



Universidad de Granada  
Departamento de Bioquímica y Biología molecular 3  
e Inmunología

**“Estudio de la señal de localización y  
comportamiento dinámico del factor de  
transcripción y *splicing* TCERG1”**

Noemí Sánchez Hernández  
IPBLN, CSIC  
Granada, 2013

Editor: Editorial de la Universidad de Granada  
Autor: Noemí Sánchez Hernández  
D.L.: GR 1889-2013  
ISBN: 978-84-9028-607-4



CARLES M<sup>a</sup> SUÑÉ NEGRE, Doctor en Farmacia, Investigador del CSIC en el Departamento de Biología Molecular del Instituto de Parasitología y Biomedicina “López Neyra” (CSIC, Granada),

CERTIFICA:

que Noemí María Sánchez Hernández, licenciada en Biología por la Universