37 \pm 8\%$ del DNA estaba recubierto por C-S₁₀-B fluorescente a 25 nM, mientras que este valor aumentó a $99 \pm 3\%$ para 150 nM C-S₁₀-B (**Figura 20b-c**).

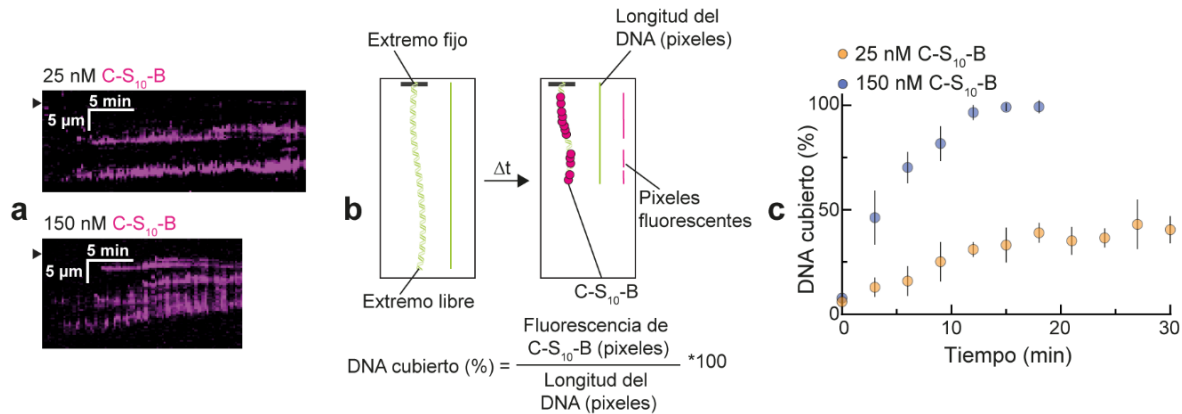


Figura 20. La concentración de C-S₁₀-B regula la velocidad de su unión al DNA. (a) Quimogramas representativos que muestran la unión de C-S₁₀-B (magenta, Alexa488) al DNA (no marcado) a 25 nM y 150 nM de C-S₁₀-B. El punto de fijación a la superficie se indica con una flecha negra. **(b)** El porcentaje de DNA cubierto con C-S₁₀-B se calculó como la razón entre la cantidad de píxeles fluorescentes y los píxeles correspondientes al DNA (no marcado). El DNA no marcado se extiende desde la barrera hasta los clústers de C-S₁₀-B que se acumulan en el extremo libre del DNA. **(c)** Recubrimiento del DNA por C-S₁₀-B fluorescente en concentración 25 nM y 150 nM. Los círculos y barras de error indican la media \pm desviación estándar para 10 moléculas de DNA.

Hasta este punto hemos caracterizado por primera vez en tiempo real la unión de C-S₁₀-B al DNA y su empaquetamiento en núcleo-cápsidas completas. Además, confirmamos la cooperatividad del ensamblaje de C-S₁₀-B, identificamos su propiedad viromimética de histéresis a la disociación, y definimos aquellas concentraciones que limitan el ensamblaje (10-50 nM). Es en este rango de concentraciones donde evaluamos si la decoración del DNA con dCas12a-S₁₀ puede acelerar el ensamblaje de las VLPs.

Decoración del DNA con dCas12a y ensamblaje de las núcleo-cápsidas

Primeramente, evaluamos si la decoración del DNA con dCas12a afecta la formación de las núcleo-cápsidas. Para esto, incubamos dCas12a con un *pool* de crRNAs complementarios a cinco o diez sitios distribuidos uniformemente sobre el λ DNA (**Figura 21a, b**). Las RNPs, marcadas fluorescentemente con puntos cuánticos (QDs), se unieron a los sitios diana (**Figura 21c**) y el número de RNP por hebra de DNA aumentó de 1-4 a 1-7 para cinco y diez sitios diana, respectivamente (**Figura 21d**). No observamos hebras de DNA con todos los sitios diana ocupados, posiblemente porque no todas las dCas12a se marcaron con anticuerpos y, a su vez, no todos los anticuerpos se conjugaron con QDs.

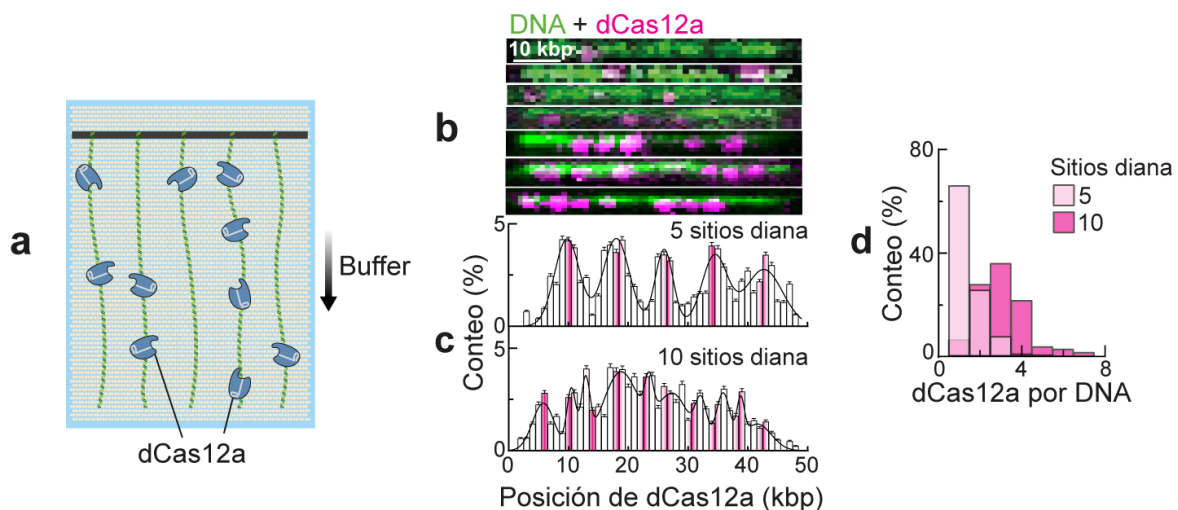


Figura 21. dCas12a se posiciona en sus sitios diana sobre el DNA vía *pools* de crRNAs. (a) Ilustración de las cortinas de DNA decoradas con dCas12a. (b) Hebras de DNA (verde, YOYO-1) decoradas con dCas12a (magenta, QD). (c) Histogramas de la distribución de dCas12a sobre el DNA. Apo-dCas12a se incubó con crRNAs complementarios a cinco (arriba, N = 561 dCas12a) o diez (debajo, N = 473 dCas12a) sitios uniformemente espaciados sobre el DNA. Las líneas magentas indican la posición de los sitios diana. Los histogramas se ajustaron a la suma de cinco o diez Gaussianas (líneas sólidas); ver Tabla S2 para los parámetros del ajuste. (d) Número de dCas12a por hebra de DNA para cinco y diez sitios de unión.

Seguidamente, incubamos el DNA predecorado con cinco o diez dCas12a junto con 25 nM de C-S₁₀-B (**Figura 22a**). La decoración con cinco y diez dCas12a aceleró la condensación del DNA dos ($t_{\text{half}} = 15 \pm 7$ min) y tres ($t_{\text{half}} = 10 \pm 3$ min) veces comparado con el DNA no decorado ($t_{\text{half}} = 31 \pm 6$ min) (**Figura 22b, c**). La decoración con diez dCas12a aumentó la cooperatividad del ensamblaje casi dos veces con respecto al DNA no decorado o decorado con cinco dCas12a, lo que sugiere un mejor alargamiento del filamento (**Figura 22d**).

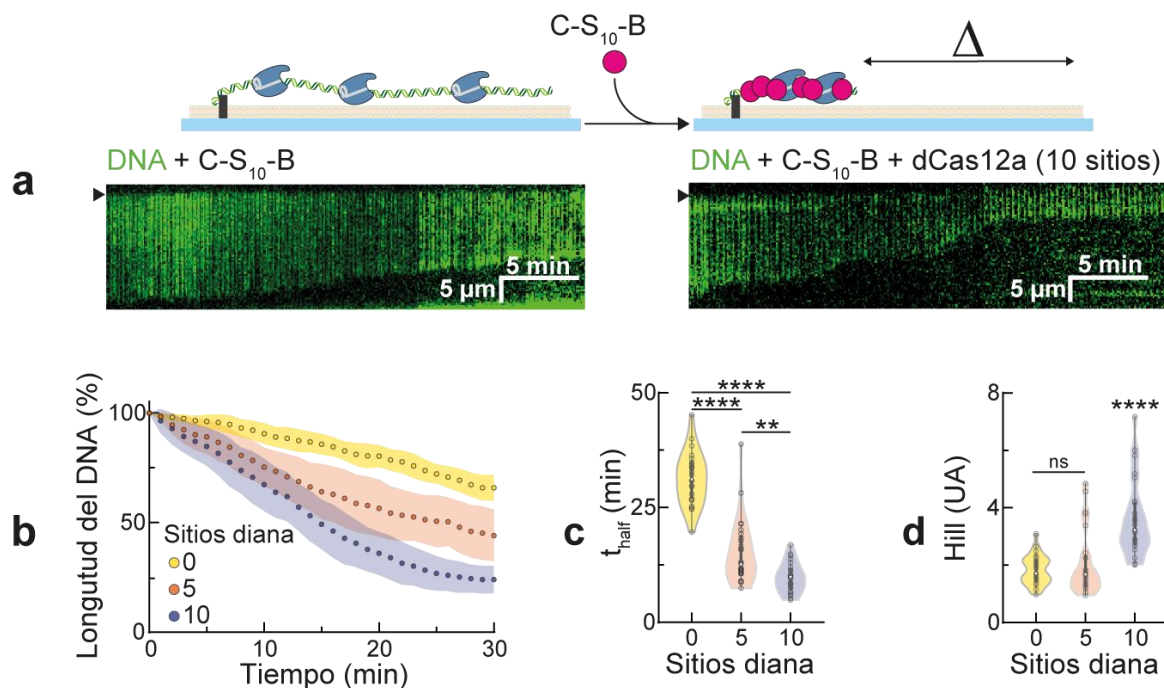


Figura 22. La decoración del DNA con dCas12a acelera la encapsidación por C-S₁₀-B.

(a) Quimogramas representativos que muestran la más rápida condensación del DNA (verde, YOYO-1) después de la decoración con dCas12a. El punto de fijación a la superficie se indica con una flecha negra. **(b)** Perfiles de condensación para DNA y DNA decorado con cinco o diez dCas12a en presencia de 25 nM C-S₁₀-B. Los círculos y áreas sombreadas indican la media \pm desviación estándar para veinticinco moléculas de DNA por condición. **(c)** El tiempo (t_{half}) requerido para alcanzar la mitad de la condensación máxima disminuye con el grado de decoración del DNA con dCas12a. Los perfiles de condensación para las moléculas que no alcanzaron el 66.5% de condensación se ajustaron a la ecuación de Hill y t_{half} se obtuvo por extrapolación. **(d)** El factor de cooperatividad (Hill) para el ensamblaje

de las núcleo-cápsidas incrementa para la decoración del DNA con diez dCas12a. Los círculos blancos y las barras gruesas en los violines representan la media y los intervalos de confianza 95% para veinticinco moléculas de DNA. ns: $p > 0.05$, (**): $p < 0.01$, (****): $p < 0.0001$.

Experimentos de doble marcaje (C-S₁₀-B con Alexa488 y dCas12a con QD) mostraron que los clústers o núcleos de C-S₁₀-B colocalizan con dCas12a en los primeros minutos del ensamblaje (**Figura 23a**). La decoración con dCas12a afectó la distribución de los núcleos de C-S₁₀-B sobre el DNA—mientras que C-S₁₀-B se acumula en el extremo libre del DNA no decorado debido al flujo del buffer, dCas12a organiza los núcleos de C-S₁₀-B alrededor de su propio sitio de unión (**Figura 23b**). El número de núcleos de C-S₁₀-B aumentó de 1-3 (1.3 ± 0.6) para DNA no decorado hasta 2-7 (3.3 ± 1.0) para DNA decorado con cinco dCas12a (**Figura 23c**). El recubrimiento del DNA con C-S₁₀-B fue tres veces más rápido para el DNA decorado ($t_{\text{half}} = 15 \pm 6$ min) con respecto al DNA no decorado ($t_{\text{half}} = 45 \pm 15$ min) (**Figura 23d, e**). Para evaluar si la filamentación de C-S₁₀-B desplaza a dCas12a, medimos el tiempo de vida de los QDs unidos al DNA (vía dCas12a) y los adsorbidos a la superficie lipídica. Ambos grupos de QDs se apagaron con cinéticas similares, indicando que el lazo R formado por el complejo dCas12a-crRNA-DNA es estable durante el empaquetamiento del DNA (**Figura 23f**).

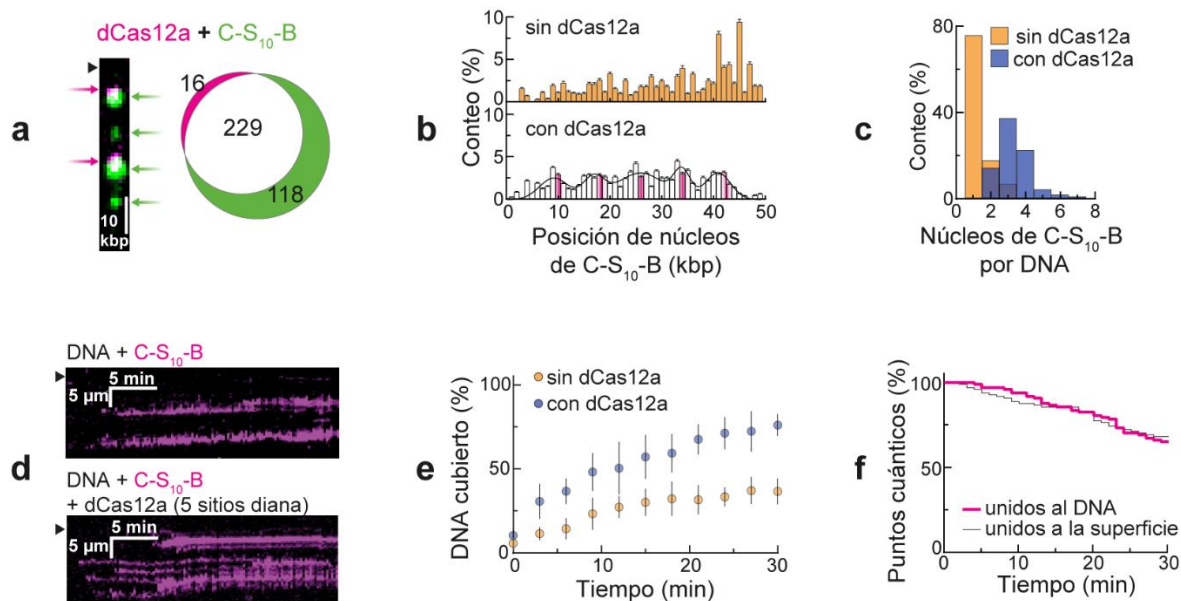


Figura 23. dCas12a favorece la nucleación de las núcleo-cápsidas. (a) Experimentos de doble marcaje que muestran la colocalización del 93% de dCas12a (magenta, QD) y el 66% de los clústers de C-S₁₀-B (verde, Alexa488). El punto de fijación a la superficie se indica con una flecha negra. **(b)** Histogramas de la distribución de los clústers de C-S₁₀-B sobre el DNA no decorado (arriba, N = 246 moléculas de DNA) o DNA decorado con cinco dCas12a (abajo, N = 484 moléculas de DNA). Las líneas magentas indican la posición de los sitios diana de dCas12a. Los histogramas se ajustaron a la suma de cinco Gaussianas (líneas sólidas); ver Tabla S2 para los parámetros del ajuste. **(c)** La pre-decoración con dCas12a aumentó el número de clústers de C-S₁₀-B por hebra (N = 484 moléculas de DNA) con respecto al DNA no decorado (N = 246 moléculas de DNA). **(d)** Quimogramas representativos que muestran que dCas12a acelera la unión de C-S₁₀-B (25 nM, magenta, Alexa488) al DNA (no marcado). El punto de fijación a la superficie se indica con una flecha negra. **(e)** Recubrimiento por C-S₁₀-B del DNA y el DNA decorado con dCas12a. Los círculos y barras de error indican la media ± desviación estándar para 10 moléculas de DNA. **(f)** Los QDs sobre el DNA (unidos a dCas12a) se apagan al mismo tiempo que los QDs unidos a la superficie lipídica en el mismo campo de vista durante el ensamblaje de C-S₁₀-B.

La aceleración de la encapsidación por dCas12a es un resultado inesperado que puede ser explicado por dos fenómenos: 1) dCas12a interactúa con C-S₁₀-B de manera específica e imprevista por nosotros, o 2) dCas12a detiene la difusión lineal de C-S₁₀-B sobre el DNA, lo que ocasiona la acumulación de C-S₁₀-B y favorece la nucleación de la VLP.

El arresto de la difusión de C-S₁₀-B por dCas12a se muestra en la **Figura 24**. En presencia del buffer circulante los clústers de C-S₁₀-B fluorescente se unen a sitios aleatorios sobre el DNA y se desplazan hacia el extremo libre o hacia sitios de nucleación. En cambio, la pre-decoración con dCas12a restringe el movimiento de C-S₁₀-B sobre el DNA y genera múltiples sitios de nucleación.

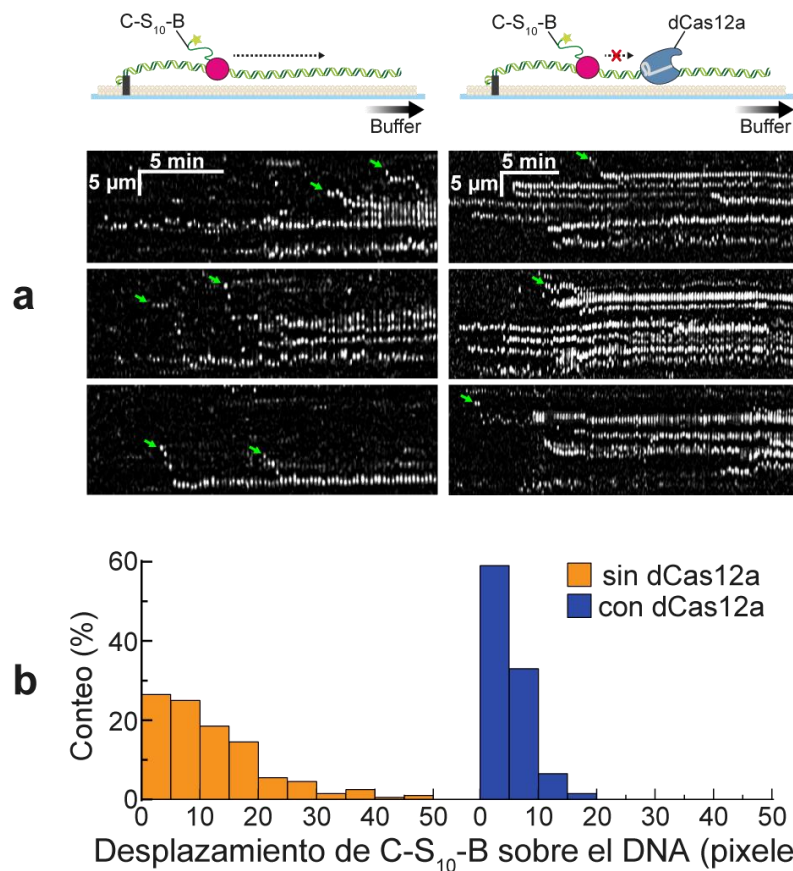


Figura 24. La decoración con dCas12a restringe la difusión lineal de C-S₁₀-B sobre el DNA. (a) Quimogramas que muestran el recubrimiento por C-S₁₀-B (blanco, Alexa488) del DNA no decorado o decorado con cinco dCas12a. Las flechas verdes indican eventos de

difusión. Para resolver los clústers de C-S₁₀-B fluorescente se procesaron las imágenes con el método de superresolución MSSR (*mean-shift super resolution*) desarrollado recientemente (Torres et al., en preparación). **(b)** La distancia recorrida por C-S₁₀-B disminuye con la decoración con dCas12a. N = 200 clústers de C-S₁₀-B.

Para descartar que una interacción específica entre dCas12a y C-S₁₀-B acelera el autoensamblaje hicimos cortinas de DNA con nucleosomas: obstáculos modelos para la difusión lineal de proteínas sobre el DNA (Gorman et al., 2010; Kanada et al., 2019; Murugan, 2018; Stigler et al., 2016). La decoración del DNA con nucleosomas (**Figura 25**) tuvo efectos idénticos a dCas12a sobre el ensamblaje de las núcleo-cápsidas: más rápida condensación del DNA (**Figura 26**) y recubrimiento con C-S₁₀-B, colocación de nucleosomas y C-S₁₀-B, e incremento del número de núcleos en etapas iniciales del ensamblaje (**Figura 27**). Estos resultados sugieren que el arresto de la difusión de C-S₁₀-B acelera el ensamblaje de las VLPs.

Concluimos hasta aquí que dCas12a se puede instalar en sitios específicos sobre el DNA mediante *pools* de crRNAs, y que la decoración con dCas12a (o cualquier otra proteína de unión estable al DNA) favorece el autoensamblaje de las núcleo-cápsidas.

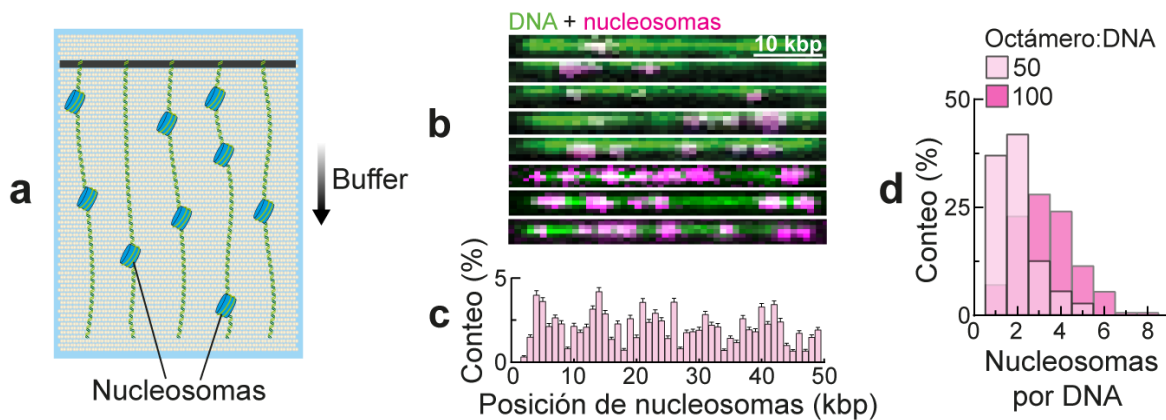


Figura 25. Los nucleosomas se posicionan sobre el DNA sin especificidad de secuencia. (a) Ilustración de la cortina de hebras de DNA decoradas con octámeros de

histonas. **(b)** Hebras de DNA (verde, YOYO-1) decoradas con nucleosomas (magenta, QD). **(c)** Histogramas con la distribución de los nucleosomas sobre el DNA para la razón molar 50:1 (octámero-DNA, N = 473 nucleosomas). **(d)** Número de nucleosomas por hebra de DNA para las razones octámero-DNA 50:1 (N = 473 nucleosomas) y 100:1 (832 nucleosomas).

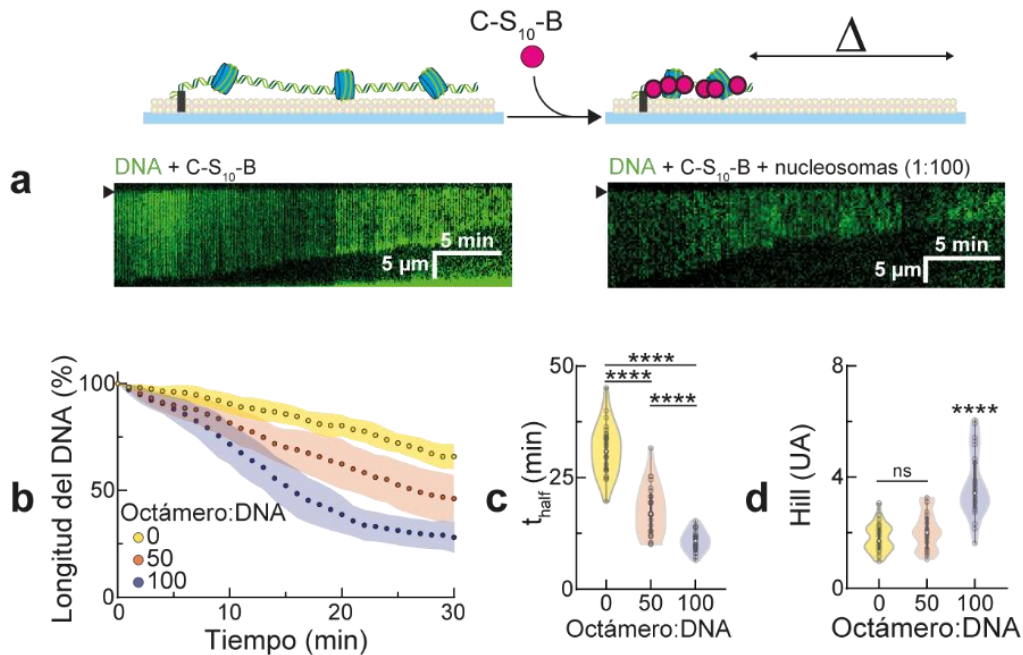


Figura 26. La decoración con nucleosomas acelera la encapsidación del DNA por C-S₁₀-B. **(a)** Quimogramas representativos que muestran la más rápida condensación del DNA (verde, YOYO-1) por C-S₁₀-B (25 nM) después de la decoración con nucleosomas (razón 100:1). El punto de fijación a la superficie se indica con una flecha negra. **(b)** Perfiles de condensación del DNA no decorado y el DNA decorado con nucleosomas a las razones octámero-DNA indicadas. Los círculos y áreas sombreadas indican la media ± desviación estándar para veinticinco moléculas de DNA. **(c)** El tiempo (t_{half}) requerido para alcanzar la mitad de la condensación máxima disminuye con el grado de decoración del DNA con nucleosomas. Los perfiles de condensación para las moléculas que no alcanzaron el 66.5% de condensación se ajustaron a la ecuación de Hill y t_{half} se obtuvo por extrapolación. **(d)** El factor de cooperatividad (Hill) incrementa para la decoración del DNA con nucleosomas en la razón 100:1. Los círculos blancos y las barras gruesas en los violines representan la

media y los intervalos de confianza 95% para veinticinco moléculas de DNA. ns: >0.05, (****): $p < 0.0001$.

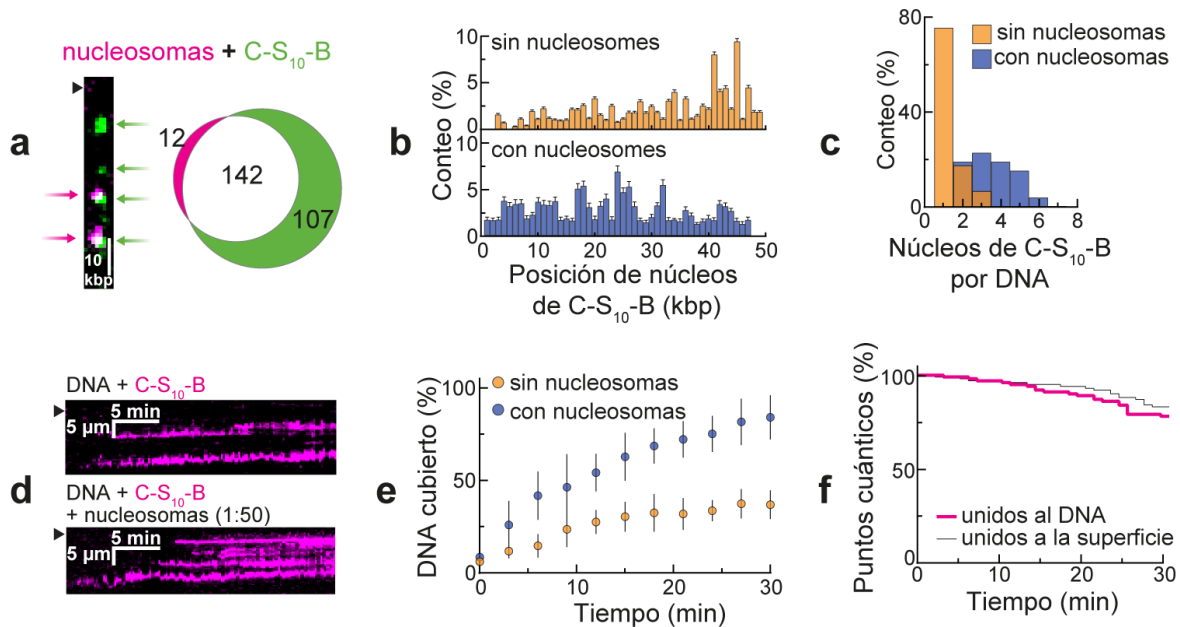


Figura 27. Los nucleosomas favorecen la nucleación de las núcleo-cápsidas. (a) Experimentos de doble marcaje que muestran la colocación del 92% de los nucleosomas (magenta, QD) y el 57% de los clústers de C-S₁₀-B (verde, Alexa488). El punto de fijación a la superficie se indica con una flecha negra. **(b)** Histogramas con la distribución de los clústers de C-S₁₀-B sobre el DNA no decorado (arriba, N = 246 moléculas de DNA) o el DNA nucleosomal (abajo, N = 362 moléculas de DNA, razón octámero-DNA 50:1). **(c)** La pre-decoración con nucleosomas aumenta el número de clústers de C-S₁₀-B sobre el DNA (N = 362 moléculas de DNA, razón 50:1) con respecto al DNA no decorado (N = 246 moléculas de DNA). **(d)** Quimogramas representativos que muestran que la unión de C-S₁₀-B (25 nM, magenta, Alexa488) al DNA (no marcado) ocurre más rápidamente en el DNA nucleosomal. El punto de fijación a la superficie se indica con una flecha negra. **(e)** Cuantificación del recubrimiento por C-S₁₀-B del DNA no decorado y el DNA nucleosomal. Los círculos y barras de error indican la media \pm desviación estándar para 10 moléculas de DNA. **(f)** El tiempo de vida de los nucleosomas (marcados con QDs) unidos al DNA es similar al tiempo de vida de los QDs unidos a la superficie lipídica en el mismo campo de vista.

Los resultados anteriores indican que la difusión sobre el DNA juega un rol principal en el ensamblaje de C-S₁₀-B. Nosotros hipotetizamos que los monómeros y oligómeros móviles de C-S₁₀-B tienen más probabilidad de polimerizarse alrededor de sus contrapartes inmóviles—ya estén inmovilizadas por un obstáculo a la difusión o por el mismo filamento de la núcleo-cápsida. A favor de esta hipótesis, Marchetti et al., (2019) reportaron que el alargamiento de las núcleo-cápsidas comienza casi exclusivamente a partir de los clústers inmóviles de C-S₁₀-B, y no de los difusionales. Además, Hernandez-Garcia et al. (2014) reportaron que la nucleación de las VLPs comienza comúnmente por uno de los extremos del DNA en templados cortos de 2.5 kbp. El arresto de la difusión de C-S₁₀-B al final del DNA podría explicar este comportamiento.

La difusión 1D—como el deslizamiento de proteínas sobre el DNA o RNA—es más rápida que la 3D y, al menos en ensayos computacionales, acelera la llegada de las proteínas de la cápsida a los sitios de ensamblaje en el genoma viral (Hu & Shklovskii, 2007; Kivenson & Hagan, 2010). La unión de proteínas virales (o incluso proteínas de la misma célula hospedera) al genoma viral podría detener la difusión 1D y crear sitios de polimerización que aceleran el ensamblaje viral. Esta hipótesis sobre el mecanismo del ensamblaje viral escapa a los objetivos de nuestra discusión y al conocimiento actual.

dCas12a-S₁₀ acelera el ensamblaje de las núcleo-cápsidas

Con el objetivo de mimetizar el rol de las señales de empaquetamiento viral, fusionamos dCas12a al dominio de polimerización S₁₀ vía rapamicina (**Figura 14**). El complejo dCas12a-S₁₀ puede interactuar directamente con C-S₁₀-B y debe acelerar el empaquetamiento del DNA con respecto a dCas12a solamente.

C-S_{10(FRB)} se incubó con la RNP (dCas12a_(FKBP)-crRNA) en presencia de rapamicina y el dímero se colocó en las cortinas de DNA (**Figura 28a**). Experimentos con C-S_{10(FRB)} fluorescente mostraron que el complejo dCas12a-S₁₀ se unió específicamente en los sitios complementarios a los crRNAs (**Figura 28b-d**).

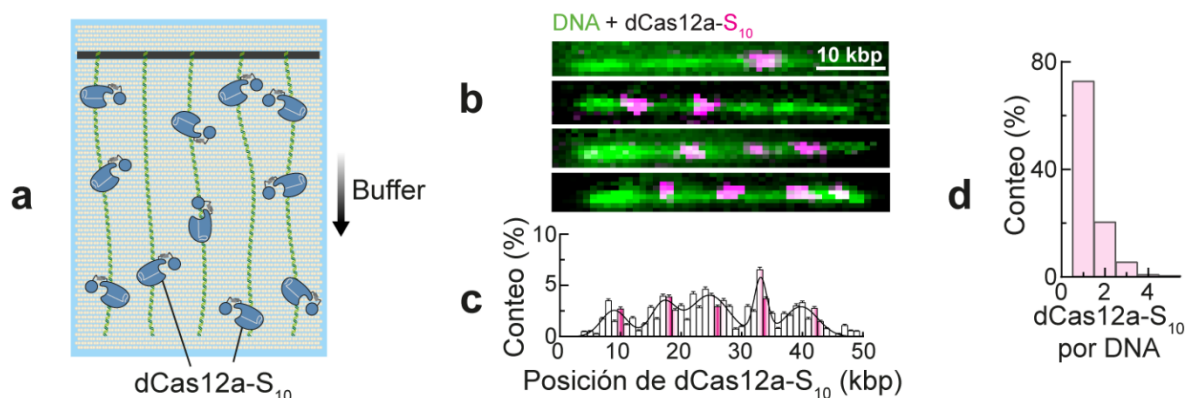


Figura 28. El complejo dCas12a-S₁₀ se posiciona en los sitios diana sobre el DNA vía *pools* de crRNA. (a) Ilustración de los sustratos de DNA decorados con dCas12a-S₁₀ formado por heterodimerización vía rapamicina. **(b)** Hebras de DNA (verde, YOYO-1) decoradas con dCas12a-S₁₀ (magenta, Atto647N). **(c)** Histogramas de la distribución de dCas12a-S₁₀ sobre el DNA con cinco sitios diana uniformemente espaciados (N = 397 dCas12a-S₁₀). Las líneas magentas indican la posición de los sitios diana. El histograma se ajustó a la suma de cinco Gaussianas (línea sólida); ver Tabla S2 para los parámetros del ajuste. **(d)** Número de dCas12a-S₁₀ por hebra de DNA (N = 397 dCas12a-S₁₀).

La decoración con dCas12a-S₁₀ aceleró la condensación del DNA ($t_{\text{half}} = 9 \pm 3$ min) dos y cinco veces con respecto a la decoración con dCas12a_(FKBP) ($t_{\text{half}} = 18 \pm 6$ min) y al DNA no decorado ($t_{\text{half}} = 46 \pm 13$ min), respectivamente (**Figura 29a-c**). La cooperatividad del ensamblaje aumentó para dCas12a-S₁₀ (3.2 ± 1.3) con respecto a dCas12a_(FKBP) (2.3 ± 0.9) y al DNA no decorado (1.4 ± 0.6), indicando una más rápida extensión del filamento sobre el DNA (**Figura 29d**).

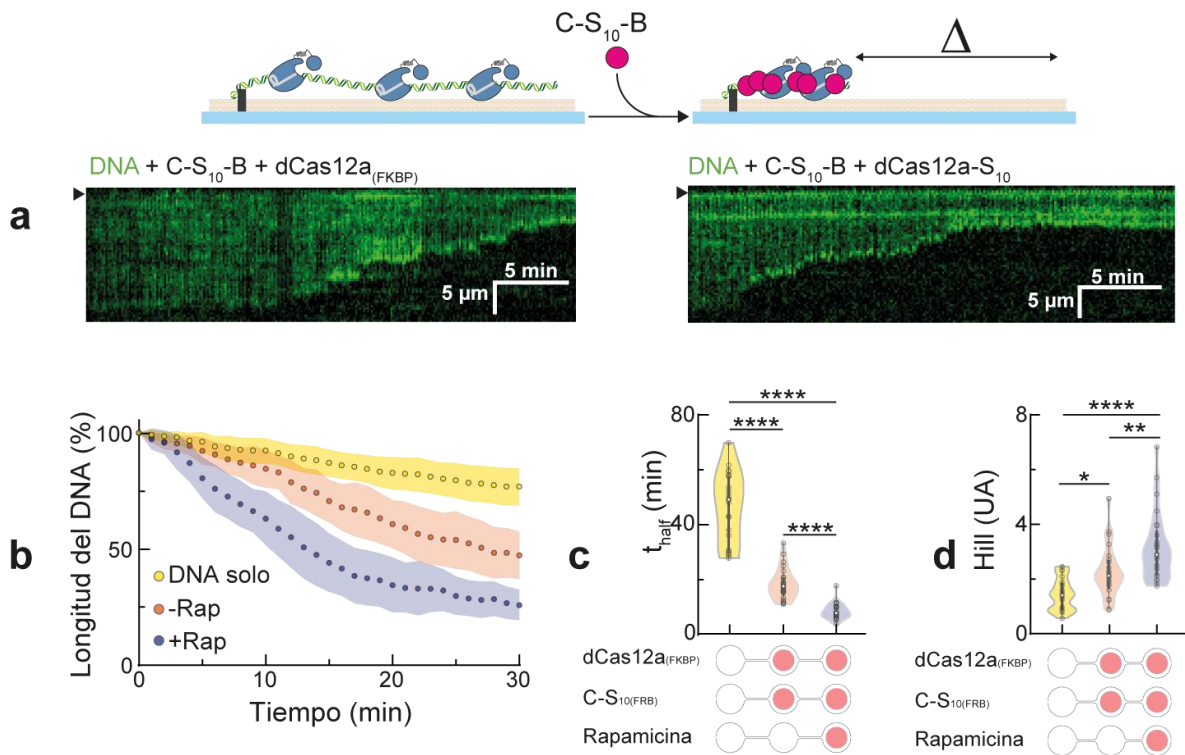


Figura 29. La decoración del DNA con dCas12a-S₁₀ favorece el autoensamblaje de las núcleo-cápsidas. (a) Quimogramas representativos que muestran que dCas12a-S₁₀ acelera la condensación del DNA (verde, YOYO-1) con respecto a dCas12a_(FKBP) en presencia de 10 nM C-S₁₀-B. El punto de fijación a la superficie se indica con una flecha negra. (b) Perfiles de condensación del DNA no decorado y el DNA decorado con cinco dCas12a_(FKBP) o cinco dCas12a-S₁₀. Los círculos y áreas sombreadas indican la media \pm desviación estándar para veinticinco moléculas de DNA. (c) El tiempo (t_{half}) requerido para alcanzar la mitad de la condensación máxima disminuye para la decoración del DNA con dCas12a-S₁₀ con respecto a dCas12a_(FKBP) y al DNA no decorado. Los perfiles de condensación para las moléculas que no alcanzaron el 66.5% de condensación se

ajustaron a la ecuación de Hill y t_{half} se obtuvo por extrapolación. **(d)** El factor de cooperatividad (Hill) para el ensamblaje de las núcleo-cápsidas incrementa con la decoración del DNA con dCas12a_(FKBP) y dCas12a-S₁₀. Los círculos blancos y las barras gruesas en los violines representan la media y los intervalos de confianza 95% para veinticinco moléculas de DNA. (*): $p < 0.05$, (**): $p < 0.01$, (****): $p < 0.0001$.

dCas12a-S₁₀ favoreció la condensación de un fragmento lineal de DNA de 2.5 kbp y de un plásmido de 9.5 kbp en ensayos de retardo de movilidad electroforética (EMSA, **Figura 30**). Estos resultados indican que dCas12a-S₁₀ nuclea los filamentos de C-S₁₀-B sobre el DNA diana y promueve el ensamblaje de las VLPs a concentraciones limitantes de C-S₁₀-B.

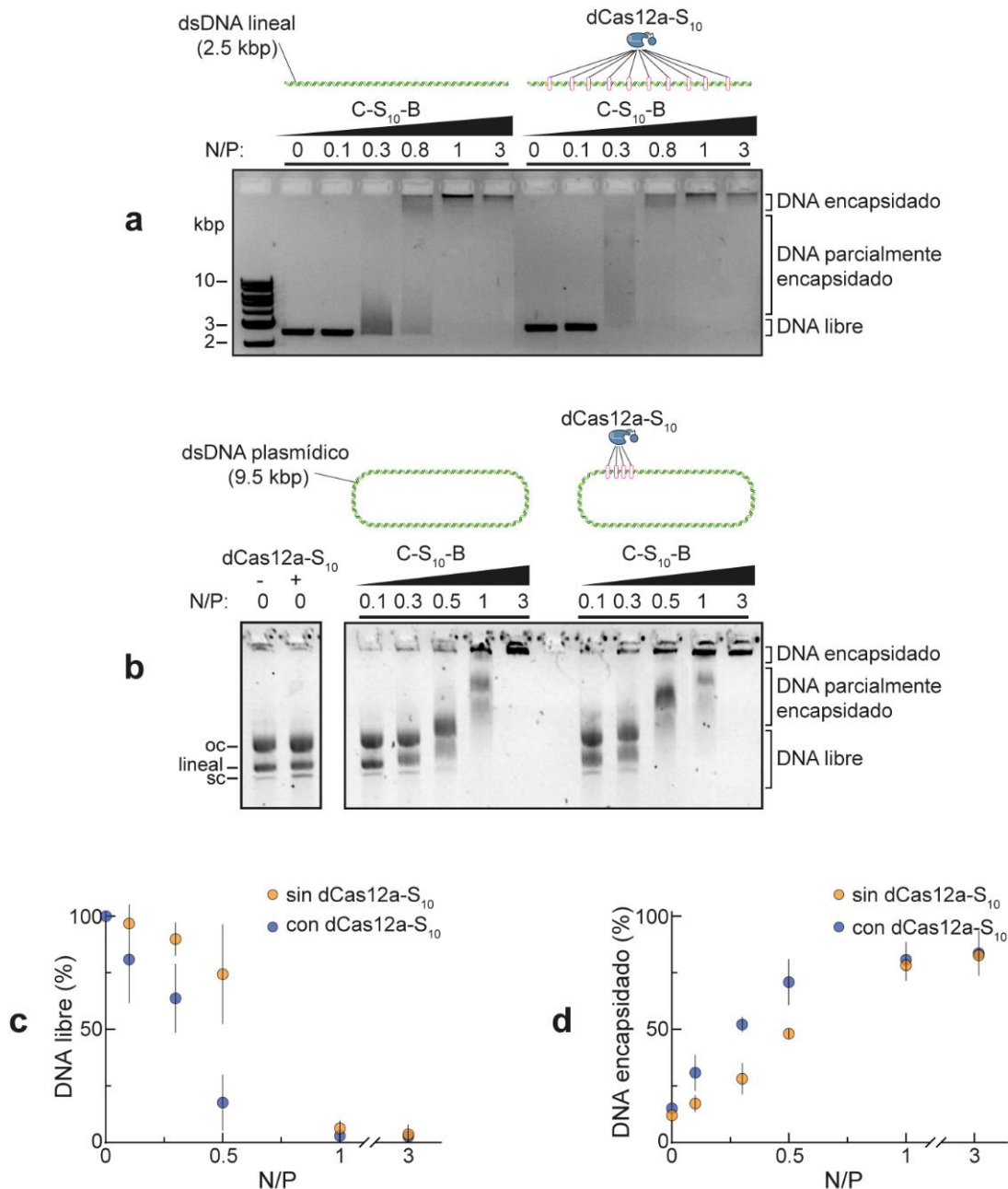


Figura 30. La decoración del DNA con dCas12a-S₁₀ favorece la encapsulación en concentraciones sub-saturantes de C-S₁₀-B. El ensamblaje de las VLPs se evaluó con un ensayo de retardo de movilidad electroforética porque la unión de C-S₁₀-B neutraliza las cargas del DNA. **(a)** El DNA templado de 2.5 kbp se decoró con dCas12a-S₁₀ en diez sitios diferentes y uniformemente espaciados. **(b)** El plásmido de DNA templado de 9.5 kbp se decoró con dCas12a-S₁₀ en cuatro sitios idénticos y cercanos entre sí. El plásmido se encuentra en las conformaciones superenrollado (sc), lineal y círculo abierto (oc). **(c)** Intensidad de la banda correspondiente al DNA libre en **(b)**. **(d)** Intensidad de la banda

correspondiente al DNA completamente encapsidado en **(b)**. Los círculos y las barras de error representan la media \pm desviación estándar para tres réplicas. N/P: razón estequiométrica entre C-S₁₀-B y su sitio de unión en el DNA (6 bp).

La decoración del DNA con dCas12a-S₁₀ o dCas12a_(FKBP) no afectó la morfología de las núcleo-cápsidas, las cuales mostraron la típica geometría tipo rodillo y tamaño de un tercio de la longitud original del DNA (**Figura 31**).

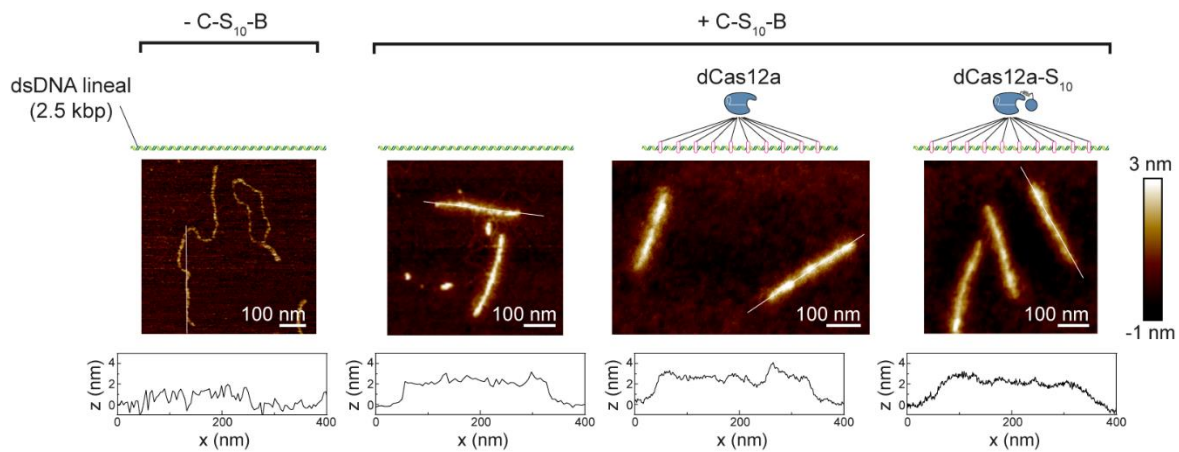


Figura 31. C-S₁₀-B empaqueta el DNA, y el DNA decorado con dCas12a y dCas12a-S₁₀ en núcleo-cápsidas artificiales. Se muestran el DNA templado (2.5 kbp, longitud de contorno: 850 nm) y las partículas obtenidas a partir del DNA no decorado y decorado (con dCas12a o dCas12a-S₁₀) después de la incubación con C-S₁₀-B. La altura de las partículas fue de 2-4 nm, similar a lo reportado por Hernandez-Garcia et al. (2014).

Ensamblaje de C-S₁₀-B sobre moléculas diana de DNA

Durante el ensamblaje viral, las proteínas de la cápsida reconocen el genoma viral en el citosol de la célula hospedera entre un exceso de ácidos nucleicos. Al menos para los virus de RNA, este exceso se ha estimado de al menos dos órdenes de magnitud (Berkowitz et al., 1996). Como se mencionó anteriormente, las señales de empaquetamiento son las responsables de tal discriminación molecular.

dCas12a-S₁₀ se puede dirigir a moléculas específicas de DNA para acelerar su empaquetamiento en núcleo-cápsidas mediante la complementariedad crRNA-DNA. Para evaluar el potencial discriminante de dCas12a-S₁₀ durante el ensamblaje de C-S₁₀-B hicimos un ensayo de cortinas con una mezcla equimolar de dos sustratos de DNA.

El DNA del bacteriófago λ (48.5 kbp) se biotiniló en ambos extremos y se cortó con la enzima NgoMIV para generar dos fragmentos (20 kbp y 28.5 kbp) (**Figura 32a, b**). Cinco, diez o veinticinco dCas12a-S₁₀ se dirigieron a la hebra diana (28.5 kbp) que se distingue de la hebra no diana por su tamaño (**Figura 32c**). La decoración con dCas12a-S₁₀ aceleró la condensación del DNA diana (más largo) con respecto al DNA no diana (más corto) en todos los casos, y con mayor velocidad a mayores grados de decoración: $t_{\text{half}} = 62 \pm 26$ min sin decoración hasta $t_{\text{half}} = 19 \pm 6$ min para la decoración con veinticinco dCas12a-S₁₀ (**Figura 32d-f**). Asimismo, el ensamblaje de C-S₁₀-B fue más cooperativo para el DNA decorado, indicando un mejor crecimiento de estas núcleo-cápsidas (**Figura 32g**).

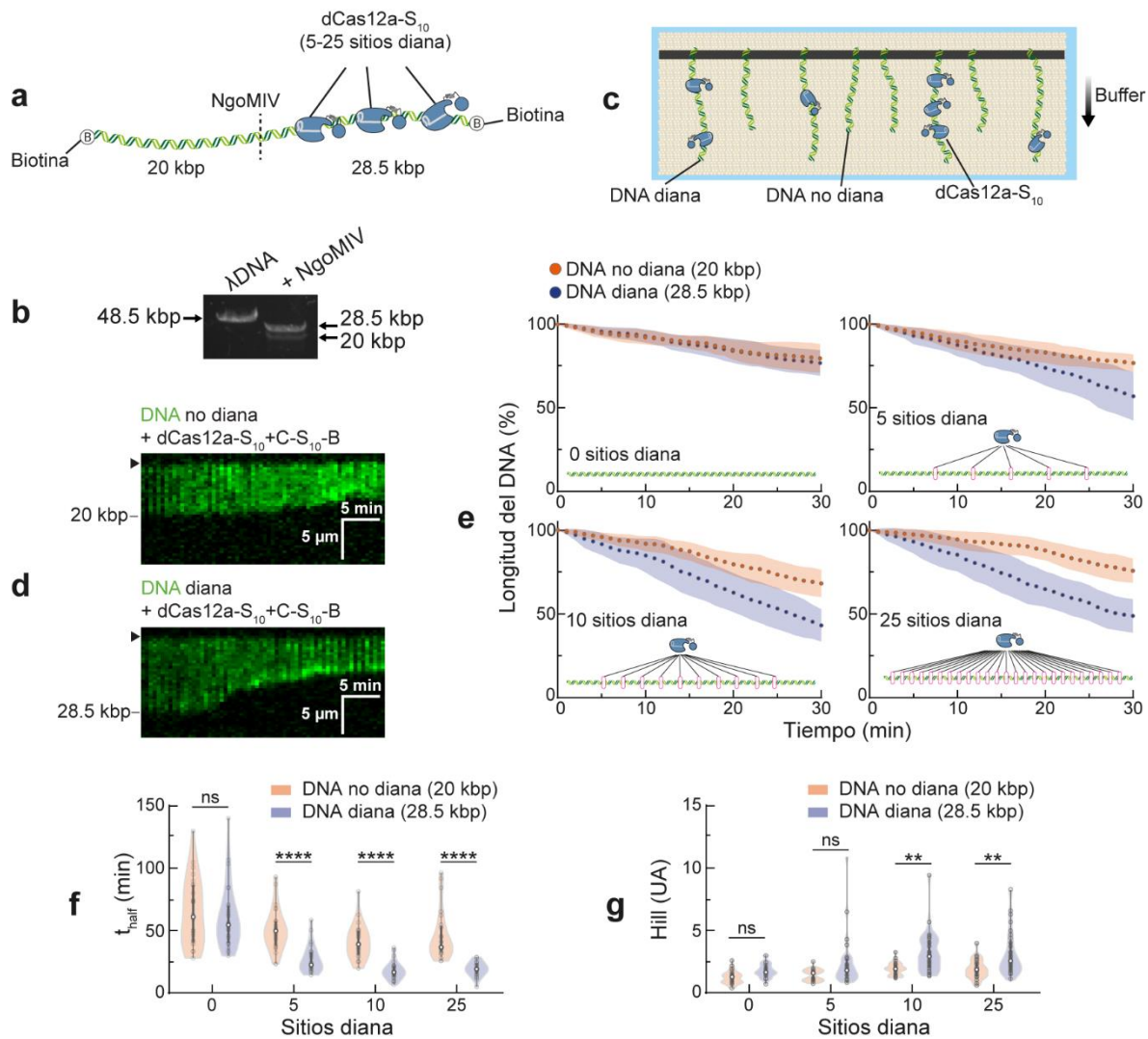


Figura 32. C-S₁₀-B encapsula selectivamente el DNA que ha sido decorado con dCas12a-S₁₀. (a) El λ DNA se ligó a un oligo biotinilado en cada extremo y se digirió con NgoMIV para generar dos sustratos de 20 kbp y 28.5 kbp, distinguibles en los ensayos de cortinas por su tamaño. (b) Gel de agarosa que muestra los dos templados de DNA obtenidos por el corte con NgoMIV. (c) La hebra diana (28.5 kbp) se decoró con cinco a veinticinco dCas12a-S₁₀ antes de la incubación con 10 nM C-S₁₀-B. (d) Quimogramas representativos que muestran la condensación de las hebras de DNA (verde, YOYO-1) no diana (arriba) y diana (abajo, con diez dCas12a-S₁₀). El punto de fijación a la superficie se indica con una flecha negra. (e) Perfiles de condensación en 10 nM de C-S₁₀-B para el DNA no diana (20 kbp) y el DNA diana (28.5 kbp) decorado con cinco a veinticinco dCas12a-S₁₀. Los círculos y áreas sombreadas indican la media \pm desviación estándar para veinticinco moléculas de DNA. (f) El tiempo (t_{half}) requerido para alcanzar la mitad de la condensación

máxima es menor para el DNA diana con respecto DNA no diana y esta diferencia aumenta con el grado de decoración con dCas12a-S₁₀ (N = 25 moléculas de DNA). Los perfiles de condensación para las moléculas que no alcanzaron el 66.5% de condensación se ajustaron a la ecuación de Hill y t_{half} se obtuvo por extrapolación. **(g)** El factor de cooperatividad (Hill) para el ensamblaje de C-S₁₀-B es mayor para el DNA diana. Los círculos blancos y las barras gruesas en los violines representan la media y los intervalos de confianza 95% para 25 moléculas de DNA. ns: $p > 0.05$, (**): $p < 0.01$, (****): $p < 0.0001$.

Consideraciones finales

En resumen, hemos demostrado que dCas12a (o cualquier proteína que se una establemente al DNA) puede servir como un sitio de nucleación para las núcleo-cápsidas, sobre todo si la proteína contiene un dominio de autoensamblaje.

Las proteínas CRISPR-Cas son especialmente útiles para programar las señales de empaquetamiento. El simple rediseño del crRNA modifica la secuencia que reclutará la proteína viromimética sin necesidad de manipular dominios proteicos. La única condición para el direccionamiento de CRISPR-Cas es la presencia del motivo adyacente al protoespaciador (PAM) flanqueando la secuencia diana. El PAM para nuestra dCas12a de *Acidaminococcus* sp. (TTTV) aparece cada 32 bp en promedio a lo largo del λ DNA de 48.5 kbp. No obstante, variantes de Cas9 y Cas12a con PAMs relajados podrían ampliar las posibilidades de nuestra estrategia (Gao et al., 2017; Leenay & Beisel, 2017; Nishimasu et al., 2018; Walton et al., 2020).

Demostramos que C-S₁₀-B se puede ingenierizar para adquirir nuevas propiedades virales. Todos los bloques de esta proteína se han modificado anteriormente para lograr distintos objetivos: desensamblaje responsivo (Beun et al., 2014), cambiar el organismo de expresión (Willems et al., 2019), direccionar las núcleo-cápsidas a ciertos organismos (resultados no publicados del grupo), cambiar la afinidad a DNA (Hernandez-Garcia et al., 2017), y ahora para reconocer secuencias específicas de DNA (Calcines-Cruz et al., 2021). El caso de C-S₁₀-B demuestra el potencial de las estrategias *bottom-up* para responder a problemas actuales de la (bio/nano)ciencia.

Conclusiones

- Se caracterizó en tiempo real el empaquetamiento de moléculas individuales de DNA en núcleo-cápsidas tipo-virus y se identificó el rango de concentraciones de C-S₁₀-B que limitan el ensamblaje.

- La decoración del DNA con múltiples dCas12a vía *pools* de crRNAs nucleó el ensamblaje de las partículas y aceleró de 100-200% el empaquetamiento del DNA en concentraciones limitantes de C-S₁₀-B. La decoración del DNA con nucleosomas tuvo un efecto idéntico a dCas12a, lo que indica que la aceleración del ensamblaje se debe al arresto de la difusión lineal de C-S₁₀-B sobre el DNA.

- Inspirados en las proteínas que reconocen las señales de empaquetamiento en virus naturales, acoplamos dCas12a al dominio de cooperatividad S₁₀ para reclutar C-S₁₀-B. La decoración del DNA con el nuevo complejo dCas12a-S₁₀ aceleró el ensamblaje de las núcleo-cápsidas dos veces con respecto a dCas12a solamente, y cinco veces con respecto al DNA no decorado.

- Múltiples complejos para el reconocimiento de señales de empaquetamiento (dCas12a-S₁₀) se pueden colocar sobre moléculas de DNA para su encapsidación preferencial con respecto a moléculas de DNA no dianas.

- Los resultados presentados muestran que las proteínas asociadas a CRISPR pueden acelerar la encapsidación de cualquier molécula de DNA, y posiblemente también RNA, en VLPs de manera específica y sencilla. Esta nueva estrategia puede ser útil para el desarrollo de sistemas de entrega de genes y bio-nanomateriales novedosos.

Perspectivas

1. Estudiar el efecto de la densidad y la posición de las señales de empaquetamiento en el ensamblaje de las VLPs programables.
2. Caracterizar el efecto de agentes que retarden la difusión de C-S₁₀-B sobre el DNA en el ensamblaje de las VLPs.
3. Estandarizar una técnica “a granel” de fluorescencia para evaluar el ensamblaje de las núcleo-cápsidas en tiempo real, como tioflavina-T o YOYO-1. Este tipo de ensayos serviría para complementar y profundizar los resultados obtenidos aquí.
4. Identificar dominios de unión a secuencias específicas de DNA o RNA (provenientes de factores de transcripción y proteínas de unión a RNA), preferentemente pequeños y de fácil producción, para crear VLPs con ensamblaje selectivo.
5. Profundizar el análisis de las cinéticas de autoensamblaje con modelos de ecuaciones diferenciales y Monte Carlo cinético. Estas aproximaciones, que actualmente llevamos a cabo con colaboradores, pueden describir mecanísticamente el ensamblaje de las VLPs, y el efecto de la decoración con dCas12a y dCas12a-S₁₀.

Bibliografía

- Azuma, Y., Edwardson, T. G. W., Terasaka, N., & Hilvert, D. (2018). Modular Protein Cages for Size-Selective RNA Packaging in Vivo. *Journal of the American Chemical Society*, *140*(2), 566–569. <https://doi.org/10.1021/jacs.7b10798>
- Bale, J. B., Gonen, S., Liu, Y., Sheffler, W., Ellis, D., Thomas, C., Cascio, D., Yeates, T. O., Gonen, T., King, N. P., & Baker, D. (2016). Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science*, *353*(6297), 389–394. <https://doi.org/10.1126/science.aaf8818>
- Banaszynski, L. A., Liu, C. W., & Wandless, T. J. (2005). Characterization of the FKBP-rapamycin-FRB ternary complex. *Journal of the American Chemical Society*, *127*(13), 4715–4721. <https://doi.org/10.1021/ja043277y>
- Berkowitz, R., Fisher, J., & Goff, S. P. (1996). RNA packaging. *Current Topics in Microbiology and Immunology*, *214*, 177–218. https://doi.org/10.1007/978-3-642-80145-7_6
- Beun, L. H., Storm, I. M., Werten, M. W. T., De Wolf, F. A., Cohen Stuart, M. A., & De Vries, R. (2014). From micelles to fibers: Balancing self-assembling and random coiling domains in pH-responsive silk-collagen-like protein-based polymers. *Biomacromolecules*, *15*(9), 3349–3357. <https://doi.org/10.1021/bm500826y>
- Borodavka, A., Tuma, R., & Stockley, P. G. (2012). Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(39), 15769–15774. <https://doi.org/10.1073/pnas.1204357109>
- Brown, R. S., Anastasakis, D. G., Hafner, M., & Kielian, M. (2020). Multiple capsid protein binding sites mediate selective packaging of the alphavirus genomic RNA. *Nature Communications*, *11*(1). <https://doi.org/10.1038/s41467-020-18447-z>
- Butterfield, G. L., Lajoie, M. J., Gustafson, H. H., Sellers, D. L., Nattermann, U., Ellis, D., Bale, J. B., Ke, S., Lenz, G. H., Yehdego, A., Ravichandran, R., Pun, S. H., King, N. P., & Baker, D. (2017). Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature*, *552*(7685), 415–420. <https://doi.org/10.1038/nature25157>
- Calcines-Cruz, C., Finkelstein, I. J., & Hernandez-Garcia, A. (2021). CRISPR-Guided Programmable Self-Assembly of Artificial Virus-Like Nucleocapsids. *Nano Letters*, 10–15. <https://doi.org/10.1021/acs.nanolett.0c04640>
- Casini, G. L., Graham, D., Heine, D., Garcea, R. L., & Wu, D. T. (2004). In vitro papillomavirus capsid assembly analyzed by light scattering. *Virology*, *325*(2), 320–327. <https://doi.org/10.1016/j.virol.2004.04.034>
- Chai, S., Lurz, R., & Alonso, J. C. (1995). The small subunit of the terminase enzyme of *Bacillus subtilis* bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. *Journal of Molecular Biology*, *252*(4), 386–398. <https://doi.org/10.1006/jmbi.1995.0505>
- Chandler-Bostock, R., Mata, C. P., Bingham, R. J., Dykeman, E. C., Meng, B., Tuthill, T. J., Rowlands, D. J., Ranson, N. A., Twarock, R., & Stockley, P. G. (2020). Assembly

- of infectious enteroviruses depends on multiple, conserved genomic RNA-coat protein contacts. *PLoS Pathogens*, *16*(12), 1–23. <https://doi.org/10.1371/journal.ppat.1009146>
- Cingil, H. E., Boz, E. B., Biondaro, G., De Vries, R., Cohen Stuart, M. A., Kraft, D. J., Van der Schoot, P., & Sprakel, J. (2017). Illuminating the Reaction Pathways of Viromimetic Assembly. *Journal of the American Chemical Society*, *139*(13), 4962–4968. <https://doi.org/10.1021/jacs.7b01401>
- Comas-Garcia, M., Datta, S. A. K., Baker, L., Varma, R., Gudla, P. R., & Rein, A. (2017). Dissection of specific binding of HIV-1 Gag to the 'packaging signal' in viral RNA. *ELife*, *6*. <https://doi.org/10.7554/eLife.27055>
- D'Souza, V., Melamed, J., Habib, D., Pullen, K., Wallace, K., & Summers, M. F. (2001). Identification of a high affinity nucleocapsid protein binding element within the Moloney murine leukemia virus ψ -RNA packaging signal: Implications for genome recognition. *Journal of Molecular Biology*, *314*(2), 217–232. <https://doi.org/10.1006/jmbi.2001.5139>
- Dang, X., Yi, H., Ham, M. H., Qi, J., Yun, D. S., Ladewski, R., Strano, M. S., Hammond, P. T., & Belcher, A. M. (2011). Virus-templated self-assembled single-walled carbon nanotubes for highly efficient electron collection in photovoltaic devices. *Nature Nanotechnology*, *6*(6), 377–384. <https://doi.org/10.1038/nnano.2011.50>
- De Beer, T., Fang, J., Ortega, M., Yang, Q., Maes, L., Duffy, C., Berton, N., Sippy, J., Overduin, M., Feiss, M., & Catalano, C. E. (2002). Insights into specific DNA recognition during the assembly of a viral genome packaging machine. *Molecular Cell*, *9*(5), 981–991. [https://doi.org/10.1016/S1097-2765\(02\)00537-3](https://doi.org/10.1016/S1097-2765(02)00537-3)
- Ding, P., Kharytonchik, S., Waller, A., Mbaekwe, U., Basappa, S., Kuo, N., Frank, H. M., Quasney, C., Kidane, A., Swanson, C., Van, V., Sarkar, M., Cannistraci, E., Chaudhary, R., Flores, H., Telesnitsky, A., & Summers, M. F. (2020). Identification of the initial nucleocapsid recognition element in the HIV-1 RNA packaging signal. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(30), 17737–17746. <https://doi.org/10.1073/pnas.2008519117>
- Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., Yang, D., Ma, C., Wang, S., Wu, D., Ma, Y., Fan, S., Wang, J., Gao, N., & Huang, Z. (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature*, *532*(7600), 522–526. <https://doi.org/10.1038/nature17944>
- Gaj, T., Gersbach, C. A., & Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. In *Trends in Biotechnology* (Vol. 31, Issue 7, pp. 397–405). Elsevier Current Trends. <https://doi.org/10.1016/j.tibtech.2013.04.004>
- Gallardo, I. F., Pasupathy, P., Brown, M., Manhart, C. M., Neikirk, D. P., Alani, E., & Finkelstein, I. J. (2015). High-Throughput Universal DNA Curtain Arrays for Single-Molecule Fluorescence Imaging. *Langmuir*, *31*(37), 10310–10317. <https://doi.org/10.1021/acs.langmuir.5b02416>
- Gao, L., Cox, D. B. T., Yan, W. X., Manteiga, J. C., Schneider, M. W., Yamano, T., Nishimasu, H., Nureki, O., Crosetto, N., & Zhang, F. (2017). Engineered Cpf1 variants with altered PAM specificities. *Nature Biotechnology*, *35*(8), 789–792. <https://doi.org/10.1038/nbt.3900>

- Garmann, R. F., Goldfain, A. M., & Manoharan, V. N. (2019). Measurements of the self-assembly kinetics of individual viral capsids around their RNA genome. *Proceedings of the National Academy of Sciences of the United States of America*, 116(45), 22485–22490. <https://doi.org/10.1073/pnas.1909223116>
- Gorman, J., Plys, A. J., Visnapuu, M. L., Alani, E., & Greene, E. C. (2010). Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nature Structural and Molecular Biology*, 17(8), 932–938. <https://doi.org/10.1038/nsmb.1858>
- Guilinger, J. P., Thompson, D. B., & Liu, D. R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature Biotechnology*, 32(6), 577–582. <https://doi.org/10.1038/nbt.2909>
- Hagan, M. F., & Elrad, O. M. (2010). Understanding the concentration dependence of viral capsid assembly kinetics - The origin of the lag time and identifying the critical nucleus size. *Biophysical Journal*, 98(6), 1065–1074. <https://doi.org/10.1016/j.bpj.2009.11.023>
- Hernandez-Garcia, A., Estrich, N. A., Werten, M. W. T., Van Der Maarel, J. R. C., LaBean, T. H., De Wolf, F. A., Cohen Stuart, M. A., & De Vries, R. (2017). Precise Coating of a Wide Range of DNA Templates by a Protein Polymer with a DNA Binding Domain. *ACS Nano*, 11(1), 144–152. <https://doi.org/10.1021/acsnano.6b05938>
- Hernandez-Garcia, A., Kraft, D. J., Janssen, A. F. J., Bomans, P. H. H., Sommerdijk, N. A. J. M., Thies-Weesie, D. M. E., Favretto, M. E., Brock, R., De Wolf, F. A., Werten, M. W. T., Van Der Schoot, P., Stuart, M. C., & De Vries, R. (2014). Design and self-assembly of simple coat proteins for artificial viruses. *Nature Nanotechnology*, 9(9), 698–702. <https://doi.org/10.1038/nnano.2014.169>
- Hernandez-Garcia, A., Werten, M. W. T., Stuart, M. C., De Wolf, F. A., & De Vries, R. (2012). Coating of single DNA molecules by genetically engineered protein diblock copolymers. *Small*, 8(22), 3491–3501. <https://doi.org/10.1002/smll.201200939>
- Hu, T., & Shklovskii, B. I. (2007). Kinetics of viral self-assembly: Role of the single-stranded RNA antenna. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 75(5), 12–16. <https://doi.org/10.1103/PhysRevE.75.051901>
- Hunter, C. A., & Anderson, H. L. (2009). What is cooperativity? *Angewandte Chemie - International Edition*, 48(41), 7488–7499. <https://doi.org/10.1002/anie.200902490>
- Ingemarsdotter, C. K., Zeng, J., Long, Z., Lever, A. M. L., & Kenyon, J. C. (2018). An RNA-binding compound that stabilizes the HIV-1 gRNA packaging signal structure and specifically blocks HIV-1 RNA encapsidation. *Retrovirology*, 15, 25. <https://doi.org/10.1186/s12977-018-0407-4>
- Inobe, T., & Nukina, N. (2016). Rapamycin-induced oligomer formation system of FRB-FKBP fusion proteins. *Journal of Bioscience and Bioengineering*, 122(1), 40–46. <https://doi.org/10.1016/j.jbiosc.2015.12.004>
- Jantz, D., & Berg, J. M. (2010). Probing the DNA-binding affinity and specificity of designed zinc finger proteins. *Biophysical Journal*, 98(5), 852–860. <https://doi.org/10.1016/j.bpj.2009.11.021>
- Jekhmane, S., De Haas, R., Paulino Da Silva Filho, O., Van Asbeck, A. H., Favretto, M. E.,

- Hernandez Garcia, A., Brock, R., & De Vries, R. (2017). Virus-Like Particles of mRNA with Artificial Minimal Coat Proteins: Particle Formation, Stability, and Transfection Efficiency. *Nucleic Acid Therapeutics*, 27(3), 159–167. <https://doi.org/10.1089/nat.2016.0660>
- Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, 31(3), 233–239. <https://doi.org/10.1038/nbt.2508>
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA – Guided. 337(August), 816–822.
- Kanada, R., Terakawa, T., Kenzaki, H., & Takada, S. (2019). Nucleosome Crowding in Chromatin Slows the Diffusion but Can Promote Target Search of Proteins. *Biophysical Journal*, 116(12), 2285–2295. <https://doi.org/10.1016/j.bpj.2019.05.007>
- Kivenson, A., & Hagan, M. F. (2010). Mechanisms of capsid assembly around a polymer. *Biophysical Journal*, 99(2), 619–628. <https://doi.org/10.1016/j.bpj.2010.04.035>
- Kraft, D. J., Kegel, W. K., & Van Der Schoot, P. (2012). A kinetic zipper model and the assembly of tobacco mosaic virus. *Biophysical Journal*, 102(12), 2845–2855. <https://doi.org/10.1016/j.bpj.2012.05.007>
- Krejchi, M. T., Atkins, E. D. T., Waddon, A. J., Fournier, M. J., Mason, T. L., & Tirrell, D. A. (1994). Chemical sequence control of β -sheet assembly in macromolecular crystals of periodic polypeptides. *Science*, 265(5177), 1427–1432. <https://doi.org/10.1126/science.8073284>
- Kutluay, S. B., Zang, T., Blanco-Melo, D., Powell, C., Jannain, D., Errando, M., & Bieniasz, P. D. (2014). Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell*, 159(5), 1096–1109. <https://doi.org/10.1016/j.cell.2014.09.057>
- Leenay, R. T., & Beisel, C. L. (2017). Deciphering, Communicating, and Engineering the CRISPR PAM. In *Journal of Molecular Biology* (Vol. 429, Issue 2, pp. 177–191). Academic Press. <https://doi.org/10.1016/j.jmb.2016.11.024>
- Li, D., Li, X., Bai, J., Liu, Y., de Vries, R., & Li, Y. (2021). Rod-shaped polypeptide nanoparticles for siRNA delivery. *International Journal of Biological Macromolecules*, 166, 401–408. <https://doi.org/10.1016/j.ijbiomac.2020.10.198>
- Lin, S. yaung, & Riggs, A. D. (1975). The general affinity of lac repressor for E. coli DNA: Implications for gene regulation in procaryotes and eucaryotes. *Cell*, 4(2), 107–111. [https://doi.org/10.1016/0092-8674\(75\)90116-6](https://doi.org/10.1016/0092-8674(75)90116-6)
- Liu, Q., Wang, C., Zheng, Y., Zhao, Y., Wang, Y., Hao, J., Zhao, X., Yi, K., Shi, L., Kang, C., & Liu, Y. (2020). Virus-like nanoparticle as a co-delivery system to enhance efficacy of CRISPR/Cas9-based cancer immunotherapy. *Biomaterials*, 258, 120275. <https://doi.org/10.1016/j.biomaterials.2020.120275>
- Marchetti, M., Kamsma, D., Cazares Vargas, E., Hernandez García, A., Van Der Schoot, P., De Vries, R., Wuite, G. J. L., & Roos, W. H. (2019). Real-Time Assembly of Viruslike Nucleocapsids Elucidated at the Single-Particle Level. *Nano Letters*, 19(8), 5746–5753. <https://doi.org/10.1021/acs.nanolett.9b02376>
- Martin, M. E., & Rice, K. G. (2007). Peptide-guided gene delivery. *AAPS Journal*, 9(1). <https://doi.org/10.1208/aapsj0901003>

- Microbiology by numbers. (2011). *Nature Reviews Microbiology*, 9(9), 628.
<https://doi.org/10.1038/nrmicro2644>
- Morton, V. L., Dykeman, E. C., Stonehouse, N. J., Ashcroft, A. E., Twarock, R., & Stockley, P. G. (2010). The impact of viral RNA on assembly pathway selection. *Journal of Molecular Biology*, 401(2), 298–308.
<https://doi.org/10.1016/j.jmb.2010.05.059>
- Murugan, R. (2018). Theory of Site-Specific DNA-Protein Interactions in the Presence of Nucleosome Roadblocks. *Biophysical Journal*, 114(11), 2516–2529.
<https://doi.org/10.1016/j.bpj.2018.04.039>
- Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O. O., Gootenberg, J. S., Mori, H., Oura, S., Holmes, B., Tanaka, M., Seki, M., Hirano, H., Aburatani, H., Ishitani, R., Ikawa, M., Yachie, N., ... Nureki, O. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*, 361(6408), 1259–1262. <https://doi.org/10.1126/science.aas9129>
- Oosawa, F., & Asakura, S. (1975). *Thermodynamics of the Polymerization of Protein*. Academic Press.
- Patel, N., Wroblewski, E., Leonov, G., Phillips, S. E. V., Tuma, R., Twarock, R., & Stockley, P. G. (2017). Rewriting nature's assembly manual for a ssRNA virus. *Proceedings of the National Academy of Sciences*, 114(46).
- Pinnapireddy, S. R., El Assy, M. R., Schlote, P., & Bakowsky, U. (2019). Glycosylated artificial virus-like hybrid vectors for advanced gene delivery. *Polymers*, 11(2), 1–9.
<https://doi.org/10.3390/polym11020243>
- Powers, E. T., & Powers, D. L. (2006). The kinetics of nucleated polymerizations at high concentrations: Amyloid fibril formation near and above the “supercritical concentration.” *Biophysical Journal*, 91(1), 122–132.
<https://doi.org/10.1529/biophysj.105.073767>
- Prevelige, P. E., Thomas, D., & King, J. (1993). Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophysical Journal*, 64(3), 824–835. [https://doi.org/10.1016/S0006-3495\(93\)81443-7](https://doi.org/10.1016/S0006-3495(93)81443-7)
- Prinz, H. (2010). Hill coefficients, dose-response curves and allosteric mechanisms. *Journal of Chemical Biology*, 3(1), 37–44. <https://doi.org/10.1007/s12154-009-0029-3>
- Privalov, P. L., Dragan, A. I., & Crane-Robinson, C. (2011). Interpreting protein/DNA interactions: Distinguishing specific from non-specific and electrostatic from non-electrostatic components. In *Nucleic Acids Research* (Vol. 39, Issue 7, pp. 2483–2491). Oxford Academic. <https://doi.org/10.1093/nar/gkq984>
- Punter, M. T. J. J. M., Hernandez-Garcia, A., Kraft, D. J., De Vries, R., & Van Der Schoot, P. (2016). Self-Assembly Dynamics of Linear Virus-Like Particles: Theory and Experiment. *Journal of Physical Chemistry B*, 120(26), 6286–6297.
<https://doi.org/10.1021/acs.jpccb.6b02680>
- Razzokov, J., Naderi, S., & Van Der Schoot, P. (2018). Nanoscale insight into silk-like protein self-assembly: Effect of design and number of repeat units. *Physical Biology*, 15(6), aadb5e. <https://doi.org/10.1088/1478-3975/aadb5e>
- Ren, H., Zhu, S., & Zheng, G. (2019). Nanoreactor design based on self-assembling

- protein nanocages. *International Journal of Molecular Sciences*, 20(3), 25–27.
<https://doi.org/10.3390/ijms20030592>
- Shibata, M., Nishimasu, H., Kodera, N., Hirano, S., Ando, T., Uchihashi, T., & Nureki, O. (2017). Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. *Nature Communications*, 8(1), 1430.
<https://doi.org/10.1038/s41467-017-01466-8>
- Singh, S., & Zlotnick, A. (2003). Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly. *Journal of Biological Chemistry*, 278(20), 18249–18255. <https://doi.org/10.1074/jbc.M211408200>
- Smeenk, J. M., Otten, M. B. J., Thies, J., Tirrell, D. A., Stunnenberg, H. G., & Van Hest, J. C. M. (2005). Controlled assembly of macromolecular β -sheet fibrils. *Angewandte Chemie - International Edition*, 44(13), 1968–1971.
<https://doi.org/10.1002/anie.200462415>
- Soniat, M. M., Myler, L. R., Schaub, J. M., Kim, Y., Gallardo, I. F., & Finkelstein, I. J. (2017). Next-Generation DNA Curtains for Single-Molecule Studies of Homologous Recombination. In *Methods in Enzymology* (1st ed., Vol. 592). Elsevier Inc.
<https://doi.org/10.1016/bs.mie.2017.03.011>
- Stewart, H., Bingham, R. J., White, S. J., Dykeman, E. C., Zothner, C., Tuplin, A. K., Stockley, P. G., Twarock, R., & Harris, M. (2016). Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome packaging. *Scientific Reports*, 6(February), 1–12.
<https://doi.org/10.1038/srep22952>
- Stigler, J., Çamdere, G., Koshland, D. E., & Greene, E. C. (2016). Single-Molecule Imaging Reveals a Collapsed Conformational State for DNA-Bound Cohesin. *Cell Reports*, 15(5), 988–998. <https://doi.org/10.1016/j.celrep.2016.04.003>
- Stockley, P. G., Ranson, N. A., & Twarock, R. (2013). A new paradigm for the roles of the genome in ssRNA viruses. *Future Virology*, 8(6), 531–543.
<https://doi.org/10.2217/fvl.12.84>
- Strohkendl, I., Saifuddin, F. A., Rybarski, J. R., Finkelstein, I. J., & Russell, R. (2018). Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a. *Molecular Cell*, 71(5), 816–824.e3. <https://doi.org/10.1016/j.molcel.2018.06.043>
- Subramanian, R. H., Smith, S. J., Alberstein, R. G., Bailey, J. B., Zhang, L., Cardone, G., Suominen, L., Chami, M., Stahlberg, H., Baker, T. S., & Tezcan, F. A. (2018). Self-Assembly of a Designed Nucleoprotein Architecture through Multimodal Interactions. *ACS Central Science*, 4(11), 1578–1586. <https://doi.org/10.1021/acscentsci.8b00745>
- Sykora, S., Cumbo, A., Belliot, G., Pothier, P., Arnal, C., Dudal, Y., Corvini, P. F. X., & Shahgaldian, P. (2015). Virus-like particles as virus substitutes to design artificial virus-recognition nanomaterials. *Chemical Communications*, 51(12), 2256–2258.
<https://doi.org/10.1039/c4cc08843c>
- Terasaka, N., Azuma, Y., & Hilvert, D. (2018). Laboratory evolution of virus-like nucleocapsids from nonviral protein cages. *Proceedings of the National Academy of Sciences of the United States of America*, 115(21), 5432–5437.
<https://doi.org/10.1073/pnas.1800527115>

- Tetter, S., Terasaka, N., Steinauer, A., Bingham, R. J., Clark, S., Scott, A. J. P., Patel, N., Leibundgut, M., Wroblewski, E., Ban, N., Stockley, P. G., Twarock, R., & Hilvert, D. (2021). Evolution of a virus-like architecture and packaging mechanism in a repurposed bacterial protein. *Science*, *372*(6547), 1220–1224. <https://doi.org/10.1126/SCIENCE.ABG2822>
- Turner, D. R., Joyce, L. E., & Butler, P. J. G. (1988). The tobacco mosaic virus assembly origin RNA. Functional characteristics defined by directed mutagenesis. *Journal of Molecular Biology*, *203*(3), 531–547. [https://doi.org/10.1016/0022-2836\(88\)90190-8](https://doi.org/10.1016/0022-2836(88)90190-8)
- Twarock, R., Bingham, R. J., Dykeman, E. C., & Stockley, P. G. (2018). A modelling paradigm for RNA virus assembly. In *Current Opinion in Virology* (Vol. 31, pp. 74–81). Elsevier B.V. <https://doi.org/10.1016/j.coviro.2018.07.003>
- Umunnakwe, C. N., Duchon, A., Nikolaitchik, O. A., Rahman, S. A., Liu, Y., Chen, J., Tai, S., Pathak, V. K., & Hu, W. S. (2021). Specific Guanosines in the HIV-2 Leader RNA are Essential for Efficient Viral Genome Packaging. *Journal of Molecular Biology*, *433*(2), 166718. <https://doi.org/10.1016/j.jmb.2020.11.017>
- Valbuena, A., Maity, S., Mateu, M. G., & Roos, W. H. (2020). Visualization of Single Molecules Building a Viral Capsid Protein Lattice through Stochastic Pathways. *ACS Nano*, *14*. <https://doi.org/10.1021/acsnano.0c03207>
- Valbuena, A., & Mateu, M. G. (2017). Kinetics of Surface-Driven Self-Assembly and Fatigue-Induced Disassembly of a Virus-Based Nanocoating. *Biophysical Journal*, *112*(4), 663–673. <https://doi.org/10.1016/j.bpj.2016.11.3209>
- Van der heijden, T., Seidel, R., Modesti, M., Kanaar, R., Wyman, C., & Dekker, C. (2007). Real-time assembly and disassembly of human RAD51 filaments on individual DNA molecules. *Nucleic Acids Research*, *35*(17), 5646–5657. <https://doi.org/10.1093/nar/gkm629>
- Van Galen, M., Higler, R., & Sprakel, J. (2019). Allosteric pathway selection in templated assembly. *Science Advances*, *5*(10), eaaw3353. <https://doi.org/10.1126/sciadv.aaw3353>
- Vargas, E. C., Stuart, M. A. C., de Vries, R., & Hernandez-Garcia, A. (2019). Template-Free Self-Assembly of Artificial De Novo Viral Coat Proteins into Nanorods: Effects of Sequence, Concentration, and Temperature. *Chemistry - A European Journal*, *14*–17. <https://doi.org/10.1002/chem.201901486>
- Walton, R. T., Christie, K. A., Whittaker, M. N., & Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*, *368*(6488), 290–296. <https://doi.org/10.1126/science.aba8853>
- Wasik, B. R., & Turner, P. E. (2013). On the biological success of viruses. *Annual Review of Microbiology*, *67*, 519–541. <https://doi.org/10.1146/annurev-micro-090110-102833>
- Werten, M. W. T., Wisselink, W. H., Jansen-van Den Bosch, T. J., De Bruin, E. C., & De Wolf, F. A. (2001). Secreted production of a custom-designed, highly hydrophilic gelatin in *Pichia pastoris*. *Protein Engineering*, *14*(6), 447–454. <https://doi.org/10.1093/protein/14.6.447>
- Willems, L., Van Westerveld, L., Roberts, S., Weitzhandler, I., Calcines Cruz, C., Hernandez-Garcia, A., Chilkoti, A., Mastrobattista, E., Van Der Oost, J., & De Vries,

- R. (2019). Nature of amorphous hydrophilic block affects self-assembly of an artificial viral coat polypeptide. *Biomacromolecules*, *20*(10), 3641–3647. <https://doi.org/10.1021/acs.biomac.9b00512>
- Wilson, K. P., Yamashita, M. M., Sintchak, M. D., Rotstein, S. H., Murcko, M. A., Boger, J., Thomson, J. A., Fitzgibbon, M. J., Black, J. R., & Navia, M. A. (1995). Comparative X-ray structures of the major binding protein for the immunosuppressant FK506 (tacrolimus) in unliganded form and in complex with FK506 and rapamycin. *Acta Crystallographica Section D Biological Crystallography*, *51*(4), 511–521. <https://doi.org/10.1107/s0907444994014514>
- Xu, R., Shi, M., Li, J., Song, P., & Li, N. (2020). Construction of SARS-CoV-2 Virus-Like Particles by Mammalian Expression System. *Frontiers in Bioengineering and Biotechnology*, *8*. <https://doi.org/10.3389/fbioe.2020.00862>
- Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I. M., Li, Y., Fedorova, I., Nakane, T., Makarova, K. S., Koonin, E. V., Ishitani, R., Zhang, F., & Nureki, O. (2016). Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell*, *165*(4), 949–962. <https://doi.org/10.1016/j.cell.2016.04.003>
- Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., Van Der Oost, J., Regev, A., Koonin, E. V., & Zhang, F. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, *163*(3), 759–771. <https://doi.org/10.1016/j.cell.2015.09.038>
- Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., Degennaro, E. M., Winblad, N., Choudhury, S. R., Abudayyeh, O. O., Gootenberg, J. S., Wu, W. Y., Scott, D. A., Severinov, K., Van Der Oost, J., & Zhang, F. (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nature Biotechnology*, *35*(1), 31–34. <https://doi.org/10.1038/nbt.3737>
- Zhang, J., & You, Y. (2020). CRISPR-Cas13a system: a novel approach to precision oncology. *Cancer Biology and Medicine*, *17*(1), 6–8. <https://doi.org/10.20892/j.issn.2095-3941.2019.0325>
- Zhao, B., Cohen Stuart, M. A., & Hall, C. K. (2016). Dock 'n roll: Folding of a silk-inspired polypeptide into an amyloid-like beta solenoid. *Soft Matter*, *12*(16), 3721–3729. <https://doi.org/10.1039/c6sm00169f>
- Zhou, K., Eiben, S., & Wang, Q. (2016). Coassembly of Tobacco Mosaic Virus Coat Proteins into Nanotubes with Uniform Length and Improved Physical Stability. *ACS Applied Materials and Interfaces*, *8*(21), 13192–13196. <https://doi.org/10.1021/acsami.6b04321>
- Zhou, W., Liu, L., Huang, J., Cai, Y., Cohen Stuart, M. A., de Vries, R., & Wang, J. (2021). Supramolecular virus-like particles by co-assembly of triblock polypeptide and PAMAM dendrimers. *Soft Matter*. <https://doi.org/10.1039/D1SM00290B>

Anexos

Tabla S1: Secuencia de la región variable de los crRNAs usados para posicionar dCas12a y dCas12a-S₁₀ en el DNA. La secuencia completa del crRNA consiste en una región constante 5' (UAAUUUCUACUCUUGUAGAU, 20 nt) seguida de la región variable 3' (24 nt) que se muestra en la tabla.

Decoración del λDNA con dCas12a o dCas12a-S₁₀	
5 sitios diana	AUGAUGUUCUGCUGGAUAUGCACU CCUGACACCGGACGGAAAGCUGAC AAUGUCGGCUAAUCGAUUUGGCCA GCUAGCAAUUAUGUGCAUCGAUU AUGAACGCAAUAUUCACAAGCAAU
10 sitios diana	CGUGAGAGCUAUCUUUCACACG AUGAUGUUCUGCUGGAUAUGCACU CGUAUGUCGCCGGAAGACUGGCUG CCUGACACCGGACGGAAAGCUGAC UGAU AUGCCGCAGAAACGUUGUAU AAUGUCGGCUAAUCGAUUUGGCCA AUGUUCAUCGUUCCUUAAGACGC GCUAGCAAUUAUGUGCAUCGAUU CCGGACAGGAGCGUAAUGUGGCAG AUGAACGCAAUAUUCACAAGCAAU
Decoración de la hebra diana de 28 kbp con dCas12a-S₁₀	
5 sitios diana	UGGCCAAAUCGAUUAGCCGACAUU GCGUCUUUAAGGAACGAUGAACAU AAUCGAUGCACAUAUUAUUGCUAGC CUGCCACAUUACGCUCCUGUCCGG AUUGCUUGUGAAUAUUGCGUUCAU
10 sitios diana	UGGCCAAAUCGAUUAGCCGACAUU GUCAGAGGCUUGUGUUUGUGUCCU GCGUCUUUAAGGAACGAUGAACAU ACUGCGCAUCGCUGGCAUCACCUU AAUCGAUGCACAUAUUAUUGCUAGC GCUUA AUGACAUUCCUUUCCCGAU CUGCCACAUUACGCUCCUGUCCGG CAGGAACGCAACCGCAGCUUAGAC AUUGCUUGUGAAUAUUGCGUUCAU AUACCGGAAGCAGAACCGGAUCAC
25 sitios diana	UUGCUUCUCUUGACCGUAGGACUU CAGUAUUAUGUAGUCUGUUUUUUA UAAACUCCUUGCAAUGUAUGUCGU UGGCCUCGAAACCACCGAGCCGGA CAUCAUUCAGUCGAACUCACACA UCACCGCAGAUGGUUAUCUGUAUG UCUGGCGAUUGAAGGGCUAAAUUC CGGGGUGGAUCUAUGAAAAACAUC AGAAGGAAGAUAUCCUCGCAUGGU

CCAACCAAUUGUAUAUCGAUACCG
ACCCUCAGAGAGAGGCUGAUCACU
ACUUAUAGUAUUGGUUGCGUAAC
UGAUUUGCCGCAGAAACGUUGUAU
AAUGUCGGCUAAUCGAUUUGGCCA
AUGUUCAUCGUUCCUUAAGACGC
GCUAGCAAUUAUGUGCAUCGAUU
CCGGACAGGAGCGUAAUGUGGCAG
AUGAACGCAAUAUUCACAAGCAAU
AGGCCACCGCAUCUCGUGCUGAAG
UUGAAGCAAUUCUGAAACCUAUUA
AGGACACAAACACAAGCCUCUGAC
AAGGUGAUGCCAGCGAUGCGCAGU
AUCGGGAAAGGAAUGUCAUUAAGC
GUCUAAGCUGCGGUUGCGUUCUG
GUGAUCCGGUUCUGCUUCCGGUUAU

Ensayo de movilidad electroforética y microscopía de fuerza atómica

DNA lineal
(10 sitios diana)

UAGAGCAUAAGCAGCGCAACACCC
AUGAUGAUUUGAACAGGAAGGCU
UCUCUGCGAGCAUAAUGCCUGCGU
GCCUCCACGUCUCACCGAGCGUG
UUGAUGGCCUCAUCCACACGCAGC
UAACCGCUUCACACUGACGCCGGA
UUGUUGGUUGCUGCACCAUCCUCU
UGUAUGAAAACGCCACCAUUCUCC
CGGCUCAGUCAUCGCCAAGCUGG
CGGACACAGUUCGGGAUGGUCAGC

DNA plasmídico (4 repeticiones de 1 sitio
diana)

CCUUGUUCACCUUGGUUGACCAGGG

Tabla S2: Parámetros del ajuste de curvas a la suma de múltiples Gaussianas. La distribución sobre el DNA de dCas12a, dCas12a-S₁₀ y los clústers de C-S₁₀-B se ajustó a la suma de cinco o diez Gaussianas en las **Figuras 21, 23 y 28**. Para cada Gaussianas se muestra la media \pm desviación estándar. La diferencia con respecto a la posición esperada se muestra en paréntesis. Las unidades son kbp.

Posición esperada	dCas12a, 5 sitios diana (Figura 20c)	dCas12a, 10 sitios diana (Figura 20c)	C-S₁₀-B clústers, 5 sitios diana (Figura 22b)	dCas12a-S₁₀, 5 sitios diana (Figura 27c)
6.2	-	5.9 \pm 3.6 (0.4)	-	-
10.3	9.7 \pm 2.1 (0.5)	10.5 \pm 1.4 (0.3)	9.0 \pm 2.8 (1.3)	8.0 \pm 2.1 (1.3)
14.3	-	12.9 \pm 1.3 (1.4)	-	-
18.4	18.0 \pm 2.2 (0.4)	18.8 \pm 6.0 (0.4)	16.7 \pm 2.0 (1.7)	17.1 \pm 2.0 (1.3)
22.6	-	23.5 \pm 1.3 (0.9)	-	-
26.6	26.0 \pm 1.5 (0.6)	27.5 \pm 5.3 (0.9)	25.8 \pm 4.5 (0.9)	24.7 \pm 3.3 (2.0)
30.7	-	32.2 \pm 1.8 (1.5)	-	-
34.7	34.5 \pm 2.1 (0.2)	35.9 \pm 2.3 (1.2)	33.8 \pm 1.6 (0.9)	33.1 \pm 1.1 (1.6)
38.7	-	38.8 \pm 1.1 (0.1)	-	-
42.8	42.7 \pm 3.1 (0.1)	41.6 \pm 4.7 (1.2)	40.7 \pm 2.4 (2.1)	39.8 \pm 2.8 (3.1)
R²	0.6978	0.86921	0.5582	0.6511

Producción científica

Se incluyen a continuación tres artículos en los que participé durante mi doctorado.

1. El primero (Calcines-Cruz et al., 2021) incluye todos los resultados presentados en la tesis y da solución a mi proyecto de doctorado.
2. El segundo (Lione et al., 2019) describe la sustitución del dominio C de C-S₁₀-B por un polipéptido tipo elastina que se expresa en *E. coli*. Mi contribución fue demostrar que el dominio tipo colágeno C se expresa pobremente en *E. coli*, de ahí que su modificación es crucial para cambiar el modo y huésped de expresión.
3. El tercero (Torres-García et al., en preparación) no se incluye más que la lista de autores a petición del autor principal. Este trabajo presenta un novedoso algoritmo de análisis de imágenes de fluorescencia para obtener superresolución. Mi contribución fue demostrar la aplicación de dicho algoritmo a las imágenes obtenidas en las cortinas de DNA, y usar moléculas de DNA decoradas con dCas12a (visualizadas por AFM) como nano-reglas.

CRISPR-Guided Programmable Self-Assembly of Artificial Virus-Like Nucleocapsids

Carlos Calcines-Cruz, Ilya J. Finkelstein,* and Armando Hernandez-Garcia*



Cite This: <https://doi.org/10.1021/acs.nanolett.0c04640>



Read Online

ACCESS |



Metrics & More



Article Recommendations

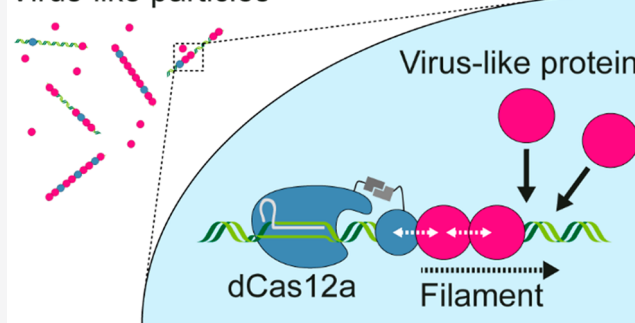


Supporting Information

ABSTRACT: Designer virus-inspired proteins drive the manufacturing of more effective, safer gene-delivery systems and simpler models to study viral assembly. However, self-assembly of engineered viromimetic proteins on specific nucleic acid templates, a distinctive viral property, has proved difficult. Inspired by viral packaging signals, we harness the programmability of CRISPR-Cas12a to direct the nucleation and growth of a self-assembling synthetic polypeptide into virus-like particles (VLP) on specific DNA molecules. Positioning up to ten nuclease-dead Cas12a (dCas12a) proteins along a 48.5 kbp DNA template triggers particle growth and full DNA encapsidation at limiting polypeptide concentrations. Particle growth rate is further increased when dCas12a is dimerized with a polymerization silk-like domain. Such improved self-assembly efficiency allows for discrimination between cognate versus noncognate DNA templates by the synthetic polypeptide. CRISPR-guided VLPs will help to develop programmable bioinspired nanomaterials with applications in biotechnology as well as viromimetic scaffolds to improve our understanding of viral self-assembly.

KEYWORDS: virus-like particles, Cas12a, assembly kinetics, DNA curtain

Virus-like particles



Virus-like particles (VLPs) mimic the capability of some viruses to encapsulate and protect genetic material from degradation by nucleases. We have previously described VLPs formed by the self-assembly of a triblock polypeptide (C-S₁₀-B) that functionally mimics the tobacco mosaic virus coat protein¹ (Figure 1a, Figure S1). C-S₁₀-B fuses three independent blocks: (1) “C” (~400 aa), a random coil collagen-like domain that consists mostly of glycine, proline, and uncharged polar amino acids;² (2) “S₁₀”, a silk-inspired polymerization domain with the sequence [(AG)₃QG]₁₀ that is responsible for C-S₁₀-B self-assembly into rodlike structures;^{3–5} and (3) “B”, a cationic dodecalysine stretch that interacts with nucleic acids and other polyanions.^{6,7} C-S₁₀-B nucleates (without sequence specificity) on double-stranded DNA (dsDNA), albeit with a preference for free DNA ends.⁸ After rate-limiting nucleation, C-S₁₀-B filaments grow rapidly through elongation,^{1,9} similar to the assembly of the tobacco mosaic virus coat protein on genomic ssRNA.

Viral coat proteins preferentially encapsulate their own genomes. They achieve such specificity by encoding one or more packaging signals along the viral DNA or RNA. These sequences bind capsid proteins with high affinity^{10–17} and decrease the energy barrier for nucleation, thereby promoting encapsidation of the viral genome among a vast excess of cellular nucleic acids.^{18–21} We reasoned that designer VLPs can also leverage a similar packaging signal to enhance nucleation at specific DNA sites. We chose the catalytically dead CRISPR-

Cas12a (dCas12a) as a programmable nucleation signal because it binds dsDNA with 50 fM affinity, has a higher DNA binding specificity than *S. pyogenes* Cas9,²² and can be directed to multiple sites along the DNA via pooled CRISPR RNAs (crRNAs).²³

Self-assembly kinetics were monitored in real-time using the single-molecule DNA curtain assay (Figure 1b).^{24,25} For this, arrays of DNA molecules (48.5 kbp, derived from λ-phage) are affixed to a lipid bilayer via a biotin–streptavidin linkage in a microfluidic flowcell. Microfabricated chromium barriers are used to organize thousands of DNA strands for high-throughput data collection and analysis. The surface-immobilized DNA is extended for fluorescent imaging via the application of mild buffer flow. DNA length was monitored in the experiment because time-dependent DNA contraction is a direct readout of VLP filamentation (Figure 1c).¹

First, we identified the minimal concentrations for efficient encapsidation of individual dsDNA substrates by C-S₁₀-B alone. At 100–300 nM, C-S₁₀-B monotonically condensed

Received: November 23, 2020
Revised: March 11, 2021

Received: November 23, 2020

Revised: March 11, 2021

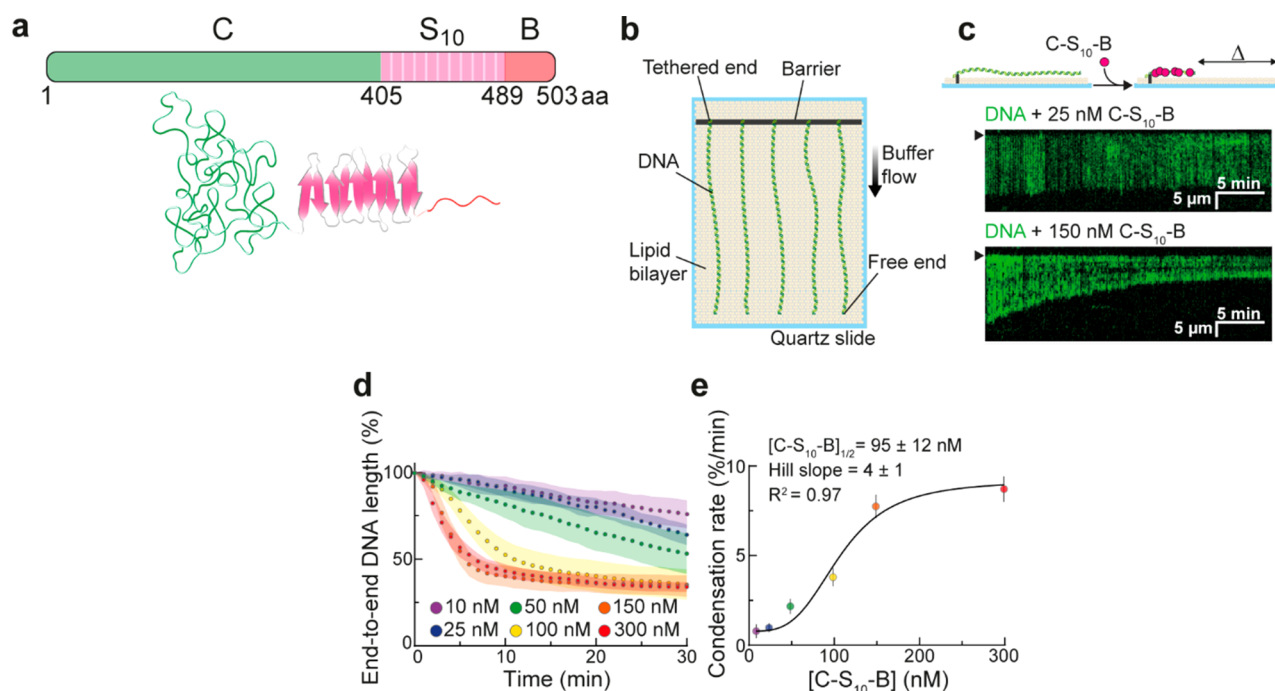


Figure 1. Synthetic polypeptide C-S₁₀-B packages DNA into nucleocapsids via molecular self-assembly. (a) Schematic depiction and modular design of C-S₁₀-B. (b) Illustration of the DNA curtain assay. Individual DNA molecules are captured on the surface of a lipid bilayer-passivated flowcell and extended by gentle buffer flow. (c) Representative kymographs showing condensation of individual DNA strands (green) at different rates determined by C-S₁₀-B concentration. The DNA is stained with the intercalating dye YOYO-1. Barrier position is indicated with black arrows on the kymographs. (d) DNA condensation profiles at the indicated C-S₁₀-B concentrations. The circles and shaded areas represent the mean and standard deviation for 25 molecules per condition, respectively. (e) DNA condensation rate in the first 3 min of assembly at different concentrations of C-S₁₀-B. The data were fit to the Hill equation (solid line). Circles and error bars are the mean ±95% confidence intervals for 25 DNA molecules per condition.

individual DNA molecules until complete assembly of linear particles with a final length of $34 \pm 3\%$ of the initial length ($N = 25$ DNA molecules at each concentration) (Figure 1d). Total DNA contraction to about one-third of the DNA original length has also been observed by previous AFM studies.¹ The DNA encapsidation rate—measured as t_{half} , the time required to achieve half of total packaging—decreased with higher C-S₁₀-B concentrations up to 50 nM and plateaued at $[C-S_{10}-B] \geq 100$ nM (Figure S2a). In addition, the DNA did not completely contract at $[C-S_{10}-B] \leq 50$ nM as compared to 100–300 nM. Together, these observations indicate incomplete C-S₁₀-B coating of the DNA at ≤ 50 nM C-S₁₀-B concentrations. Initial condensation rate increased with C-S₁₀-B concentration in a sigmoid-shaped curve distinctive of nucleated self-assemblies (Figure 1e). DNA coating was significantly impaired at 25 nM C-S₁₀-B compared with 150 nM (Figure S2b, c). These results are consistent with a dynamic equilibrium between C-S₁₀-B nucleation-filamentation and dissociation from DNA, akin to RAD51 and other dynamic filaments.²⁶

Next, we evaluated whether positioning dCas12a on the DNA can improve encapsidation at low C-S₁₀-B concentrations. The DNA was uniformly decorated with five or ten dCas12a-crRNA ribonucleoproteins (RNPs) (Figure 2a). We confirmed site-specific target binding by imaging fluorescent RNPs along the DNA molecule (Figure S3a, b). We observed 1–4 target-bound RNPs on DNA substrates harboring five binding sites (1.2 ± 0.5 RNP per DNA, $N = 456$ DNA molecules) and 1–7 RNPs on DNA substrates with ten binding sites (2.6 ± 1.1 RNP per DNA, $N = 185$ DNA molecules) (Figure S3c). Because not all RNPs are decorated with fluorescent QDs, our results are a lower bound on the true target site occupancy. To monitor how

encapsidation varies with RNP density, we injected 25 nM C-S₁₀-B into flowcells where the DNA was predecorated with either five or ten RNPs (Figure 2b, c). DNA contraction rate increased 2- and 3-fold for DNA with five ($t_{\text{half}} = 15 \pm 7$ min; $N = 25$ DNA molecules) and ten ($t_{\text{half}} = 10 \pm 3$ min; $N = 25$ DNA molecules) RNPs with respect to nondecorated DNA ($t_{\text{half}} = 31 \pm 6$ min; $N = 25$ DNA molecules) (Figure 2d). We conclude that dCas12a RNPs can be installed at specific sites along the DNA template to accelerate encapsidation by C-S₁₀-B.

We reasoned that dCas12a organizes large C-S₁₀-B clusters that further polymerize into filaments. Two-color imaging confirmed that fluorescent C-S₁₀-B colocalizes with the RNPs at early stages of encapsidation (Figure 2e). Positioning five RNPs on the DNA increased the number of fluorescent C-S₁₀-B puncta per DNA strand from 1–3 (1.3 ± 0.6 , $N = 166$ DNA molecules) to 2–7 (3.3 ± 1.0 , $N = 97$ DNA molecules) for DNA and dCas12a-decorated DNA, respectively (Figure 2f). C-S₁₀-B protomers can freely diffuse on the DNA.²⁷ In the presence of buffer flow, these molecules slide and assemble into large clusters at the free DNA ends. In contrast, C-S₁₀-B accumulated at dCas12a sites on the decorated DNA (Figure S3d). C-S₁₀-B binding was also more rapid on predecorated DNA substrates ($t_{\text{half}} = 15 \pm 6$ min; $N = 10$ DNA molecules) than on nondecorated DNA ($t_{\text{half}} = 45 \pm 15$ min; $N = 10$ DNA molecules) (Figure 2g, h). C-S₁₀-B filamentation was unable to displace dCas12a, which forms a stable RNA:DNA loop (R-loop) with the DNA substrate (Figure S3e). Because dCas12a may act as a roadblock for C-S₁₀-B linear diffusion on the DNA, we also tested whether other DNA-binding proteins will accumulate C-S₁₀-B clusters. Notably, DNA decoration with nucleosomes also accelerated DNA encapsidation and particle

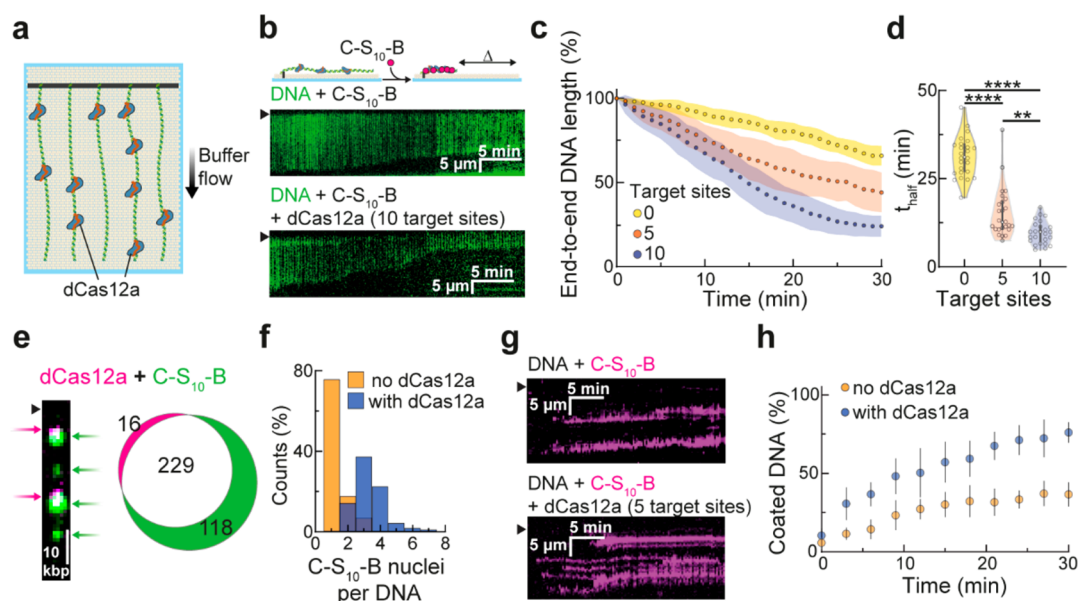


Figure 2. Seeding assembly with dCas12a improves DNA encapsidation by C-S₁₀-B. (a) Illustration of dCas12a-decorated DNA substrates. dCas12a is incubated with pools of crRNAs and directed to five or ten sites uniformly distributed along the DNA prior to C-S₁₀-B injection. (b) Representative kymographs show faster encapsidation at 25 nM C-S₁₀-B after the DNA (green) is decorated with dCas12a. Barrier position is indicated with arrows on the kymographs. (c) Condensation profiles at 25 nM C-S₁₀-B for DNA, and DNA decorated with dCas12a targeting five or ten sequences along the template. The circles and shaded areas represent the mean and standard deviation for 25 molecules per condition, respectively. (d) Violin plots showing the time (t_{half}) required to reach half of maximum condensation for each DNA molecule analyzed in panel c. We extrapolated t_{half} (after curve fitting to the Hill equation) for molecules that did not reach 66.5% encapsidation during the experiment (30 min). (***) and (****) indicate $p < 0.01$ and $p < 0.0001$, respectively. (e) Double labeling experiments show that 93% of dCas12a (magenta) colocalized with 66% of fluorescent C-S₁₀-B clusters (green). Barrier position is indicated with a black arrow. (f) Positioning five dCas12a on the DNA increases the number of fluorescent C-S₁₀-B clusters relative to undecorated DNA ($N = 484$ and 246 clusters, respectively). (g) Kymographs showing binding of C-S₁₀-B (magenta) on undecorated (top) or dCas12a-decorated (bottom) DNA (unlabeled) at 25 nM C-S₁₀-B. Barrier position is indicated with black arrows on the kymographs. (h) Extent of DNA that is coated by fluorescent C-S₁₀-B. Points and error bars indicate the mean and standard deviation, respectively ($N = 10$ DNA molecules per condition).

growth (Figure S4). We conclude that dCas12a stalls C-S₁₀-B sliding on DNA, which enhances C-S₁₀-B collisions in its vicinity and triggers particle nucleation. In support of this model, Marchetti et al.²⁷ observed that immobile C-S₁₀-B clusters, and not their sliding counterparts, initiate filament growth.

To further promote and accelerate DNA packaging at limiting C-S₁₀-B concentrations, we physically coupled dCas12a to the diblock polypeptide C-S₁₀ via the rapamycin-inducible dimerization of FKBP and FRB domains (Figure 3a). We cloned and purified a dCas12a_(FKBP) C-terminal fusion and verified that this construct retains target-specific DNA binding (Figure S5). The FRB domain was fused to a truncated C-S₁₀ polypeptide that lacks the “B” DNA-binding module. Both proteins, dCas12a_(FKBP) and C-S₁₀(FRB), were incubated with rapamycin and the dimerized complex was injected into the flowcell prior to incubation with 10 nM C-S₁₀-B (Figure 3b, c). Positioning five dimerized dCas12a-S₁₀ complexes along the DNA substrate accelerated encapsidation by C-S₁₀-B 2-fold relative to dCas12a_(FKBP) alone ($t_{\text{half}} = 9 \pm 3$ min; $N = 25$ DNA molecules vs 18 ± 6 min; $N = 25$ DNA molecules), and 5-fold relative to the undecorated DNA ($t_{\text{half}} = 46 \pm 13$ min; $N = 25$ DNA molecules) (Figure 3d, e). Positioning dCas12a-S₁₀ on a 2.5 kbp linear dsDNA or a 9.5 kbp plasmid also improved encapsidation at limiting C-S₁₀-B concentrations in ensemble electrophoretic mobility shift assays (EMSA) (Figure S6). These results indicate that initiating nucleation via interspersed dCas12a RNPs fused to the self-assembly domain S₁₀ accelerates DNA packaging at subsaturating C-S₁₀-B concentrations. Importantly, decoration with dCas12a or dCas12a-S₁₀ did not

disrupt the morphology of the artificial nucleocapsids, and full particles of 1/3 of the template DNA length were observed by AFM (Figure S7).

Finally, we assessed whether targeting dCas12a-S₁₀ to a specific DNA template can trigger VLP encapsidation in the presence of other (noncognate) DNA molecules. For this assay, we immobilized an equimolar mixture of two DNA templates on the flowcell surface; one was 20 kbp and the second was 28.5 kbp (Figure 4a). dCas12a-S₁₀ was directed to 5, 10, or 25 sites on the 28.5 kbp template (Figure 4b). As seen in Figure 4c, d and Figure S8, dCas12a-S₁₀ selectively accelerated the packaging of the target DNA template and the encapsidation rate increased 3-fold from zero target sites ($t_{\text{half}} = 62 \pm 26$ min; $N = 25$ DNA molecules) to a maximum of 25 target sites ($t_{\text{half}} = 19 \pm 6$ min; $N = 25$ DNA molecules).

Taken together, our data show that a target DNA-bound dCas12a (or any strongly DNA-bound roadblock protein) can serve as a viral-like packaging signal—especially if the roadblock protein contains a self-assembly or polymerization domain. RNA-guided CRISPR-Cas proteins are especially attractive as artificial packaging signals because they can be targeted to any DNA sequence that is proximal to a protospacer adjacent motif (PAM). For AsdCas12a, the PAM consensus sequence is TTTV; this PAM appears on average every 32 bp on 48.5 kbp long λ -phage DNA. Other CRISPR-Cas enzymes, including Cas9, can also serve as nucleation signals. Relaxed PAM variants, both for Cas9 and Cas12a, will further increase the targeting possibilities for rapid and sequence specific VLP assembly.^{28–31}

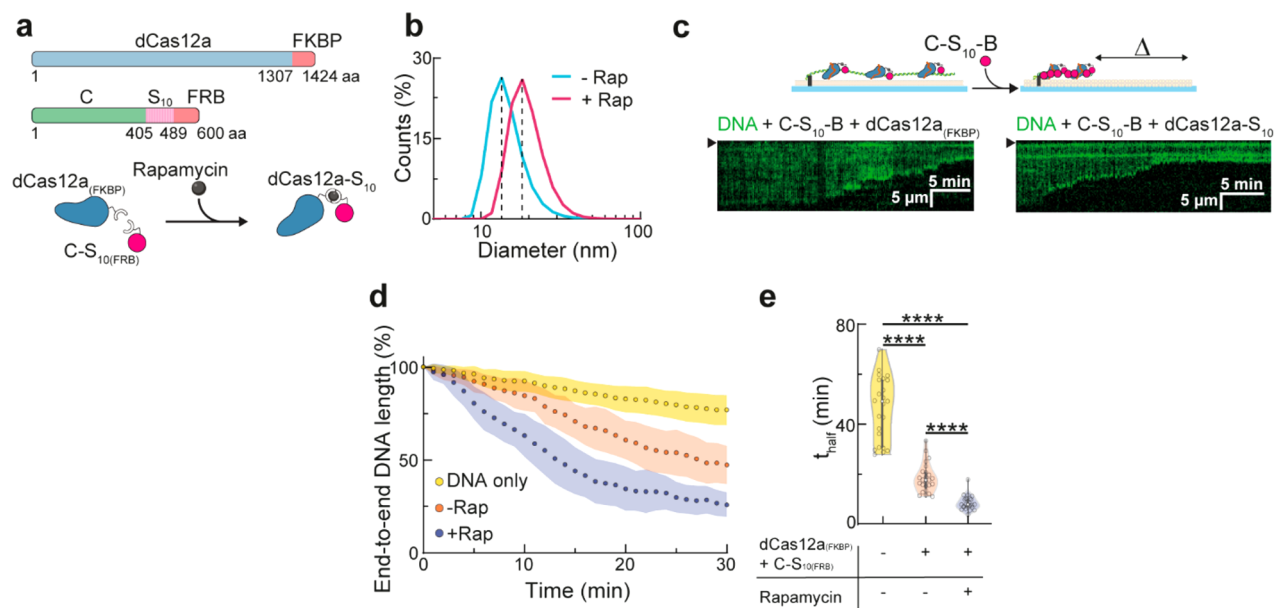


Figure 3. Coupling of dCas12a to the polymerization domain C-S₁₀ improves C-S₁₀-B self-assembly on DNA. (a) Schematic depiction of the modular design of dCas12a_(FKBP) and C-S₁₀_(FRB) and their dimerization via rapamycin to form the dCas12a-S₁₀ complex. (b) Dynamic light scattering experiment showing rapamycin (Rap) induced dimerization of dCas12a_(FKBP) and C-S₁₀_(FRB). The shift in population size occurred immediately after rapamycin addition. (c) Representative kymographs showing that decorating DNA with dCas12a-S₁₀ accelerates DNA packaging relative to decoration with dCas12a_(FKBP) alone at 10 nM C-S₁₀-B. Barrier position is indicated with arrows on the kymographs. (d) Condensation profiles at 10 nM C-S₁₀-B for undecorated DNA, and DNA decorated with five dCas12a_(FKBP) or five dCas12a-S₁₀. The circles and shaded areas represent the mean and standard deviation for 25 molecules per condition, respectively. (e) Violin plots showing the time (t_{half}) required to reach half of maximum condensation for each DNA strand analyzed in panel d. Extrapolation (after curve fitting to the Hill equation) was used to estimate t_{half} for molecules that did not reach 66.5% encapsidation during the experiment (30 min). (****) indicates $p < 0.0001$.

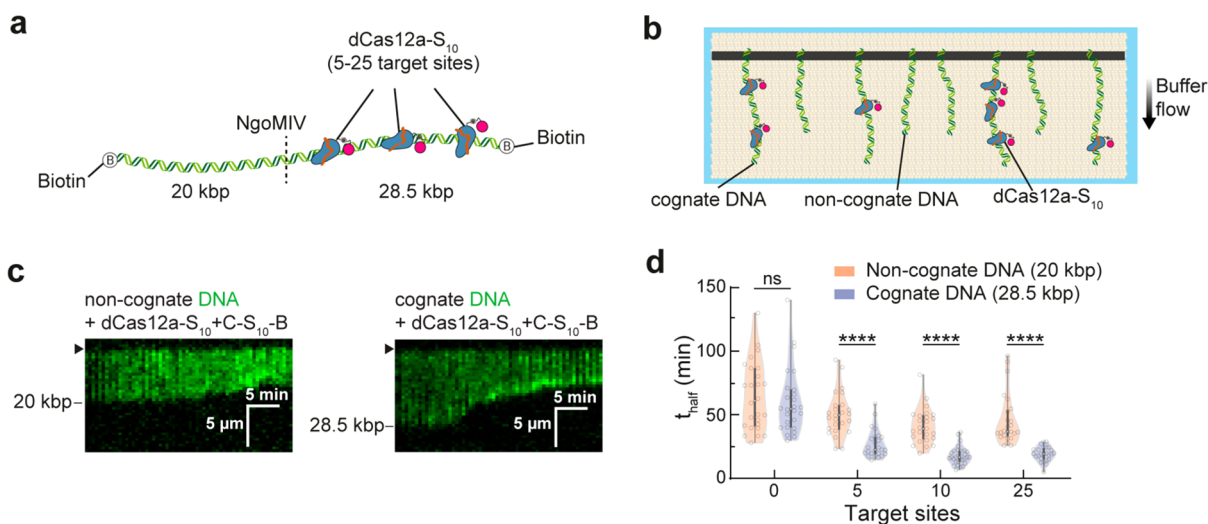


Figure 4. C-S₁₀-B selectively assembles on DNA that is decorated with dCas12a-S₁₀. (a) DNA was ligated with biotinylated oligos at both ends followed by cleavage with NgoMIV to generate two biotinylated DNA substrates (20 kbp and 28.5 kbp) distinguishable by size after YOYO-1 staining. (b) 28.5 kbp DNA (cognate DNA) was decorated with 5 to 25 dCas12a-S₁₀ prior to incubation with 10 nM C-S₁₀-B. (c) Representative kymographs showing encapsidation of the noncognate (20 kbp) and cognate (28.5 kbp) DNA substrates after decoration with ten dCas12a-S₁₀. Barrier position is indicated with arrows on the kymographs. (d) Violin plots showing the time (t_{half}) required to reach half of maximum condensation for both DNA substrates after decoration of the cognate DNA with 5 to 25 dCas12a-S₁₀. Extrapolation (after curve fitting to the Hill equation) was used to estimate t_{half} for molecules that did not reach 66.5% encapsidation during the experiment (30 min). (ns) and (****) indicate $p > 0.05$ and $p < 0.0001$, respectively.

Additionally, this work shows that the C-S₁₀-B protein can be re-engineered to more closely resemble virus-like features.

In summary, we have used CRISPR-dCas12a and coupled it to the polymerization domain S₁₀ to form a “packaging signal recognition complex” (in analogy with the assembly of viral capsid proteins around the viral genome) that triggers binding of

C-S₁₀-B and packaging of target DNA sequences. Importantly, such packaging signal recognition complex (and hence artificial particle nucleation and growth) can be easily redirected toward different DNA sequences by simple design of the crRNA without need of the more cumbersome manipulation of protein domains. Moreover, our results highlight the importance of

having multiple strong and specific interactions in templated nucleoprotein self-assemblies and should prove useful for developing new tailor-made nanomaterials for diverse biotechnological applications.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.nanolett.0c04640>.

Methods regarding proteins, crRNA synthesis, DNA curtain assay, imaging, dynamic light scattering, electrophoretic mobility shift assay, atomic force microscopy, and statistics; figures including results for C-S₁₀-B self-assembly on the DNA and DNA decoration with CRISPR-dCas12a, nucleosomes, and CRISPR-dCas12a-S₁₀; tables with crRNA sequences and parameters for multiple Gaussian fitting (PDF)

Video S1, DNA coating at 25 nM C-S₁₀-B (AVI)

Video S2, DNA coating at 150 nM C-S₁₀-B (AVI)

■ AUTHOR INFORMATION

Corresponding Authors

Ilya J. Finkelstein – Department of Molecular Biosciences, University of Texas at Austin, Austin, Texas 78712, United States; Institute for Cellular and Molecular Biology and Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas 78712, United States; orcid.org/0000-0002-9371-2431; Email: armandohg@iquimica.unam.mx

Armando Hernandez-Garcia – Department of Chemistry of Biomacromolecules, Institute of Chemistry, National Autonomous University of Mexico, Mexico City C.P. 04510, Mexico; orcid.org/0000-0002-2401-8139; Email: ifinkelstein@cm.utexas.edu

Author

Carlos Calcines-Cruz – Department of Chemistry of Biomacromolecules, Institute of Chemistry, National Autonomous University of Mexico, Mexico City C.P. 04510, Mexico; orcid.org/0000-0003-4524-8993

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.nanolett.0c04640>

Author Contributions

A.H.G. and I.J.F. conceived the idea for this study and designed the experiments together with C.C.C. C.C.C. performed all the experiments and analyzed the data jointly with I.J.F. and A.H.G. I.J.F., A.H.G., and C.C.C. wrote the manuscript, which was read and approved by all authors.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge a CONACyT-University of Texas (CONTEX) grant used to carry out this research. A.H.G. also thanks the IA200119 DGAPA-PAPIIT grant. This work was partially funded by the NIH (GM124141 to I.J.F.) and the Welch Foundation (F-1016 to I.J.F.). C.C.C. acknowledges the support of CONACyT for supporting his graduate studies at UT-Austin. We thank David Moreno Gutiérrez for providing some of the C-S₁₀-B protein and all the members of A.H.G.'s and I.J.F.'s laboratories for valuable discussions. We also thank the

technicians and appreciate the use of the facilities at the Institute of Chemistry at UNAM, and Tom Wandless at Stanford University for the kind donation of FKBP and FRB plasmids.

■ ABBREVIATIONS

crRNA, CRISPR RNA
FKBP, FK506 binding protein
FRB, FKBP rapamycin binding protein
PAM, protospacer adjacent motif
RNP, ribonucleoprotein particle
VLP, virus-like particle

■ REFERENCES

- (1) Hernandez-Garcia, A.; Kraft, D. J.; Janssen, A. F. J.; Bomans, P. H. H.; Sommerdijk, N. A. J. M.; Thies-Weesie, D. M. E.; Favretto, M. E.; Brock, R.; de Wolf, F. A.; Werten, M. W. T.; van der Schoot, P.; Stuart, M. C.; de Vries, R. Design and Self-Assembly of Simple Coat Proteins for Artificial Viruses. *Nat. Nanotechnol.* **2014**, *9* (9), 698–702.
- (2) Werten, M. W. T.; Wisselink, W. H.; Jansen-van den Bosch, T. J.; de Bruin, E. C.; de Wolf, F. A. Secreted Production of a Custom-Designed, Highly Hydrophilic Gelatin in *Pichia Pastoris*. *Protein Eng., Des. Sel.* **2001**, *14* (6), 447–454.
- (3) Krejchi, M.; Atkins, E.; Waddon, A.; Fournier, M.; Mason, T.; Tirrell, D. Chemical Sequence Control of β -Sheet Assembly in Macromolecular Crystals of Periodic Polypeptides. *Science* **1994**, *265* (5177), 1427–1432.
- (4) Smeenk, J. M.; Otten, M. B. J.; Thies, J.; Tirrell, D. A.; Stunnenberg, H. G.; van Hest, J. C. M. Controlled Assembly of Macromolecular β -Sheet Fibrils. *Angew. Chem., Int. Ed.* **2005**, *44* (13), 1968–1971.
- (5) Zhao, B.; Cohen Stuart, M. A.; Hall, C. K. Dock 'n Roll: Folding of a Silk-Inspired Polypeptide into an Amyloid-like Beta Solenoid. *Soft Matter* **2016**, *12* (16), 3721–3729.
- (6) Martin, M. E.; Rice, K. G. Peptide-Guided Gene Delivery. *AAPS J.* **2007**, *9* (1), E18–E29.
- (7) Hernandez-Garcia, A.; Werten, M. W. T.; Stuart, M. C.; de Wolf, F. A.; de Vries, R. Coating of Single DNA Molecules by Genetically Engineered Protein Diblock Copolymers. *Small* **2012**, *8* (22), 3491–3501.
- (8) Hernandez-Garcia, A.; Cohen Stuart, M. A.; de Vries, R. Templated Co-Assembly into Nanorods of Polyanions and Artificial Virus Capsid Proteins. *Soft Matter* **2018**, *14* (1), 132–139.
- (9) Punter, M. T. J. M.; Hernandez-Garcia, A.; Kraft, D. J.; de Vries, R.; van der Schoot, P. Self-Assembly Dynamics of Linear Virus-Like Particles: Theory and Experiment. *J. Phys. Chem. B* **2016**, *120* (26), 6286–6297.
- (10) Chai, S.; Lurz, R.; Alonso, J. C. The Small Subunit of the Terminase Enzyme of Bacillus Subtilis Bacteriophage SPP1 Forms a Specialized Nucleoprotein Complex with the Packaging Initiation Region. *J. Mol. Biol.* **1995**, *252* (4), 386–398.
- (11) Turner, D. R.; Joyce, L. E.; Butler, P. J. G. The Tobacco Mosaic Virus Assembly Origin RNA. *J. Mol. Biol.* **1988**, *203* (3), 531–547.
- (12) D'Souza, V.; Melamed, J.; Habib, D.; Pullen, K.; Wallace, K.; Summers, M. F. Identification of a High Affinity Nucleocapsid Protein Binding Element within the Moloney Murine Leukemia Virus Ψ -RNA Packaging Signal: Implications for Genome Recognition. *J. Mol. Biol.* **2001**, *314* (2), 217–232.
- (13) de Beer, T.; Fang, J.; Ortega, M.; Yang, Q.; Maes, L.; Duffy, C.; Berton, N.; Sippy, J.; Overduin, M.; Feiss, M.; Catalano, C. E. Insights into Specific DNA Recognition during the Assembly of a Viral Genome Packaging Machine. *Mol. Cell* **2002**, *9* (5), 981–991.
- (14) Borodavka, A.; Tuma, R.; Stockley, P. G. Evidence That Viral RNAs Have Evolved for Efficient, Two-Stage Packaging. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (39), 15769–15774.
- (15) Kutluay, S. B.; Zang, T.; Blanco-Melo, D.; Powell, C.; Jannain, D.; Errando, M.; Bieniasz, P. D. Global Changes in the RNA Binding

Specificity of HIV-1 Gag Regulate Virion Genesis. *Cell* **2014**, *159* (5), 1096–1109.

(16) Stewart, H.; Bingham, R. J.; White, S. J.; Dykeman, E. C.; Zothner, C.; Tuplin, A. K.; Stockley, P. G.; Twarock, R.; Harris, M. Identification of Novel RNA Secondary Structures within the Hepatitis C Virus Genome Reveals a Cooperative Involvement in Genome Packaging. *Sci. Rep.* **2016**, *6* (1), 22952.

(17) Patel, N.; Wroblewski, E.; Leonov, G.; Phillips, S. E. V.; Tuma, R.; Twarock, R.; Stockley, P. G. Rewriting Nature's Assembly Manual for a ssRNA Virus. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (46), 12255–12260.

(18) Morton, V. L.; Dykeman, E. C.; Stonehouse, N. J.; Ashcroft, A. E.; Twarock, R.; Stockley, P. G. The Impact of Viral RNA on Assembly Pathway Selection. *J. Mol. Biol.* **2010**, *401* (2), 298–308.

(19) Kraft, D. J.; Kegel, W. K.; van der Schoot, P. A Kinetic Zipper Model and the Assembly of Tobacco Mosaic Virus. *Biophys. J.* **2012**, *102* (12), 2845–2855.

(20) Comas-Garcia, M.; Datta, S. A.; Baker, L.; Varma, R.; Gudla, P. R.; Rein, A. Dissection of Specific Binding of HIV-1 Gag to the "packaging Signal" in Viral RNA. *eLife* **2017**, *6*, No. e27055.

(21) Comas-Garcia, M. Packaging of Genomic RNA in Positive-Sense Single-Stranded RNA Viruses: A Complex Story. *Viruses* **2019**, *11* (3), 253.

(22) Strohkendl, I.; Saifuddin, F. A.; Rybarski, J. R.; Finkelstein, I. J.; Russell, R. Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a. *Mol. Cell* **2018**, *71* (5), 816–824.e3.

(23) Zetsche, B.; Heidenreich, M.; Mohanraju, P.; Fedorova, I.; Kneppers, J.; DeGennaro, E. M.; Winblad, N.; Choudhury, S. R.; Abudayyeh, O. O.; Gootenberg, J. S.; Wu, W. Y.; Scott, D. A.; Severinov, K.; van der Oost, J.; Zhang, F. Multiplex Gene Editing by CRISPR–Cpf1 Using a Single CrRNA Array. *Nat. Biotechnol.* **2017**, *35* (1), 31–34.

(24) Gallardo, I. F.; Pasupathy, P.; Brown, M.; Manhart, C. M.; Neikirk, D. P.; Alani, E.; Finkelstein, I. J. High-Throughput Universal DNA Curtain Arrays for Single-Molecule Fluorescence Imaging. *Langmuir* **2015**, *31* (37), 10310–10317.

(25) Soniat, M. M.; Myler, L. R.; Schaub, J. M.; Kim, Y.; Gallardo, I. F.; Finkelstein, I. J. Next-Generation DNA Curtains for Single-Molecule Studies of Homologous Recombination. *Methods Enzymol.* **2017**, *592*, 259–281.

(26) Miné, J.; Disseau, L.; Takahashi, M.; Cappello, G.; Dutreix, M.; Viovy, J.-L. Real-Time Measurements of the Nucleation, Growth and Dissociation of Single Rad51–DNA Nucleoprotein Filaments. *Nucleic Acids Res.* **2007**, *35* (21), 7171–7187.

(27) Marchetti, M.; Kamsma, D.; Cazares Vargas, E.; Hernandez García, A.; van der Schoot, P.; de Vries, R.; Wuite, G. J. L.; Roos, W. H. Real-Time Assembly of Viruslike Nucleocapsids Elucidated at the Single-Particle Level. *Nano Lett.* **2019**, *19* (8), 5746–5753.

(28) Gao, L.; Cox, D. B. T.; Yan, W. X.; Manteiga, J. C.; Schneider, M. W.; Yamano, T.; Nishimasu, H.; Nureki, O.; Crosetto, N.; Zhang, F. Engineered Cpf1 Variants with Altered PAM Specificities Increase Genome Targeting Range. *Nat. Biotechnol.* **2017**, *35* (8), 789–792.

(29) Leenay, R. T.; Beisel, C. L. Deciphering, Communicating, and Engineering the CRISPR PAM. *J. Mol. Biol.* **2017**, *429* (2), 177–191.

(30) Nishimasu, H.; Shi, X.; Ishiguro, S.; Gao, L.; Hirano, S.; Okazaki, S.; Noda, T.; Abudayyeh, O. O.; Gootenberg, J. S.; Mori, H.; Oura, S.; Holmes, B.; Tanaka, M.; Seki, M.; Hirano, H.; Aburatani, H.; Ishitani, R.; Ikawa, M.; Yachie, N.; Zhang, F.; Nureki, O. Engineered CRISPR-Cas9 Nuclease with Expanded Targeting Space. *Science* **2018**, *361* (6408), 1259–1262.

(31) Walton, R. T.; Christie, K. A.; Whittaker, M. N.; Kleinstiver, B. P. Unconstrained Genome Targeting with Near-PAMless Engineered CRISPR-Cas9 Variants. *Science* **2020**, *368* (6488), 290–296.

Supplementary information

CRISPR-guided programmable self-assembly of artificial virus-like nucleocapsids

Carlos Calcines-Cruz¹, Ilya J. Finkelstein^{2,3,4,}, and Armando Hernandez-Garcia^{1,*}*

¹ Department of Chemistry of Biomacromolecules, Institute of Chemistry, National Autonomous University of Mexico, Mexico City C.P. 04510, Mexico

² Department of Molecular Biosciences, University of Texas at Austin, Austin, Texas 78712, USA

³ Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas 78712, USA

⁴ Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas 78712, USA

SUPPLEMENTARY FIGURES

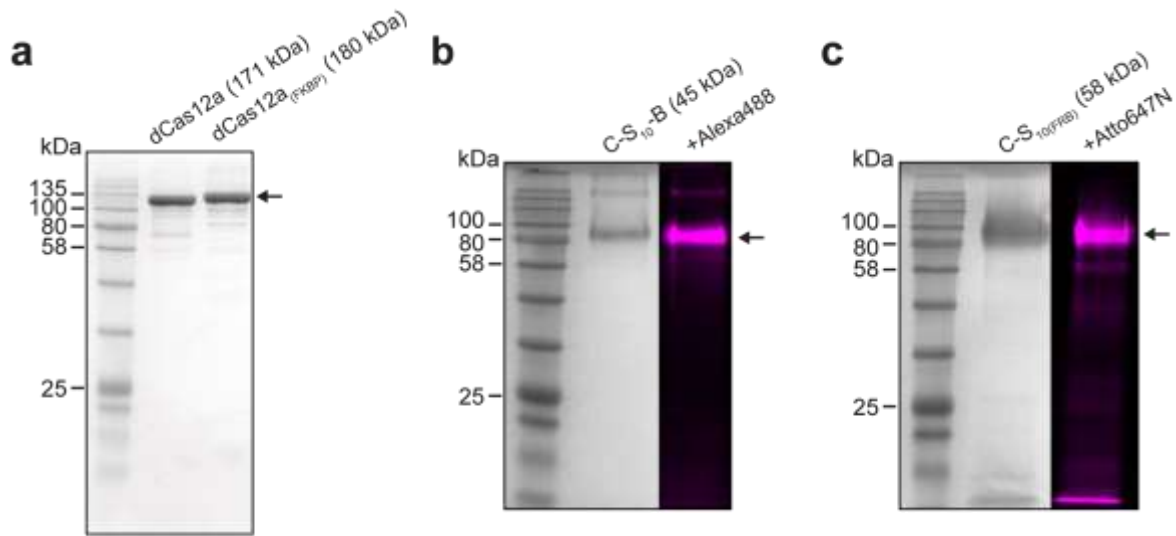


Figure S1: Proteins used in this study. **a:** SDS-PAGE gel with dCas12a (171 kDa) and the fusion protein containing dCas12a and the FKBP domain (dCas12a_(FKBP), 180 kDa). **b:** SDS-PAGE gel with Coomassie stained (left) and maleimide-Alexa488 labeled (right) C-S₁₀-B (45 kDa). **c:** SDS-PAGE gel with Coomassie-stained (left) and maleimide-Atto647N labeled (right) fusion protein containing C-S₁₀ and the FRB domain (C-S_{10(FRB)}, 58 kDa). Both C-S₁₀-B and C-S_{10(FRB)} show aberrant migration in gel because of the high proline content (22%) in the C domain.

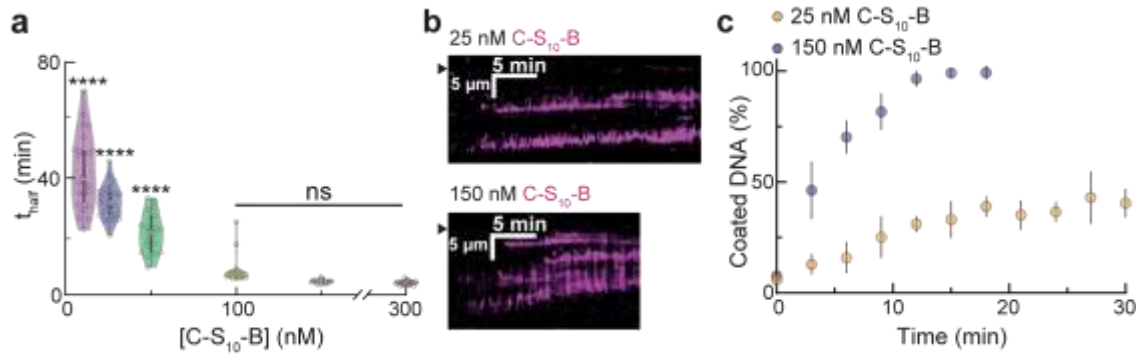


Figure S2: C-S₁₀-B concentration regulates DNA condensation rate and extent of compaction. **a:** The time (t_{half}) required to reach half of maximum condensation for each DNA molecule analyzed in Fig. 1d. We fit condensation profiles (Fig. 1d) to the Hill equation and estimated t_{half} for molecules that did not reach 66.5% encapsidation during the experiment (30 min). (ns) and (****) indicate $p > 0.05$ and $p < 0.0001$, respectively. **b:** Kymographs showing binding of C-S₁₀-B (magenta, Alexa488-labeled) on DNA (unlabeled) at the indicated C-S₁₀-B concentrations. Black arrows indicate the chromium barriers. **c:** Extent of DNA that is coated by fluorescent C-S₁₀-B. Points and error bars indicate the mean and standard deviation, respectively (N = 10 DNA molecules per condition).

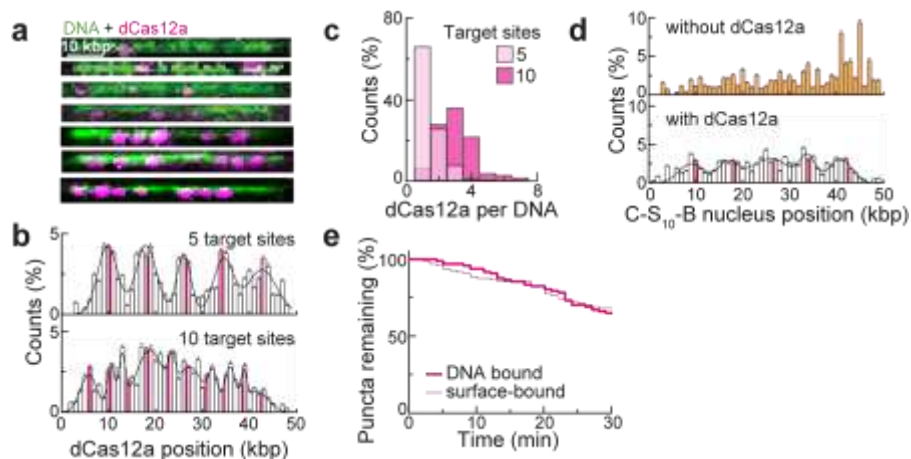


Figure S3: Characterization of dCas12a binding to DNA substrates prior to encapsidation by C-S₁₀-B. **a:** DNA (green, stained with YOYO-1) decorated with dCas12a (magenta, quantum dot-labeled). **b:** Histograms of dCas12a binding positions along the DNA. Apo dCas12a was reconstituted with a mixture of five (top) or ten (bottom) crRNAs. The top and bottom histograms summarize 561 and 473 dCas12a molecules, respectively. The expected binding positions are shown via the magenta lines in the histograms. The black lines are the fit to a sum of five (top) or ten (bottom) Gaussians; see Table S2 for fit parameters. **c:** Number of fluorescent dCas12a proteins per DNA strand after targeting five (N = 561) or ten (N = 473) binding sites. **d:** Distribution of C-S₁₀-B clusters on DNA without any dCas12a (top, N = 246) or pre-bound with dCas12a (bottom, targeting five sites, N = 484). dCas12a target sites are shown via strong magenta lines in the histogram. Black line: fit to a sum of five Gaussians; see Table S2 for fit parameters. **e:** dCas12a lifetime on DNA (with 25 nM C-S₁₀-B) is comparable to that of quantum dots on the flowcell surface (N = 100 each). These results indicate that dCas12a does not dissociate from DNA during the course of the encapsidation experiments.

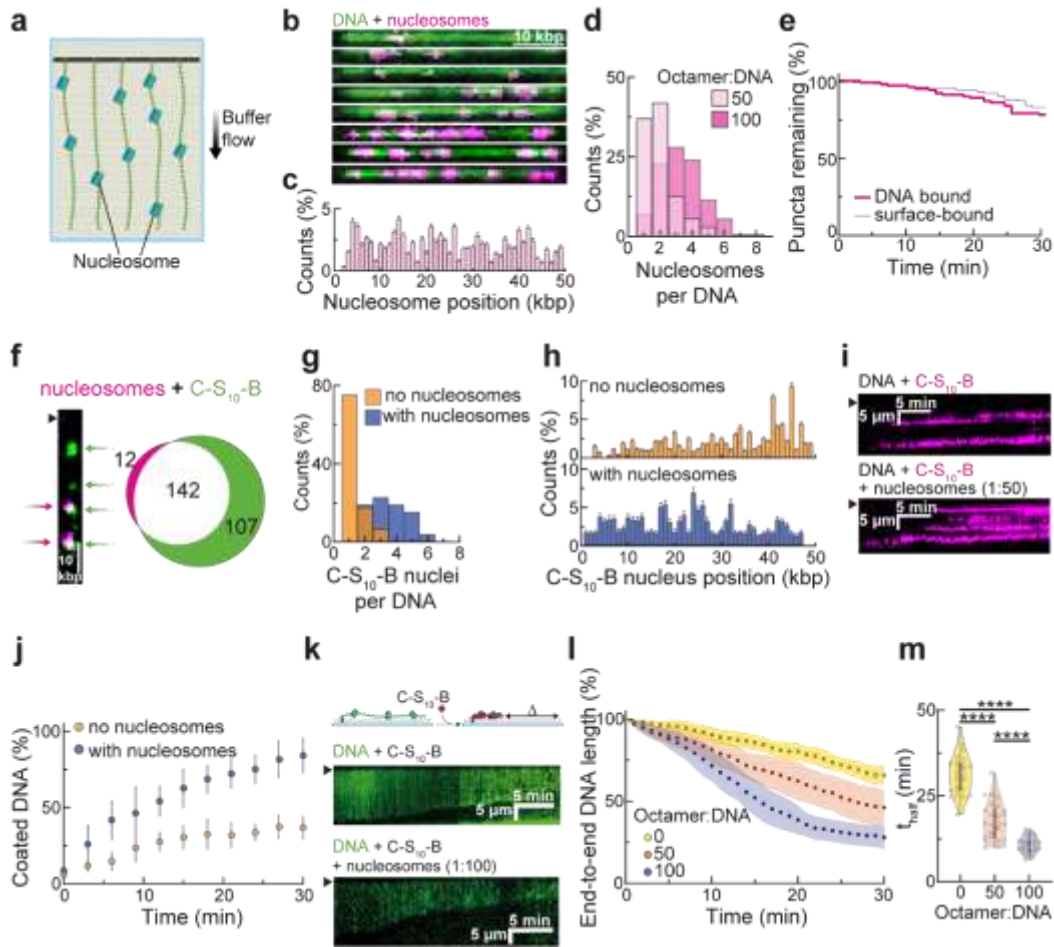


Figure S4: Nucleosomes accelerate DNA encapsidation by C-S₁₀-B. **a:** Illustration of nucleosome-decorated DNA. The nucleosomes were assembled via salt dialysis of histone octamers onto the DNA substrate. **b:** DNA (green, stained with YOYO-1) decorated with fluorescent nucleosomes (magenta, quantum dot-labeled). **c:** Histograms of nucleosome binding positions along the DNA at the 50:1 ratio (N = 279 nucleosomes). **d:** Histograms of the number of fluorescent nucleosomes per DNA strand at the 50:1 (N = 279) and 100:1 (N = 832) octamer:DNA molar ratios. **e:** Nucleosomes do not dissociate from DNA during encapsidation with 25 nM C-S₁₀-B. Surface-immobilized quantum dots are included as a comparison (N = 100 for each condition). **f:** Double labeling experiments show that 92% of nucleosomes (magenta) co-localized with 57% of fluorescent C-S₁₀-B clusters (green).

Barrier position is indicated with a black arrow. **g**: Histograms of the number of C-S₁₀-B clusters on DNA (N = 246 clusters) and nucleosomal DNA (N = 362 clusters; 50:1 octamer:DNA molar ratio). **h**: Distribution of C-S₁₀-B clusters along the DNA with (N = 362) or without (N = 246) nucleosomes deposited at a 50:1 octamer:DNA molar ratio. **i**: Kymographs showing binding of C-S₁₀-B (magenta, Alexa488-labeled) on non-decorated (top) or nucleosome-decorated (bottom) DNA at 25 nM C-S₁₀-B. Black arrows denote the barrier position. **j**: Coating of DNA by fluorescent C-S₁₀-B is faster after nucleosome deposition. Points and error bars indicate the mean and standard deviation, respectively (N = 10 DNA molecules per condition). **k**: Representative kymographs showing faster encapsidation by C-S₁₀-B for nucleosomal DNA (green) reconstituted at the 100:1 histone octamer:DNA molar ratio. Black arrows denote the barrier position. **l**: Condensation profiles at 25 nM C-S₁₀-B for DNA, and nucleosomal DNA reconstituted at the indicated histone octamer:DNA molar ratios. Shown are the mean and standard deviation for 25 DNA molecules per condition. **m**: Violin plots showing the time (t_{half}) required to reach half of maximum condensation for each DNA strand analyzed in (l). Extrapolation after curve fitting to the Hill equation was used to estimate t_{half} for molecules that did not reach 66.5% encapsidation during the experiment (30 min). (***) indicates $p < 0.0001$.

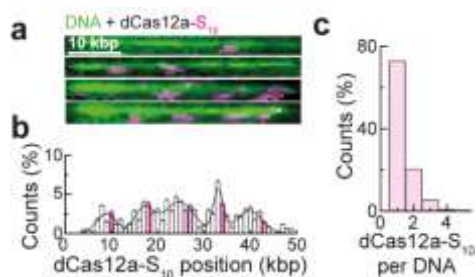


Figure S5: Decoration of DNA with the previously dimerized complex dCas12a-S₁₀. **a:** Representative DNA molecules (green, stained with YOYO-1) decorated with dCas12a-S₁₀ (magenta, Atto647N-labeled). **b:** Distribution of dCas12a-S₁₀ when binding was targeted to five sites along the DNA (N = 397 dCas12a-S₁₀). The expected binding positions of dCas12a-S₁₀ are shown via the magenta lines in the histogram. Black line: the fit to a sum of five Gaussians; see Table S2 for fit parameters. **c:** Histogram of the number of Atto647N-labeled dCas12a-S₁₀ per DNA strand (N = 397 dCas12a-S₁₀ molecules) along the DNA.

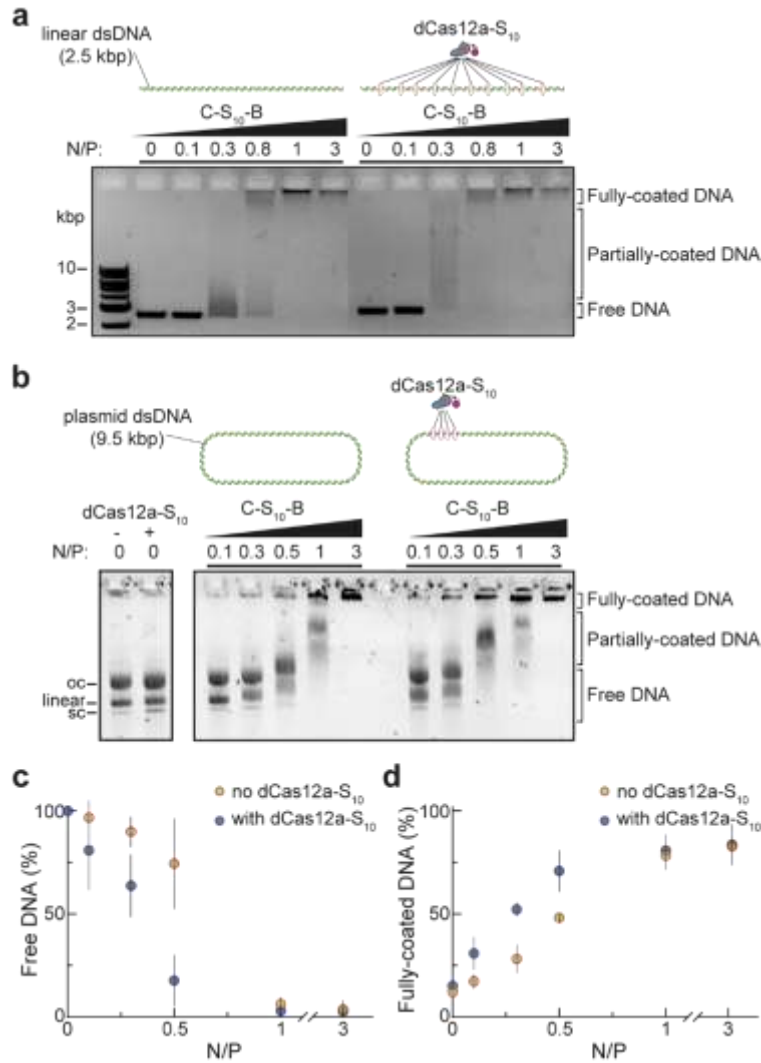


Figure S6: Positioning multiple dCas12a-S₁₀ on DNA improves encapsidation at sub-saturating C-S₁₀-B concentrations. C-S₁₀-B binding is assessed via electrophoretic mobility shift assays. **a:** A linear 2.5 kbp dsDNA fragment was decorated with ten dCas12a-S₁₀ uniformly distributed along the template via a pooled crRNA library. **b:** dCas12a-S₁₀ was directed to four identical sites (via a single crRNA) on a 9.5 kbp pPIC9 plasmid (in supercoiled, linear, and nicked open-circular conformations). **c:** Intensity of the bands corresponding to the free plasmid observed in (b). Shown are the mean and standard deviation for three replicates. **d:** Intensity of the band corresponding to the fully-coated plasmid observed in (b). Shown are the mean and standard deviation for three replicates. N/P:

stoichiometric ratio between C-S₁₀-B molecules and the number of available DNA binding sites (assuming a 6 bp footprint per C-S₁₀-B monomer).

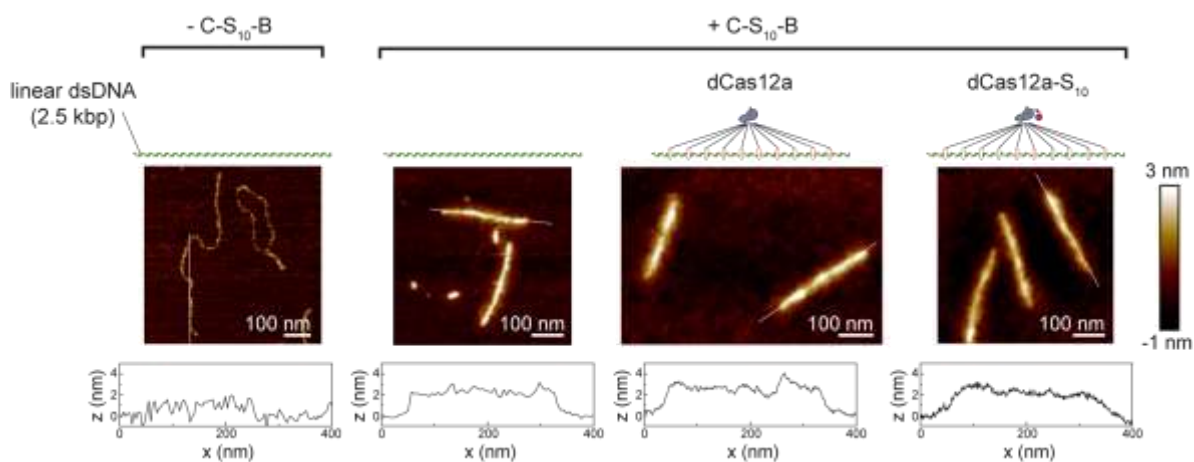


Figure S7: C-S₁₀-B packages DNA and DNA decorated with dCas12a or dCas12a-S₁₀ into artificial nucleocapsids. Shown are the 2.5 kbp (850 nm) DNA fragment, and particles obtained with DNA or decorated DNA after an overnight incubation with C-S₁₀-B. The height profiles of the particles are shown below.

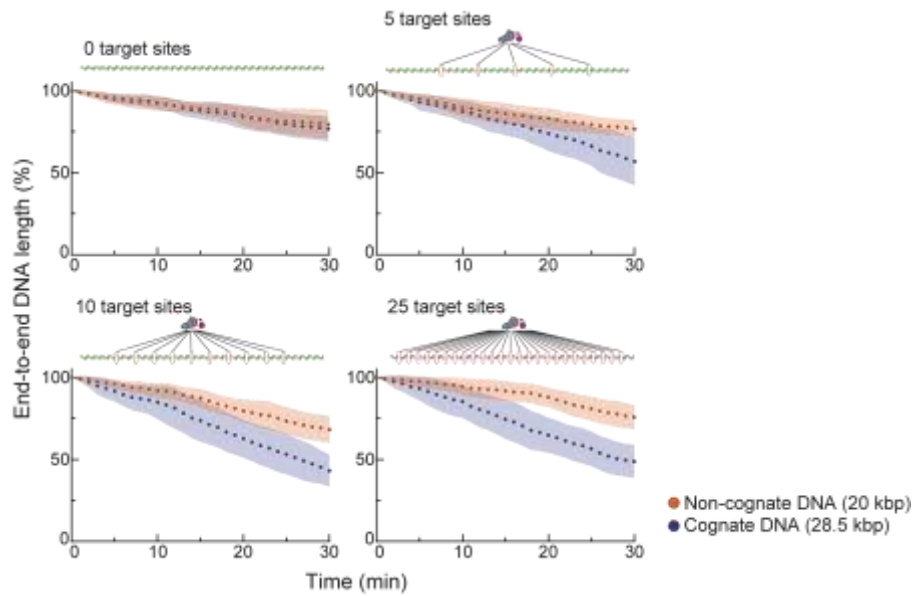


Figure S8: Encapsidation of cognate and non-cognate DNA templates after selective DNA decoration with dCas12a-S₁₀. Condensation profiles at 10 nM C-S₁₀-B for the non-cognate (20 kbp) and cognate (28.5 kbp) DNA strands after dCas12a-S₁₀ binding at zero, five, ten or twenty-five sites along the cognate DNA. Circles and shaded areas indicate the mean and standard deviation, respectively (N = 25 DNA molecules per condition).

Table S1: Variable sequence of crRNAs used for positioning dCas12a/dCas12a-S₁₀ on the DNA. All crRNAs consist of a 5' constant region (UAAUUUCUACUCUUGUAGAU, 20 nt) followed by the 3' variable region (24 nt) shown in the table.

Full λ DNA	
5 sites	AUGAUGUUCUGCUGGAUAUGCACU CCUGACACCGGACGGAAAGCUGAC AAUGUCGGCUAAUCGAUUUGGCCA GCUAGCAAUUAUGUGCAUCGAUU AUGAACGCAAUAUUCACAAGCAAU
10 sites	CGUGAGAGCUAUCCCUUCACCACG AUGAUGUUCUGCUGGAUAUGCACU CGUAUGUCGCCGGAAAGACUGGCUG CCUGACACCGGACGGAAAGCUGAC UGAUAUGCCGCAGAAACGUUGUAU AAUGUCGGCUAAUCGAUUUGGCCA AUGUUCAUCGUUCCUAAAAGACGC GCUAGCAAUUAUGUGCAUCGAUU CCGGACAGGAGCGUAAUGUGGCAG AUGAACGCAAUAUUCACAAGCAAU
28 kbp cognate strand	
5 sites	UGGCCAAAUCGAUUAGCCGACAUU GCGUCUUUAAGGAACGAUGAACAU AAUCGAUGCACAUAUUAUUGCUAGC CUGCCACAUAACGCUCCUGUCCGG AUUGCUUGUGAAUAUUGCGUUCAU
10 sites	UGGCCAAAUCGAUUAGCCGACAUU GUCAGAGGCUUGUGUUUGUGUCCU GCGUCUUUAAGGAACGAUGAACAU ACUGCGCAUCGCUGGCAUCACCUU AAUCGAUGCACAUAUUAUUGCUAGC GCUUAAUGACAUAUCCUUUCCCGAU CUGCCACAUAACGCUCCUGUCCGG CAGGAACGCAACCGCAGCUUAGAC AUUGCUUGUGAAUAUUGCGUUCAU AUACCGGAAGCAGAACCGGAUCAC
25 sites	UUGCUUCUCUUGACCGUAGGACUU CAGUAUUAUGUAGUCUGUUUUUUA UAAACUCCUUGCAAUGUAUGUCGU UGGCCUCGAAACCACCGAGCCGGA CAUCAUCCAGUCGAACUCACACA

UCACCGCAGAUGGUUAUCUGUAUG
 UCUGGCGAUUGAAGGGCUAAAUUC
 CGGGGUGGAUCUAUGAAAAACAUC
 AGAAGGAAGAUAUCCUCGCAUGGU
 CCAACCAAUGUAUAUCGAUACCG
 ACCUCAGAGAGAGGCUGAUCACU
 ACUUAUAGUAUUGGUUGCGUAAC
 UGAUAUGCCGCAGAAACGUUGUAU
 AAUGUCGGCUAAUCGAUUUGGCCA
 AUGUUCAUCGUUCCUUAAGACGC
 GCUAGCAAUUAUGUGCAUCGAUU
 CCGGACAGGAGCGUAAUGUGGCAG
 AUGAACGCAAUAUUCACAAGCAAU
 AGGCCACCGCAUCUCGUGCUGAAG
 UUGAAGCAAUCUGAAACCUAUUA
 AGGACACAAACACAAGCCUCUGAC
 AAGGUGAUGCCAGCGAUGCGCAGU
 AUCGGGAAAGGAAUGUCAUUAAGC
 GUCUAAGCUGCGGUUGCGUCCUG
 GUGAUCCGGUUCUGCUUCCGGUAU

Electrophoretic mobility shift assay

Linear dsDNA
 (10 sites)

UAGAGCAUAAGCAGCGCAACACCC
 AUGAUGAUUUGAACAGGAAGGCU
 UCUCUGCGAGCAUAAUGCCUGCGU
 GCCUCCCACGUCUCACCGAGCGUG
 UUGAUGGCCUCAUCCACACGCAGC
 UAACCGCUUCACACUGACGCCGGA
 UUGUUGGUUGCUGCACCAUCCUCU
 UGUAUGAAAACGCCCACCAUUCCC
 CGGCUCAGUCAUCGCCAAGCUGG
 CGGACACAGUUCGGGAUGGUCAGC

Plasmid DNA (4 repeats of 1 site)

CCUUGUUCACCUUGGUUGACCAGGG

Table S2: Parameters of multiple Gaussian fitting. Binding of dCas12a, dCas12a-S₁₀ and C-S₁₀-B clusters was fit to the sum of five or ten Gaussians in Figs. S3b, S3d and S5b. For each Gaussian we report the mean \pm standard deviation, and the offset with respect to the expected position in parenthesis (in kbp).

Expected positions	dCas12a, 5 target sites (Fig. S3b top)	dCas12a, 10 target sites (Fig. S3b bottom)	C-S₁₀-B clusters, 5 target sites (Fig. S3d)	dCas12a-S₁₀, 5 target sites (Fig. S5b)
6.2	-	5.9 \pm 3.6 (0.4)	-	-
10.3	9.7 \pm 2.1 (0.5)	10.5 \pm 1.4 (0.3)	9.0 \pm 2.8 (1.3)	8.0 \pm 2.1 (1.3)
14.3	-	12.9 \pm 1.3 (1.4)	-	-
18.4	18.0 \pm 2.2 (0.4)	18.8 \pm 6.0 (0.4)	16.7 \pm 2.0 (1.7)	17.1 \pm 2.0 (1.3)
22.6	-	23.5 \pm 1.3 (0.9)	-	-
26.6	26.0 \pm 1.5 (0.6)	27.5 \pm 5.3 (0.9)	25.8 \pm 4.5 (0.9)	24.7 \pm 3.3 (2.0)
30.7	-	32.2 \pm 1.8 (1.5)	-	-
34.7	34.5 \pm 2.1 (0.2)	35.9 \pm 2.3 (1.2)	33.8 \pm 1.6 (0.9)	33.1 \pm 1.1 (1.6)
38.7	-	38.8 \pm 1.1 (0.1)	-	-
42.8	42.7 \pm 3.1 (0.1)	41.6 \pm 4.7 (1.2)	40.7 \pm 2.4 (2.1)	39.8 \pm 2.8 (3.1)
R²	0.6978	0.86921	0.5582	0.6511

VIDEO CAPTIONS

Video 1: DNA coating at 25 nM C-S₁₀-B. Binding of fluorescent C-S₁₀-B (magenta, at 25 nM) on a single DNA molecule (unlabeled). Only two clusters of C-S₁₀-B are observed, a larger one (top, and first to form) and a smaller one at the free end of the DNA. The video is displayed at a rate 150X (10 frames per second) from that at which it was recorded (1 frame/15 seconds). The scale (10 kbp) is shown with a white vertical line.

Video 2: DNA coating at 150 nM C-S₁₀-B. Binding of fluorescent C-S₁₀-B (magenta, at 150 nM) on a single DNA molecule (unlabeled). Almost total coating of the DNA occurs halfway through the experiment. The video is displayed at a rate 150X (10 frames per second) from that at which it was recorded (1 frame/15 seconds). The scale (10 kbp) is shown with a white vertical line.

EXPERIMENTAL SECTION

Plasmids, protein expression and purification

The FK506 Binding Protein (FKBP) and FKBP-Rapamycin Binding domain (FRB) sequences were extracted from pGEX-2T plasmids (kindly provided by Tom Wandless, Stanford University) and inserted at the C-terminus of nuclease inactivated version of AsCas12a (D908A) and C-S₁₀, respectively.

The pET19 plasmids with dCas12a and dCas12a_(FKBP) (both with an N-terminus 6XHis-SUMO tag) were used to transform BL21 (DE3) *Escherichia coli*. Colonies were grown on Luria Bertani plates with 100 µg/mL ampicillin and used to inoculate Terrific Broth liquid culture in Fernbach flasks. Cultures were grown at 37°C until OD₆₀₀ ~0.8 and then transferred to 12°C for 1 h before induction with 1 mM IPTG. Cultures were grown for 24 h at 12°C and the pellets were collected by centrifugation and stored at -70°C.

The pellets were resuspended in buffer A (20 mM Tris-HCl pH 8.0, 250 mM NaCl, 10 mM imidazole) supplemented with 1 mM PMSF and sonicated on ice. The lysate was clarified by centrifugation at 30 000 g for 45 min and the supernatant containing the protein of interest was filtered through 0.2 µm syringe filters. The supernatant was then loaded into a 5 mL HisTrap (GE Healthcare) column, washed with buffer B (20 mM Tris-HCl pH 8.0, 1 M NaCl, 25 mM imidazole) and eluted with buffer C (20 mM Tris-HCl pH 8.0, 1 M NaCl, 250 mM imidazole). The eluted fraction was loaded into a dialysis membrane with 2.4 µM SUMO protease and dialyzed overnight at 4°C in buffer D for dCas12a (50 mM phosphate buffer pH 6.0, 100 mM KCl, 5 mM MgCl₂, 10% glycerol, 2 mM DTT) and buffer E for dCas12a_(FKBP) (20 mM HEPES-KOH pH 7.2, 100 mM KCl, 5 mM MgCl₂, 10% glycerol, 2 mM DTT). The 0.2 µm filtered cleaved product was injected in a 5 mL HiTrap

SP HP (GE Healthcare) cationic exchange column and eluted with a gradient to buffer F (20 mM HEPES-KOH pH 7.5, 2 M KCl). The fractions containing the protein of interest were pooled, concentrated with 100 kDa Amicon centrifugal filter units and injected into a size exclusion HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated with storage buffer G (20 mM HEPES-KOH pH 7.5, 500 mM KCl, 10% glycerol). Purified proteins were concentrated to 10-12 μ M, flash frozen in liquid nitrogen and stored at -70°C until use. Protein identity and purity were assessed with western blots and SDS-PAGE, respectively.

pPIC9 plasmids with C-S₁₀-B and C-S₁₀(FRB) were linearized with SacI and electroporated into histidine auxotrophic *Pichia pastoris* GS115. Mut⁺ colonies expressing the protein were used to inoculate 300 mL of MGY and grown for 24 h until OD₆₀₀ ~6.0. Cells were pelleted and resuspended in MM medium for protein expression for 72 h, with 1% methanol addition every 24 h. Culture supernatant was collected by centrifugation, the medium pH was adjusted to 8.0 with NaOH and 1 mM PMSF, 5 mM EDTA was added to inhibit proteolysis followed by 0.2 μ m filtration.

For C-S₁₀-B, the culture supernatant was saturated to 50% with ammonium sulfate and incubated overnight at 4°C. The protein precipitate was resuspended in Milli-Q water at 65°C and the precipitation step was repeated once. The precipitate was resuspended in Milli-Q water at 65°C before 50 mM NaCl and 40% acetone addition. The sample was centrifuged to discard precipitates and the concentration of acetone in the supernatant was increased to 80% to selectively precipitate C-S₁₀-B. The precipitate was air-dried for 15 min, resuspended, and dialyzed overnight against Milli-Q water at 4°C, followed by flash

freeze and lyophilization. Before use, the lyophilized protein was dissolved in Milli-Q water at 0.1 mg/mL and incubated at 70°C for 15 min.

For C-S_{10(FRB)}, the culture supernatant was concentrated with 30 kDa Amicon stirred cells and proteins were precipitated with 80% saturation ammonium sulfate at 4°C overnight. The protein precipitate was resuspended in PBS pH 7.5 and the precipitation step was repeated once. The precipitate was resuspended in PBS pH 7.5 and 40% acetone was added followed by centrifugation. The concentration of acetone in the supernatant was increased to 80% to selectively precipitate C-S_{10(FRB)}. The precipitate was resuspended to ~80 μM in PBS pH 7.5, dialyzed overnight against PBS pH 7.5 at 4°C and flash frozen. Protein identity and purity of C-S_{10-B} and C-S_{10(FRB)} were assessed with mass spectrometry (MALDI-TOF) and SDS-PAGE, respectively.

crRNA pools

The 24 nt forward oligo with the promoter for T7 RNA polymerase (GAAATTAATACGACTCACTATAGG), and the 68 nt reverse oligos with the 24 nt sequence complementary to the forward oligo plus the 44 nt crRNA sequence (see Table S1) were purchased from IDT. The T7 promoter forward oligo was mixed with pools of reverse oligos (1.5:1.0 molar ratio) and hybridized by incubation at 75°C for 5 min and then cooling to 25°C over 25 min in annealing buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA). The pool of partially double stranded DNA templates was used for *in vitro* transcription with HiScribe™ T7 Quick High Yield RNA Synthesis Kit (New England BioLabs) and the crRNA was purified with TRizol (Ambion). crRNA integrity and purity were assessed with 15% acrylamide Urea-PAGE.

Preparation of DNA substrates

To prepare the DNA template for curtain assays, λ DNA (125 μ g, NEB) was incubated with 2 μ M biotinylated oligo complementary to one of the two 12 nt cohesive ends in λ DNA in T4 DNA ligase reaction buffer (NEB) and hybridized at 70°C for 15 min followed by cooling to 15°C over 2 h. T4 DNA ligase (2000 units, NEB) was added and the mixture was incubated at room temperature overnight. The ligase was inactivated with 2 M NaCl and the sample was injected into a Sephacryl S-1000 size exclusion column (GE Healthcare) to remove excess oligos and DNA ligase.

For experiments with nucleosomal DNA, purified biotinylated DNA was precipitated with sodium acetate (pH 5.5) and 0.3 M isopropanol to 1:1 v/v on ice, and centrifuged at 15 000 g for 30 min. The precipitate was resuspended to 0.8 nM and incubated with human histone octamers (3xHA H2A and H2B, H3, H4; Histone Source) at molar ratios of 50:1 and 100:1 (histone octamer to λ DNA) in 2 M TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 2 M NaCl) with 1 mM DTT. The 100 μ L mixture was loaded in dialysis buttons (10 kDa MWCO, BioRad) and dialyzed against dialysis buffer (10 mM Tris-HCl pH 7.6, 1 mM EDTA, 1 mM DTT and NaCl). The first 1 h dialysis step was done against 1.5 M NaCl, followed by 2 h dialysis steps against 1, 0.8, 0.6 and 0.4 mM with a final overnight step against 0.2 M NaCl at 4°C.

For experiments with two DNA strands, λ DNA was ligated with biotinylated oligos complementary to both cohesive ends. Hybridization and ligation were performed as described above except that CutSmart buffer (NEB) supplemented with 1 mM rATP was used instead of T4 ligase buffer. The mixture was heated at 65°C for 10 min to inactivate the ligase, and NgoMIV was added for DNA cleavage at 37°C for 3 h. Lastly, 2 M NaCl

was added to the solution prior to size exclusion chromatography. Complete cleavage by NgoMIV was assessed with gel electrophoresis. All DNA substrates were stored at 4°C until use.

Single-tethered DNA curtain assay

A solution containing 20 µL of a liposome stock solution (97.7% DOPC, 2.0% DOPE-mPEG2k and 0.3% DOPE-biotin) and 980 µL of lipid buffer (10 mM Tris-HCl pH 8, 100 mM NaCl) was injected into the assembled flowcell and incubated for 30 min at room temperature. The flowcell was washed with BSA buffer (40 mM Tris-HCl pH 8, 2 mM MgCl₂, 0.2 mg/mL BSA) and incubated for 10 min before injection of 0.1 mg/mL streptavidin in BSA buffer followed by 10 min incubation. The biotinylated λDNA was diluted in BSA buffer and injected into the flowcell for tethering, excess DNA was removed with buffer. Imaging buffer consisted of BSA buffer supplemented with 100 mM NaCl, 5 mM MgCl₂, 2 mM DTT and gloxy solution (500 units of catalase, 70 units of glucose oxidase and 1% glucose w/v). YOYO-1 was added to the imaging buffer when DNA staining was required.

Total internal reflection fluorescence images were acquired with an inverted Nikon Ti-E microscope. Excitation (488 nm) and emission signals were split with a 638 nm dichroic beam splitter (Chroma) and captured by two EM-CCD cameras (Andor iXon DU897).

dCas12a or dCas12a_(FKBP) ribonucleoprotein complexes were prepared by mixing the protein with crRNA pools at 1:10 molar ratio in reaction buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM MgCl₂, 2% glycerol, 2 mM DTT) followed by incubation at 37°C for

30 min. The complexes were diluted to 10 nM in imaging buffer, injected in the flowcell and incubated at room temperature for 30 min.

Labeling of dCas12a and nucleosomes was carried out *in situ* by injection into the flowcell of anti-FLAG or anti-HA coupled quantum dots (QD₇₀₅), respectively. C-S₁₀-B and C-S_{10(FRB)} were labeled via maleimide reaction with Alexa488 and Atto647N, respectively. Both C-S₁₀-B and C-S_{10(FRB)} contain an N-terminal cysteine residue and C-S_{10(FRB)} has an additional cysteine in the FRB moiety.

Images were processed with FIJI software. DNA condensation kinetics were obtained by measuring the DNA length at each time step with rate 1 frame/min. DNA coating by fluorescent C-S₁₀-B was obtained by measuring the number of fluorescent pixels on the vertical path along the DNA. For this experiment, 100% indicates a DNA that has been fully coated with C-S₁₀-B.

Dynamic light scattering

Dimerization of dCas12a_(FKBP) and C-S_{10(FRB)} via rapamycin was assessed by dynamic light scattering with a Zetasizer μ V (Malvern). Both proteins were diluted to 2 μ M in reaction buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM MgCl₂, 2% glycerol, 1 mM DTT) for a total volume of 20 μ L and loaded in the cell. Measurements were taken before and immediately after (\sim 30 s) addition of 5 μ M rapamycin.

Electrophoretic mobility shift assay

The crRNA-loaded dCas12a-S₁₀ complex was incubated with 100 ng DNA (linear with 10 different target sites or plasmid with 4 identical target sites) at stoichiometry 10:1 ([dCas12a]:[target site]) for 1 h before C-S₁₀-B addition at 0 to 3 N/P ratios ([C-S₁₀-

B]:[DNA binding sites of 6 bp]) for 3 h (linear DNA) or overnight incubation (plasmid). Encapsidation was assessed on agarose gel electrophoresis at 4°C and DNA bands were stained with SYBR Green.

Gel images were processed with FIJI. To quantify the EMSAs we calculated the % of the DNA that remained free and the % of the DNA that was stalled at the well (for each lane). This quantification renders how much of the total DNA (per lane) is either in its free form or completely encapsidated.

Atomic force microscopy

crRNA-loaded dCas12a and dCas12a-S₁₀ were mixed with a 2.5 kbp linear dsDNA fragment (10 target sites) at stoichiometry 10:1 ([dCas12a]:[target site]) for 1 h. C-S₁₀-B was added at N/P = 3 and incubated overnight. The sample with 5 nM DNA was deposited on the mica and incubated for 10 min before rinsing with 0.5 mL Milli-Q water and air-drying. Images were acquired on a Digital Instruments NanoScope V (Bruker) in ScanAssyst™ mode and processed with NanoScope Analysis 1.20.

Statistics and data analysis

All results are given as mean ± standard deviation except for Fig. 1e (mean ± 95% confidence intervals). Non-parametric statistical tests for two (Mann-Whitney) and multiple (Kruskal-Wallis) means were performed on GraphPad Prism 8. Multiple Gaussian fitting was performed with OriginPro 8.5. Errors bars for binding distributions of dCas12a, nucleosomes, C-S₁₀-B clusters and dCas12a-S₁₀ were generated by bootstrap analysis.

To calculate t_{half} values, the curves with DNA compaction versus time were fit to the Hill equation (Equation 1). Here, $DNA\ length_0$ and $DNA\ length_f$ are the lengths of the fully

extended DNA (100%) and the fully compacted DNA (33%), respectively. *Hill* defines the steepness of the curve and t_{half} is the time required to achieve the midpoint between 100% and 33% (= 66.5%).

$$\text{DNA length} = \text{DNA length}_f + \left(\frac{\text{DNA length}_0 - \text{DNA length}_f}{1 + (t_{half}/\text{time})^{\text{Hill}}} \right) \quad (1)$$

Compaction rate versus C-S₁₀-B concentration (Fig. 1e) was also fit to the Hill equation (Equation 2). Here, $rate_0$ and $rate_f$ stand for the condensation rates at the lowest and highest C-S₁₀-B concentrations, respectively. *Hill* defines the steepness of the curve and $[C-S_{10}-B]_{50}$ is the C-S₁₀-B concentration at which the midpoint between $rate_0$ and $rate_f$ was achieved.

$$\text{rate} = \text{rate}_0 + \left(\frac{\text{rate}_f - \text{rate}_0}{1 + ([C-S_{10}-B]_{50}/[C-S_{10}-B])^{\text{Hill}}} \right) \quad (2)$$

Nature of Amorphous Hydrophilic Block Affects Self-Assembly of an Artificial Viral Coat Polypeptide

Lione Willems,[†] Larissa van Westerveld,[†] Stefan Roberts,[§] Isaac Weitzhandler,[§] Carlos Calcines Cruz,^{||} Armando Hernandez-Garcia,^{||} Ashutosh Chilkoti,[§] Enrico Mastrobattista,[⊥] John van der Oost,[‡] and Renko de Vries^{*,†,⊕}

[†]Physical Chemistry and Soft Matter and [‡]Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

[§]Department of Biomedical Engineering, Duke University, Durham, North Carolina 27708, United States

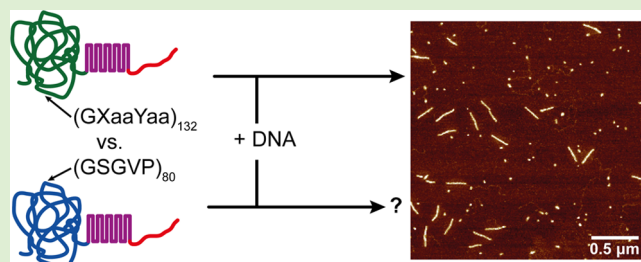
^{||}Institute of Chemistry, Department of Chemistry of Biomacromolecules, National Autonomous University of Mexico, 04510 Mexico City, Mexico

[⊥]Department of Pharmaceutics, Utrecht Institute of Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands

Supporting Information

ABSTRACT: Consensus motifs for sequences of both crystallizable and amorphous blocks in silks and natural structural analogues of silks vary widely. To design novel silklike polypeptides, an important question is therefore how the nature of either the crystallizable or the amorphous block affects the self-assembly and resulting physical properties of silklike polypeptides. We address herein the influence of the amorphous block on the self-assembly of a silklike polypeptide that was previously designed to encapsulate single DNA molecules into rod-shaped viruslike particles. The polypeptide

has a triblock architecture, with a long N-terminal amorphous block, a crystallizable midblock, and a C-terminal DNA-binding block. We compare the self-assembly behavior of a triblock with a very hydrophilic collagen-like amorphous block (GXaaYaa)₁₃₂ to that of a triblock with a less hydrophilic elastin-like amorphous block (GSGVP)₈₀. The amorphous blocks have similar lengths and both adopt a random coil structure in solution. Nevertheless, atomic force microscopy revealed significant differences in the self-assembly behavior of the triblocks. If collagen-like amorphous blocks are used, there is a clear distinction between very short polypeptide-only fibrils and much longer fibrils with encapsulated DNA. If elastin-like amorphous blocks are used, DNA is still encapsulated, but the polypeptide-only fibrils are now much longer and their size distribution partially overlaps with that of the encapsulated DNA fibrils. We attribute the difference to the more hydrophilic nature of the collagen-like amorphous block, which more strongly opposes the growth of polypeptide-only fibrils than the elastin-like amorphous blocks. Our work illustrates that differences in the chemical nature of amorphous blocks can strongly influence the self-assembly and hence the functionality of engineered silklike polypeptides.



INTRODUCTION

Self-assembly of proteins is increasingly exploited to design scaffolds for cell growth, nanoparticles for drug delivery, biosensors, and nanoreactors.^{1,2} A relatively straightforward approach to the de novo design of self-assembling structural proteins is the use of well-conserved motifs found in natural structural proteins such as elastin, silk, and collagen. These consensus motifs can be repeated and combined in many ways to design sequences for novel self-assembling polypeptides for a wide range of applications.^{1,3}

An important class of self-assembling polypeptides is silklike polypeptides that typically consist of both crystallizable and amorphous blocks. Consensus motifs for sequences of both crystallizable and amorphous blocks in natural silks and structural silk analogues, such as spider silks,⁴ silkworm

silks,⁵ and squid sucker ring teeth,⁶ vary widely. An obvious question is therefore how the nature of the crystallizable and the amorphous blocks affects the self-assembly and the resulting physical properties of these materials. The same question also arises for de novo design of silklike polypeptides, but so far this issue has not yet been addressed in the literature.

We here consider the influence of the nature of the amorphous block on self-assembly of a triblock polypeptide designed previously to encapsulate single DNA molecules into rod-shaped viruslike particles (VLPs).⁷ The triblock polypeptide C-S^Q₁₀-K₁₂ features a dodeca-lysine block (K₁₂) for binding

Received: April 15, 2019

Revised: August 8, 2019

Published: August 16, 2019

to the nucleic acid template, a silklike midblock $S_{10}^Q = (\text{GAGAGAGQ})_{10}$ that self-assembles into the rod-shaped core of the VLP via stacking of β -rolls,^{8–11} and a collagen-like amorphous block $C = (\text{GXaaYaa})_{132}$ for colloidal stabilization of the VLP particles. Note that the Xaa and Yaa residues in the collagen-like C block are chosen to be highly hydrophilic and mostly uncharged such that the resulting polypeptide adopts a random coil conformation rather than a triple helix.¹²

Apart from a general understanding of the influence of the nature of amorphous blocks on the properties of silk polypeptides, we also have more specific reasons for exploring other kinds of amorphous blocks for the $C\text{-}S_{10}^Q\text{-}K_{12}$ triblock polypeptide. First of all, the nature of the amorphous block strongly impacts the production of the polypeptide in host organisms: while the $C\text{-}S_{10}^Q\text{-}K_{12}$ polypeptide can be obtained at a high yield by secreted expression in the yeast *Pichia pastoris*,⁷ expression in *Escherichia coli* is problematic (see SI-1). Next, with respect to applications, since the amorphous block is on the outside of a VLP, it is the amorphous block that will, to a large extent, determine the initial biological response of cells and tissues to the VLP. Therefore, it is of interest to explore amorphous blocks that have been very well characterized with respect to their biological response, such as elastin-like polypeptides. Elastin-like polypeptides (ELPs) with the general sequence $E_n^X = (\text{GXGVP})_n$ are known to be expressed at high yield in *E. coli*,¹³ whereas for *P. pastoris*, expression levels for ELPs are relatively low, especially for ELPs with more hydrophobic guest residues X.¹⁴ For studying our previously designed triblock polypeptide with an elastin-like rather than a collagen-like amorphous block, we therefore choose production in *E. coli*.

To study the impact of the nature of the amorphous block, we compare self-assembly and co-assembly with DNA of a new polypeptide $E_{80}^S\text{-}S_{10}^Q\text{-}K_{12}$ (produced in *E. coli*) with that of the $C\text{-}S_{10}^Q\text{-}K_{12}$ polypeptide (previously produced in *P. pastoris*). To make a good comparison with the amorphous collagen-like C block, we use an elastin-like polypeptide with a length of 80 pentamers (E_{80}^S) to match the length of the approximately 400 amino-acid-long C block. As a guest residue in the elastin-like polypeptide, we use serine ($X = S$) since it is uncharged and hydrophilic. For this guest residue, at the conditions of co-assembly of the polypeptide with DNA, the ELP block will be far below its transition temperature and adopt a random coil configuration.

MATERIALS AND METHODS

Materials. The $C\text{-}S_{10}^Q\text{-}K_{12}$ polypeptide was produced and purified as described previously.⁷ For the full amino acid sequence of the $C\text{-}S_{10}^Q\text{-}K_{12}$ polypeptide, in particular, the choice of Xaa and Yaa amino acids in the collagen-like amorphous block $C = (\text{GXaaYaa})$, see ref 7.

A pET24a(+) PRe-RDL cloning vector was constructed as previously described.¹⁵ Custom oligonucleotides were synthesized by Integrated DNA Technologies Inc. All restriction enzymes, the calf-intestinal phosphatase (CIP), and the Quick Ligation kit were purchased from New England Biolabs, and the T4 DNA ligase buffer from Invitrogen. DNA miniprep, gel purification, and PCR purification kits were from Qiagen Inc. Chemically competent *E. coli* cells (strains EBSAlpha and BL21 (DE3)) were ordered from EdgeBioSystems. The NoLimits 2000 bp linear dsDNA, PageBlue protein staining solution, SYBR Safe DNA stain, 6 \times DNA loading dye, GeneRuler 1 kb DNA ladder, and 50 \times TAE buffer were purchased from Thermo Fisher Scientific. The 10% mini-protean TGX precast protein gels and Precision Plus Protein All Blue Prestained Protein Standard were ordered from Bio-Rad, and agarose was bought from Brunschwig Chemie.

Preparation of Oligolysine-Encoding Cloning Vector. Two complementary custom oligonucleotides encoding the oligolysine K_{12} were designed: 5'-cAAGAAAAAGAAGAAAAAGAAGAAGAAGAA-AAAGAAGgg3' and 5'-CTTCTTTTCTTCTTCTTCTTTTCTTCTTTTCTTgcc-3'. To anneal the two oligonucleotides, they were diluted in T4 DNA ligase buffer to a concentration of 2 μM , heated to 95 $^\circ\text{C}$ for 2 min, and then slowly cooled to room temperature over 3 h. Linear pET24a(+) PRe-RDL cloning vector was prepared by the digestion of approximately 2 μg of vector for 2 h at 37 $^\circ\text{C}$ with 5 U of *BseRI*; the 5' ends of the linearized vector was dephosphorylated using 10 U of CIP for 30 min at 37 $^\circ\text{C}$. Next, a PCR purification kit was used to remove the enzymes and the annealed oligonucleotides were ligated into the linearized vector by incubating a mixture of both in 1 \times Quick ligase buffer for 5 min at room temperature in the presence of Quick ligase. The ligation product was then transformed into EBSAlpha chemically competent cells and the cells were plated on TBdry plates supplemented with 45 $\mu\text{g}/\text{mL}$ kanamycin. Colonies containing the plasmid with a copy of the oligolysine K_{12} were selected by colony PCR and confirmed by DNA sequencing.

Gene Construction by PRe-RDL. The method of recursive directional ligation by plasmid reconstruction (PRe-RDL)¹⁵ was used to construct a plasmid encoding for the $E_{80}^S\text{-}S_{10}^Q\text{-}K_{12}$ polypeptide. To obtain the so-called A-fragment, a previously produced¹⁶ PRe-RDL cloning vector coding for $E_{80}^S\text{-}S_{10}^Q$ was digested with *BglII* and *AclI* enzymes for 3 h at 37 $^\circ\text{C}$ and the *BglII* \times *AclI* fragment containing the $E_{80}^S\text{-}S_{10}^Q$ gene was purified by gel purification. The B-fragment was obtained by digesting the cloning vector coding for K_{12} with *BglII* and *BseRI* for 3 h at 37 $^\circ\text{C}$, followed by gel purification of the *BglII* \times *BseRI* fragment containing the K_{12} gene. Ligation of the A- and B-fragments using Quick ligase in 1 \times Quick ligase buffer for 5 min at room temperature resulted in the formation of the $E_{80}^S\text{-}S_{10}^Q\text{-}K_{12}$ gene. The ligation product was transformed into EBSAlpha chemically competent cells, and the cells were plated on TBdry plates supplemented with 45 $\mu\text{g}/\text{mL}$ kanamycin. The sequence was confirmed by DNA sequencing. Apart from the methionine start codon, the exact amino acid sequence of the encoded polypeptide is $(\text{GSGVP})_{80}(\text{GAGAGAGQ})_{10}\text{GK}_{12}\text{G}$, and this corresponds to an expected molecular weight of 39 159.97 Da.

Expression and Purification. To express the triblock polypeptide $E_{80}^S\text{-}S_{10}^Q\text{-}K_{12}$, the plasmid as described above was transformed into *E. coli* BL21(DE3) chemically competent cells, which were then used to inoculate a 10 mL starter culture of TB medium supplemented with 45 $\mu\text{g}/\text{mL}$ kanamycin. The starter culture was incubated overnight at 37 $^\circ\text{C}$ on a shaker at 200 rpm. Next, the starter culture was used to inoculate a culture of 2 L TB medium supplemented with 45 $\mu\text{g}/\text{mL}$ kanamycin. Cells were grown for a total time of 24 h at 37 $^\circ\text{C}$ on a shaker at 200 rpm. To induce polypeptide expression, isopropyl β -D-1-thiogalactopyranoside at a final concentration of 1 mM was added to the medium 8 h after the inoculation. After polypeptide expression, cells were pelleted by centrifugation at 2000g for 10 min at 10 $^\circ\text{C}$ and resuspended in 25 mL of 20 mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (HEPES) (pH 8.0). The cells were lysed by once pressing through a cooled French press at 1.5 bar. The lysate was centrifuged at 29 000g for 12 min at 4 $^\circ\text{C}$ to pellet insoluble cellular debris, and the supernatant (soluble lysate) was collected for further purification. To remove nucleic acids, the soluble lysate (~30 mL) was mixed with 4 mL of 10% (w/v) poly(ethylenimine) and centrifuged at 29 000g for 12 min at 4 $^\circ\text{C}$. The polypeptide in the remaining supernatant (soluble cleared lysate) was further purified exploiting both the thermal properties of the elastin-like block (E_{80}^S) by using inverse transition cycling (ITC),¹³ and the strong positive charge of the oligolysine block (K_{12}) in cation-exchange chromatography.

For the cation-exchange chromatography, the soluble cleared lysate was loaded onto a UNO S6 column (Bio-Rad) connected to a BioLogic DuoFlow chromatography system supplied with a QuadTec detector from Bio-Rad. Unbound molecules were removed from the column by washing with four column volumes of 20 mM HEPES buffer (pH 8.0). Bound molecules were eluted by using a salt gradient

from 0.0 to 1.0 M of NaCl in a 20 mM HEPES buffer (pH 8.0) applied over 10 column volumes. During elution, 1 mL fractions of the samples were collected from the column. Fractions containing the polypeptide were first dialyzed against Milli-Q water and then subjected to one round of ITC using the following protocol: (1) aggregation of the ELP-containing polypeptide was induced at room temperature by the addition of 2.5–3.0 M ammonium sulfate to the dialyzed sample, followed by incubation at 37 °C for 15 min. (2) Sample was centrifuged at 40 °C for 12 min at 29 000g to pellet the aggregated polypeptides. (3) Polypeptides in the pellet were solubilized in 50 mM HEPES buffer (pH 8.0). To facilitate solubilization, samples were vortexed at 1000 rpm for 90 min at room temperature. (4) To remove any remaining insoluble matter, the samples were first cooled on ice and then centrifuged at 4 °C for 12 min at 29 000g. Finally, all polypeptides were dialyzed against Milli-Q water, lyophilized, and stored at room temperature.

Characterization. Purity of the polypeptides was assessed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). SDS-PAGE was carried out using 10% mini-protean TGX precast protein gels, 1× Laemmli running buffer and Precision Plus Protein All Blue Prestained Protein Standard, and gels were stained with PageBlue protein staining solution. For MALDI-TOF MS, a matrix solution was prepared of 15 mg/mL 2,5-dihydroxyacetophenone and 4.5 mg/mL diammonium hydrogen citrate in 75% ethanol. Next, 1 volume of this matrix solution was mixed with 1 volume of 1 mg/mL polypeptide and 1 volume of 3% (v/v) trifluoroacetic acid, and 1 μ L of this mixture was dried on an 800 μ m spot of an MTP AnchorChip 384 target (Bruker). Analysis was then carried out on an UltrafleXtreme mass spectrometer (Bruker).

Atomic Force Microscopy (AFM). Polypeptide stock solutions of 100 μ M were prepared in a 10 mM phosphate buffer (pH 7.4). To promote solubilization of the freeze-dried polypeptides, they were incubated at 65 °C for 1 h. This temperature is high enough to disrupt aggregates and assemblies of the polypeptides, but not so high as to cause irreversible changes to their assembly after cooling down.¹⁷ Next, polypeptides were further diluted in 10 mM phosphate buffer (pH 7.4) to a concentration of 1.8 μ M and, if so indicated, mixed with DNA at a final DNA concentration of 1 μ g/mL, which corresponds to a polypeptide-to-DNA molar charge ratio N/P of 7. All samples were incubated for 24 h at room temperature to allow formation of complexes. For atomic force microscopy (AFM), 5 μ L of a sample that was prepared as described in the previous section was deposited on a clean silicon surface and incubated for 2 min. Next, the surface was rinsed with 1 mL of Milli-Q water to remove salts and nonadsorbed particles and dried slowly under a gentle N₂ stream. Samples were analyzed on a NanoScope MultiMode 8 system (Bruker) in the ScanAsyst (PeakForce Tapping) imaging mode, using ScanAsyst-Air cantilevers (Bruker). Areas of 5 \times 5 μ m were scanned at 512 \times 512 pixels and a line rate of 0.977 Hz. All images were subjected to a second-order flattening using NanoScope Analysis 1.40 software. If so indicated, lengths of fibrils were measured using FiberApp software.¹⁸ Settings used for FiberApp are: $\alpha = 0$, $\beta = 500$, $\gamma = 20$, $\kappa_1 = 20$, $\kappa_2 = 10$, $\text{step} = 1$ pixel, iterations = 100, "Use A* pathfinding algorithm". Weight-averaged fibril lengths were calculated assuming the mass of the fibrils was proportional to their length. The standard error of weighted means (sem_w) was approximated using the method of block averages. Data were subdivided into five subsets, weight-averaged lengths were calculated for the subsets, and the standard error of the weighted means was taken to be the standard deviation of the weight-averaged lengths for the subsets.

Gel Shift Assays. To quantify binding of the $E_{80}^S-S_{10}^Q-K_{12}$ triblock polypeptide to DNA, linear dsDNA (2000 bp) was incubated at room temperature for 2 h with increasing concentrations of the polypeptide. To that purpose, a stock solution of 100 μ M polypeptide in 10 mM phosphate buffer (pH 7.4) was prepared in the same way as for AFM. The stock solution was then diluted to the desired polypeptide concentrations and mixed with DNA. The DNA concentration was

always 15 ng/ μ L, and the polypeptide concentrations were 0.00, 1.54, 2.31, 3.08, 3.85, 5.77, 7.69, and 26.92 μ M. The corresponding N/P ratios were 0, 0.4, 0.6, 0.8, 1.0, 1.5, 2.0, and 7.0. For each mixture, an aliquot containing 52.5 ng of DNA was loaded onto a 1% agarose gel that was supplemented with 1× SYBR Safe DNA stain. After electrophoresis for 90 min at 60 V, the gel was scanned using a Bio-Rad Gel Doc EZ Imager.

RESULTS AND DISCUSSION

Cloning, Expression, and Purification. The plasmid encoding the triblock polypeptide $E_{80}^S-S_{10}^Q-K_{12}$ was constructed by recursive directional ligation by plasmid reconstruction (Pre-RDL), as described before.¹⁵ Figure 1a,b shows a schematic picture of the final $E_{80}^S-S_{10}^Q-K_{12}$ plasmid and a schematic representation of the structure of the corresponding polypeptide when co-assembled with DNA.

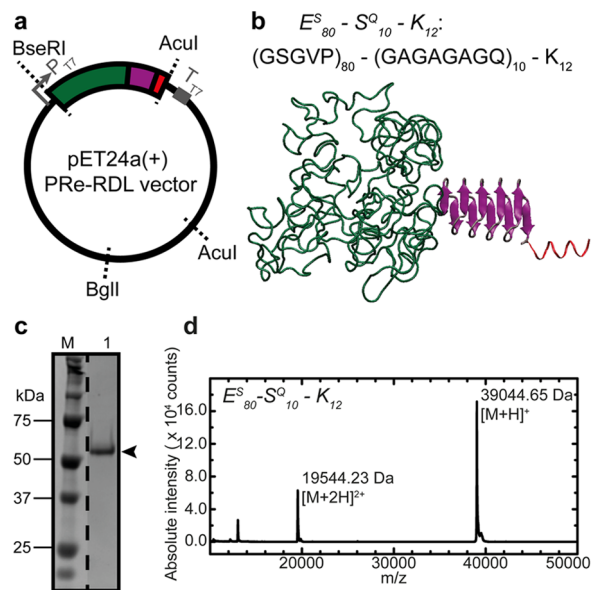


Figure 1. Expression and characterization of the $E_{80}^S-S_{10}^Q-K_{12}$ triblock polypeptide. (a) Pre-RDL plasmid for the expression of the $E_{80}^S-S_{10}^Q-K_{12}$ triblock polypeptide. The elastin-like polypeptide block E_{80}^S is shown in green, the silklike S_{10}^Q midblock in purple, and the oligolysine K_{12} binding block in red. (b) Cartoon of the structure of the triblock polypeptide when co-assembled with DNA. Colors as in (a). Silklike midblock is shown in β -solenoid configuration, which it supposedly adopts after fibrillar assembly onto DNA.^{10,11} Also given is the amino acid sequence of the triblock polypeptide. (c) SDS-PAGE of purified $E_{80}^S-S_{10}^Q-K_{12}$; the arrowhead indicates the polypeptide. Lane M is a molecular weight marker. Gel was stained using PageBlue. (d) MALDI-TOF MS analysis of the polypeptide.

The $E_{80}^S-S_{10}^Q-K_{12}$ triblocks were purified using cation-exchange chromatography followed by inverse transition cycling (ITC).¹³ Purity and molar mass of the newly purified polypeptides were confirmed by SDS-PAGE and MALDI-TOF MS (Figure 1c,d). Note that, due to the hydrophilic random coil nature of the large amorphous blocks, the triblocks have an anomalously low mobility in SDS-PAGE compared to the SDS-PAGE molecular weight standards, exactly as has been observed before for the $C-S_{10}^Q-K_{12}$ triblocks.⁷

The experimental mass, as determined from MALDI-TOF MS, is approximately 100 Da lower than expected if we assume cleavage of the N-terminal methionine.¹⁹ Possibly, also the subsequent glycine (\sim 75 Da) has been cleaved off, but this or

other small modifications are not expected to strongly affect the overall behavior of the polypeptides.

Influence of the Nature of the Amorphous Block on DNA-Binding Properties. To assess the impact of the nature of the amorphous block on the DNA-binding properties of the artificial viral capsid polypeptides, an electrophoretic mobility gel shift assay (Figure 2) was performed and compared to the

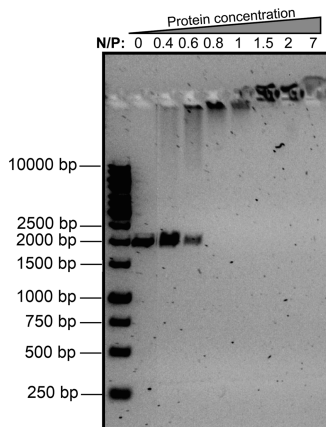


Figure 2. Gel shift assay of $E^S_{80}-S^Q_{10}-K_{12}$. The DNA (linear dsDNA, 2000 bp) concentration was kept constant at $15 \text{ ng}/\mu\text{L}$, and the DNA was incubated with increasing concentrations of the $E^S_{80}-S^Q_{10}-K_{12}$ triblock. The polypeptide-to-DNA molar charge ratios (N/P) are indicated above each lane, and the sizes of the DNA marker lane are labeled on the left. The total amount of DNA loaded into each well is 52.5 ng.

earlier results for the $C-S^Q_{10}-K_{12}$ triblocks.⁷ The gel shift assay shown in Figure 2 clearly demonstrates that binding of $E^S_{80}-S^Q_{10}-K_{12}$ is highly cooperative with high-mobility complexes (naked DNA; linear dsDNA, 2000 bp) coexisting, with low-mobility complexes (encapsulated DNA). The shift to low-mobility (encapsulated) DNA was observed for polypeptide-

to-DNA charge ratios N/P (#nitrogens in polypeptides per #phosphates in DNA) larger than ~ 0.6 . At high N/P ratio, the complexes are observed to run slightly toward the anode. While the observation of highly cooperative binding is similar to that for the $C-S^Q_{10}-K_{12}$ triblocks, for the latter (under similar conditions), encapsulated DNA was only observed for N/P larger than ~ 3 .⁷ Knowing that the amorphous block itself is not directly involved in binding to the DNA, this hints at a significant indirect influence of the nature of the amorphous block on complex formation.

Influence of the Nature of the Amorphous Block on VLP Size Distributions. We hypothesized that the indirect influence of the nature of the amorphous block on the DNA-binding properties of the artificial viral capsid polypeptides occurs through influencing the assembly of the silklike midblocks into fibrils: it is the fibril formation of the silklike midblock that mediates binding cooperativity.⁷ Therefore, we next investigated the co-assembly of $E^S_{80}-S^Q_{10}-K_{12}$ with DNA by AFM and compared it with the corresponding data for $C-S^Q_{10}-K_{12}$. Note that in earlier papers we have shown that VLP lengths obtained using AFM on dried complexes are consistent with those obtained using other methods, in particular, cryo-electron microscopy and static light scattering,⁷ as well as AFM in liquid.²⁰

Assembly and AFM imaging were performed at a polypeptide concentration of $1.8 \mu\text{M}$ and an excess of polypeptide over DNA (charge ratio N/P = 7). The AFM images of $E^S_{80}-S^Q_{10}-K_{12}$ and $C-S^Q_{10}-K_{12}$ in the absence and presence of a 2000 bp linear DNA template are shown in Figure 3.

Previously, we found⁷ that for $C-S^Q_{10}-K_{12}$, the length of fibrils co-assembled with DNA was equal to the contour length of the DNA template, divided by a factor of approximately 3 and which was called the “packing factor”. For a 2000 bp template, the DNA contour length divided by a packing factor of 3 equals approximately 230 nm. For the $C-S^Q_{10}-K_{12}$ polypeptide, fibrils with this length are only found in the

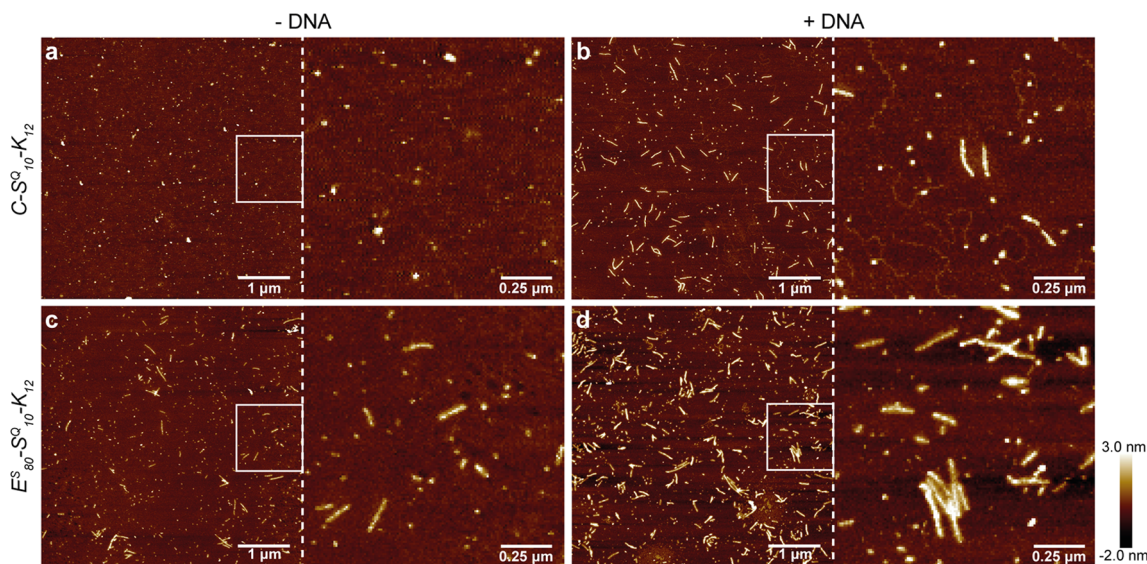


Figure 3. Self-assembly and co-assembly of $C-S^Q_{10}-K_{12}$ and $E^S_{80}-S^Q_{10}-K_{12}$ triblocks with and without a DNA template. Triblocks were solubilized at $65 \text{ }^\circ\text{C}$, diluted to a concentration of $1.8 \mu\text{M}$, and incubated at room temperature for 24 h in the absence or presence of linear dsDNA of 2000 bp. (a) $C-S^Q_{10}-K_{12}$ polypeptide incubated in the absence of DNA. (b) $C-S^Q_{10}-K_{12}$ polypeptide incubated in the presence of DNA. (c) $E^S_{80}-S^Q_{10}-K_{12}$ polypeptide incubated in the absence of DNA. (d) $E^S_{80}-S^Q_{10}-K_{12}$ polypeptide incubated in the presence of DNA. In each of the four conditions, the right panel is a digital magnification to highlight the formed fibrils.

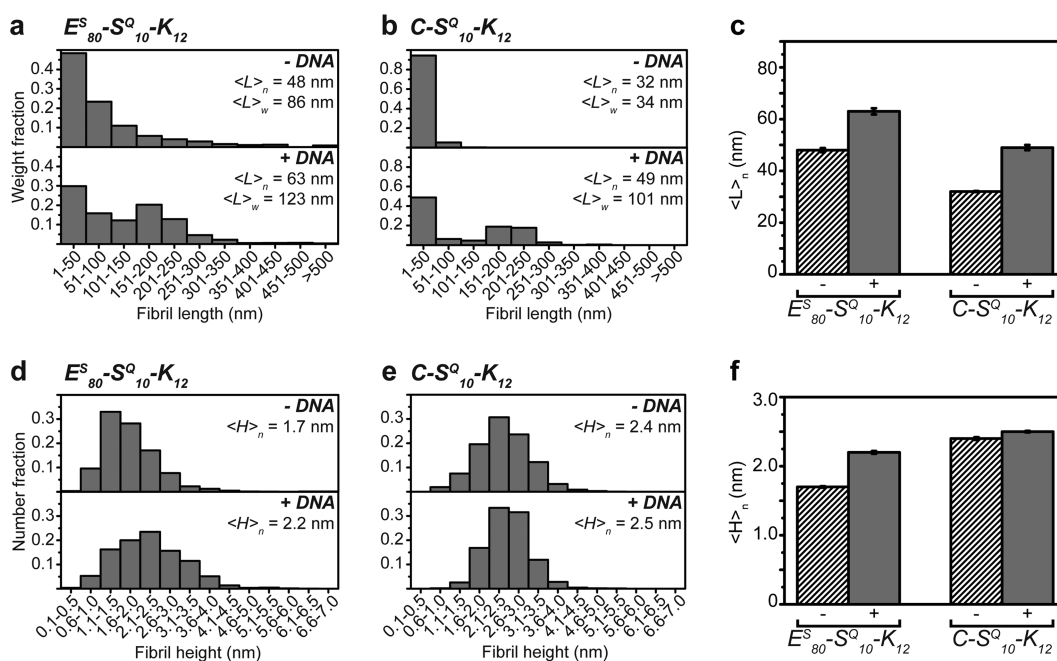


Figure 4. Length and height distributions of fibrils formed by $C-S^Q_{10}-K_{12}$ and $E^S_{80}-S^Q_{10}-K_{12}$ triblocks in the presence and absence of linear dsDNA (2000 bp), as derived from the AFM images. Representative AFM images are shown in Figure 3. Length distributions are shown as the weight fraction of polypeptide incorporated into fibrils within a given length interval. Height distributions are shown as the number fraction of polypeptide within a given interval. Length distributions for (a) $E^S_{80}-S^Q_{10}-K_{12}$ and (b) $C-S^Q_{10}-K_{12}$, and height distributions for (d) $E^S_{80}-S^Q_{10}-K_{12}$ and (e) $C-S^Q_{10}-K_{12}$ are all split in two graphs, with the top graphs showing the distribution in the absence of DNA and the bottom graphs in the presence of DNA. Note that in (b) the y-axis scale differs from that in (a). (c) Number-averaged fibril lengths of both polypeptides summarized in a bar diagram and (f) number-averaged fibril heights of both polypeptides summarized in a bar diagram. The striped bars represent the samples without DNA, and the filled bars represent the samples with DNA. The error bars in (c) and (f) are \pm sem (see values in Table 1).

Table 1. Average Fibril Lengths and Heights for $E^S_{80}-S^Q_{10}-K_{12}$ and $C-S^Q_{10}-K_{12}$ Polypeptides Assembled with or without 2000 bp dsDNA, Obtained by Analyzing AFM Images^a

		<i>N</i>	$\langle L \rangle_n$ (nm)	$\langle L \rangle_w$ (nm)	$\langle H \rangle_n$ (nm)
$E^S_{80}-S^Q_{10}-K_{12}$	1.8 μ M	2551	48 (\pm 0.8)	86 (\pm 18.9)	1.7 (\pm 0.01)
	1.8 μ M + DNA	2443	63 (\pm 1.2)	123 (\pm 9.4)	2.2 (\pm 0.02)
$C-S^Q_{10}-K_{12}$	1.8 μ M	1478	32 (\pm 0.2)	34 (\pm 1.2)	2.4 (\pm 0.02)
	1.8 μ M + DNA	2708	49 (\pm 1.0)	101 (\pm 7.9)	2.5 (\pm 0.01)

^aAll averages (\pm sem) were calculated using the data obtained from AFM images.

presence of the DNA template (Figure 3a,b). Indeed, for $C-S^Q_{10}-K_{12}$, in the absence of DNA, fibrils are mostly shorter than 50 nm. In contrast, for $E^S_{80}-S^Q_{10}-K_{12}$, in the absence of DNA, fibrils with lengths higher than 100 nm are also formed (Figure 3c). In the presence of DNA, the number of fibrils longer than 100 nm drastically increases (Figure 3d), indicating DNA encapsulation, but the encapsulation is somewhat obscured by the presence of nontemplated fibrils as we will show in a more quantitative analysis below.

We performed a quantitative analysis of the AFM images by measuring both the lengths and heights of the fibrils. To that purpose, all fibrils in three or four AFM images of $5 \times 5 \mu$ m surfaces were measured per condition. Histograms of fibril lengths and heights are shown in Figure 4. Figure 4 also shows bar diagrams of the number-averaged lengths and heights for the fibrils. Average values of fibril lengths and heights, and the number of fibrils analyzed per condition are provided in Table 1. The histograms of fibril lengths are reported as the weight fraction of polypeptides incorporated in fibrils within a certain length range, viz., the weight of the incorporated polypeptides versus the total weight of the polypeptides in the condition. The fibrils have a very wide distribution of lengths; therefore,

both the number and weight-averaged lengths are reported in Table 1 ($\langle L \rangle_n$ and $\langle L \rangle_w$).

Confirming our qualitative analysis of the AFM images, the quantitative analysis shows that $C-S^Q_{10}-K_{12}$ indeed forms only short fibrils (with length < 50 nm) in the absence of DNA (Figure 4b, top) and that in the presence of the DNA template, longer fibrils with a length of 151–250 nm are enriched (Figure 4b, bottom; also see Figure 4c). For the $E^S_{80}-S^Q_{10}-K_{12}$ polypeptides, the length distribution is much broader in the absence of the template (Figure 4a, top); nevertheless, enrichment in longer fibrils is still observed in the presence of the DNA template (Figure 4a, bottom; also see Figure 4c). The fact that an enrichment in 151–250 nm fibrils is only observed in the presence of the DNA template indicates encapsulation of this template. For the $E^S_{80}-S^Q_{10}-K_{12}$ polypeptide, the peak corresponding to encapsulated DNA occurs at somewhat smaller lengths (151–200 nm) than the expected length of around 230 nm, corresponding to a packing factor of around 3. This could indicate that the packing factor for this polypeptide is somewhat higher than that for the $C-S^Q_{10}-K_{12}$ polypeptide. Further evidence of successful DNA encapsulation by the $E^S_{80}-S^Q_{10}-K_{12}$ polypeptide is provided by the height

distributions, which show a shift to thicker fibrils in the presence of the DNA template (Figure 4d–f, Table 1). Replacing the collagen-like hydrophilic block with the elastin-like block thus preserves the fundamental features of VLP formation with DNA.

However, the fibril length distributions for the two polypeptides (Figure 4a–c and Table 1) also highlight an important difference. Very clearly, without a template, the E_{80}^S - S_{10}^Q - K_{12} fibrils formed in the absence of DNA have a length distribution that tails off at much larger lengths than the C - S_{10}^Q - K_{12} fibrils. Indeed, at 1.8 μ M, the weight-averaged length in the absence of a template is $\langle L \rangle_w = 86 \pm 19$ nm for E_{80}^S - S_{10}^Q - K_{12} , whereas it is $\langle L \rangle_w = 34 \pm 1$ nm for C - S_{10}^Q - K_{12} . As a consequence, when co-assembling with DNA, the separation in length between fibrils encapsulating DNA and fibrils not encapsulating DNA is much more pronounced for the C - S_{10}^Q - K_{12} polypeptides (compare Figure 4a,b, bottom panels).

Origin of Differences in Self-Assembly is Caused by Differences in Amorphous Blocks. We here find that the amorphous block plays a crucial role in preventing non-templated fibril formation at low concentrations by opposing fibril formation at low concentrations. We also find that the efficiency by which the amorphous block prevents (non-templated) fibril formation at low concentrations critically depends on their chemical nature.

Both amorphous blocks studied here adopt a random coil conformation and have a length of about 400 amino acids, but the collagen-like block C much more strongly opposes fibril formation at low concentrations than the elastin-like block E_{80}^S . We speculate that the key difference between the two polypeptides that gives rise to this different behavior is their relative hydrophilicity. Increased hydrophilicity translates into larger random coil sizes of the amorphous blocks and stronger intermolecular repulsions between these more hydrated random coils that are densely packed along the assembled silk fibrils. This may be expected to lead to the observed prevention of nontemplated fibril formation at low concentrations.

Indeed, the collagen-like and elastin-like blocks have grand average of hydropathy (GRAVY) values²¹ that are significantly different from each other. The GRAVY for E_{80}^S is 0.200, indicating that this amorphous block has an intermediate hydrophilicity and is close to its theta point, the transition point between a polymer in good solvent and a polymer in a poor solvent, consistent with the well-known aqueous demixing phase behavior exhibited by elastin-like polypeptides.²² On the other hand, the collagen-like C polypeptide is extremely hydrophilic with a GRAVY value of -1.717 .

In future work, it would therefore be interesting to change the nature of the guest residue X in the elastin-like polypeptide consensus motif GXGVP to obtain a polypeptide with a hydrophilicity approaching that of the collagen-like C block. The most hydrophilic amino acids, however, are charged residues, which may interfere with the electrostatic interactions that drive template binding. Therefore, the most promising candidates would be zwitterionic elastin-like polypeptides, for example, with sequences as having recently been investigated.²³

CONCLUDING REMARKS

We have shown that the assembly- and co-assembly behaviors of the C - S_{10}^Q - K_{12} and E_{80}^S - S_{10}^Q - K_{12} triblock polypeptides are significantly different. We expect that differences between the chemical nature of amorphous blocks may similarly influence

the assembly of a broad range of natural silk polypeptides and engineered silklike polypeptides.

As the much more hydrophilic collagen-like polypeptide C is better able to prevent self-assembly into fibrils at low concentrations than the much less hydrophilic elastin-like polypeptide E_{80}^S , we hypothesize that hydrophilicity of the amorphous block is a key variable in determining the assembly of natural silk polypeptides and engineered silklike polypeptides. Indeed, a critical role for hydrophilicity in controlling self-assembly is reported for many other polypeptide block copolymers, with increased hydrophilicity invariably reducing self-assembly.^{24–26}

Such a role for the hydrophilicity of amorphous blocks in silklike polypeptides is also consistent with the observation that long hydrophilic blocks surrounding silklike self-assembly blocks may completely prevent fibril formation.²⁷

An important conclusion is therefore that the chemical nature of the amorphous blocks, in particular their hydrophilicity, plays an important role in determining the assembly of both natural silks and engineered silklike polypeptides. This also means that it would be very interesting to more systematically study a broader range of amorphous structural polypeptides as amorphous blocks in the context of silklike polypeptides, to better understand the mutual influence of crystallizable and amorphous blocks on self-assembly. These amorphous blocks could include either natural ones such as resilin and abductin^{28,29} or designed sequences such as PAS and XTEN.^{30,31}

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biomac.9b00512.

Expression of triblock polypeptide C - S_{10}^Q - K_{12} in *E. coli*: SDS-PAGE and Western blot (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: renko.devries@wur.nl.

ORCID

Enrico Mastrobattista: 0000-0002-6745-2015

Renko de Vries: 0000-0001-8664-3135

Author Contributions

L.W. designed, biosynthesized, and characterized the polypeptides, and performed AFM experiments. S.R. and I.W. contributed to biosynthesis of the polypeptides. L.v.W. contributed to AFM experiments. C.C.C. and A.H.-G. performed the work described in the Supporting Information. A.C. advised on polypeptide design, E.M. and J.v.d.O. contributed to the design of the project, R.d.V. designed the project, L.W. and R.d.V. wrote the manuscript, and all authors have edited the manuscript and have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

L.W. and R.d.V. acknowledge funding by the Dutch Science Organisation (NWO), grant 712.014.003.

REFERENCES

- (1) Chow, D.; Nunalee, M. L.; Lim, D. W.; Simnick, A. J.; Chilkoti, A. Peptide-Based Biopolymers in Biomedicine and Biotechnology. *Mater. Sci. Eng., R* **2008**, *62*, 125–155.
- (2) Carter, N. A.; Geng, X.; Grove, T. Z. Design of Self-Assembling Protein-Polymer Conjugates. In *Protein-based Engineered Nanostructures*; Cortajarena, A.; Grove, T., Eds.; Springer, 2016; *Advances in Experimental Medicine and Biology*, Vol. 940; pp 179–214.
- (3) Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L. Protein-Based Block Copolymers. *Biomacromolecules* **2011**, *12*, 269–289.
- (4) Hardy, J. G.; Romer, L. M.; Scheibel, T. R. Polymeric Materials Based on Silk Proteins. *Polymer* **2008**, *49*, 4309–4327.
- (5) Vepari, C.; Kaplan, D. L. Silk as a Biomaterial. *Prog. Polym. Sci.* **2007**, *32*, 991–1007.
- (6) Hiew, S. H.; Miserez, A. Squid Sucker Ring Teeth: Multiscale Structure-Property Relationships, Sequencing, and Protein Engineering of a Thermoplastic Biopolymer. *ACS Biomater. Sci. Eng.* **2017**, *3*, 680–693.
- (7) Hernandez-Garcia, A.; Kraft, D. J.; Janssen, A. F. J.; Bomans, P. H. H.; Sommerdijk, N. A. J. M.; Thies-Weesie, D. M. E.; Favretto, M. E.; Brock, R.; de Wolf, F. A.; Wertén, M. W. T.; van der Schoot, P.; Cohen Stuart, M.; de Vries, R. Design and Self-Assembly of Simple Coat Proteins for Artificial Viruses. *Nat. Nanotechnol.* **2014**, *9*, 698–702.
- (8) Schor, M.; Martens, A.; DeWolf, F. A.; Cohen Stuart, M.; Bolhuis, P. G. Prediction of Solvent Dependent β -Roll Formation of a Self-Assembling Silk-like Protein Domain. *Soft Matter* **2009**, *5*, 2658–2665.
- (9) Schor, M.; Bolhuis, P. G. The Self-Assembly Mechanism of Fibril-Forming Silk-Based Block Copolymers. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10457–10467.
- (10) Zhao, B.; Cohen Stuart, M. A.; Hall, C. K. Dock 'n Roll: Folding of a Silk-Inspired Polypeptide into an Amyloid-like Beta Solenoid. *Soft Matter* **2016**, *12*, 3721–3729.
- (11) Zhao, B.; Cohen Stuart, M. A.; Hall, C. K. Navigating in Foldonia: Using Accelerated Molecular Dynamics to Explore Stability, Unfolding and Self-Healing of the β -Solenoid Structure Formed by a Silk-like Polypeptide. *PLoS Comput. Biol.* **2017**, *13*, No. e1005446.
- (12) Wertén, M. W. T.; Wisselink, W. H.; Jansen-van den Bosch, T. J.; de Bruin, E. C.; de Wolf, F. Secreted Production of a Custom-Designed, Highly Hydrophilic Gelatin in *Pichia Pastoris*. *Protein Eng., Des. Sel.* **2001**, *14*, 447–454.
- (13) Hassouneh, W.; Christensen, T.; Chilkoti, A. Elastin-like Polypeptides as a Purification Tag for Recombinant Proteins. In *Current Protocols in Protein Science*; John Wiley & Sons Inc.: Hoboken, NJ, 2010; pp 6.11.1–6.11.16.
- (14) Schipperus, R.; Eggink, G.; De Wolf, F. A. Secretion of Elastin-like Polypeptides with Different Transition Temperatures by *Pichia Pastoris*. *Biotechnol. Prog.* **2012**, *28*, 242–247.
- (15) McDaniel, J. R.; MacKay, J. A.; Quiroz, F. G.; Chilkoti, A. Recursive Directional Ligation by Plasmid Reconstruction Allows Rapid and Seamless Cloning of Oligomeric Genes. *Biomacromolecules* **2010**, *11*, 944–952.
- (16) Willems, L.; Roberts, S.; Weitzhandler, I.; Chilkoti, A.; Mastrobattista, E.; van der Oost, J.; de Vries, R. Inducible Fibril Formation of Silk–Elastin Diblocks. *ACS Omega* **2019**, *4*, 9135–9143.
- (17) Cazares-Vargas, E.; Cohen Stuart, M. A.; de Vries, R.; Hernandez-Garcia, A. Template-Free Self-Assembly of Artificial De Novo Viral Coat Proteins into Nanorods: Effects of Sequence, Concentration, and Temperature. *Chem. - Eur. J.* **2019**, *25*, 11058–11065.
- (18) Usov, I.; Mezzenga, R. FiberApp: An Open-Source Software for Tracking and Analyzing Polymers, Filaments, Biomacromolecules, and Fibrous Objects. *Macromolecules* **2015**, *48*, 1269–1280.
- (19) Hirel, P. H.; Schmitter, M. J.; Dessen, P.; Fayat, G.; Blanquet, S. Extent of N-Terminal Methionine Excision from *Escherichia Coli* Proteins Is Governed by the Side-Chain Length of the Penultimate Amino Acid. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8247–8251.
- (20) Marchetti, M.; Kamsma, D.; Cazares-Vargas, E.; Hernandez-Garcia, A.; van der Schoot, P.; de Vries, R.; Wuite, G. J. L.; Roos, W. H. Real-time assembly of virus-like nucleocapsids elucidated at the single-particle level. *Nano Lett.* **2019**, 5746.
- (21) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydrophobic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105–132.
- (22) Hassouneh, W.; Zhulina, E. B.; Chilkoti, A.; Rubinstein, M. Elastin-like Polypeptide Diblock Copolymers Self-Assemble into Weak Micelles. *Macromolecules* **2015**, *48*, 4183–4195.
- (23) Banskota, S.; Yousefpour, P.; Kirmani, N.; Li, X.; Chilkoti, A. Long Circulating Genetically Encoded Intrinsically Disordered Zwitterionic Polypeptides for Drug Delivery. *Biomaterials* **2019**, *192*, 475–485.
- (24) Trabbic-Carlson, K.; Meyer, D. E.; Liu, L.; Piervincenzi, R.; Nath, N.; LaBean, T.; Chilkoti, A. Effect of Protein Fusion on the Transition Temperature of an Environmentally Responsive Elastin-like Polypeptide: A Role for Surface Hydrophobicity? *Protein Eng., Des. Sel.* **2004**, *17*, 57–66.
- (25) Christensen, T.; Hassouneh, W.; Trabbic-Carlson, K.; Chilkoti, A. Predicting Transition Temperatures of Elastin-like Polypeptide Fusion Proteins. *Biomacromolecules* **2013**, *14*, 1514–1519.
- (26) Weitzhandler, I.; Dzuricky, M.; Hoffmann, I.; Garcia Quiroz, F.; Gradzielski, M.; Chilkoti, A. Micellar Self-Assembly of Recombinant Resilin-/Elastin-Like Block Copolypeptides. *Biomacromolecules* **2017**, *18*, 2419–2426.
- (27) Beun, L. H.; Storm, I. M.; Wertén, M. W. T.; De Wolf, F. A.; Cohen Stuart, M. A.; De Vries, R. From Micelles to Fibers: Balancing Self-Assembling and Random Coiling Domains in pH-Responsive Silk-Collagen-like Protein-Based Polymers. *Biomacromolecules* **2014**, *15*, 3349–3357.
- (28) Roberts, S.; Dzuricky, M.; Chilkoti, A. Elastin-like Polypeptides as Models of Intrinsically Disordered Proteins. *FEBS Lett.* **2015**, *589*, 2477–2486.
- (29) Rauscher, S.; Pomès, R. Structural Disorder and Protein Elasticity. In *Fuzziness: Structural Disorder in Protein Complexes*; Fuxreiter, M.; Tompa, P., Eds.; Springer: New York, NY, 2012; Vol. 1, pp 59–183.
- (30) Schlapschy, M.; Binder, U.; Börger, C.; Theobald, I.; Wachinger, K.; Kisling, S.; Haller, D.; Skerra, A. PASylation: A Biological Alternative to PEGylation for Extending the Plasma Half-Life of Pharmaceutically Active Proteins. *Protein Eng., Des. Sel.* **2013**, *26*, 489–501.
- (31) Schellenberger, V.; Wang, C. W.; Geething, N. C.; Spink, B. J.; Campbell, A.; To, W.; Scholle, M. D.; Yin, Y.; Yao, Y.; Bogin, O.; et al. A Recombinant Polypeptide Extends the in Vivo Half-Life of Peptides and Proteins in a Tunable Manner. *Nat. Biotechnol.* **2009**, *27*, 1186–1190.

Nanoscopic resolution within a single imaging frame

Esley Torres García ^{1,2}, Raúl Pinto Cámara ², Alejandro Linares ², Damián Martínez ², Víctor Abonza ², Eduardo Brito-Alarcón ², **Carlos Calcines-Cruz** ³, Gustavo Valdés ⁴, David Torres ², Martina Jabłoński ⁵, Héctor Torres ⁶, José Luis Martínez ⁷, Haydee Hernández ², Joseph Ocelotl ², Yasel Garcés ^{2,7}, Joseph Dubrovsky ⁶, Alberto Darszon ⁷, Mariano G. Buffone ⁵, Roberto Rodríguez Morales ⁸, Juan Manuel Rendon-Mancha ¹ Christopher D. Wood ², Armando Hernández García ⁴, Diego Krapf ⁹, Álvaro H. Crevenna ¹⁰, and Adan Guerrero ^{2,*}.

¹ Centro de Investigación en Ciencias, Instituto de Investigación en Ciencias Básicas y Aplicadas, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, México.

² Laboratorio Nacional de Microscopía Avanzada, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México.

³ Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México. Ciudad de México, México.

⁴ Departamento de Química de Biomacromoléculas, Instituto de Química. Universidad Nacional Autónoma de México. Ciudad de México, México.

⁵ Instituto de Biología y Medicina Experimental (IBYME-CONICET), Buenos Aires, Argentina.

⁶ Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, México.

⁷ Departamento de Genética del Desarrollo y Fisiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México. Cuernavaca, Morelos, México.

⁸ Instituto de Cibernética, Matemática y Física, Ciudad de la Habana, Cuba.

⁹ Electrical and Computer Engineering and School of Biomedical Engineering, Colorado State University

¹⁰ European Molecular Biology Laboratory, Neurobiology and Epigenetics Unit, Monterotondo, Italy.

* Correspondence contact: adan.guerrero@ibt.unam.mx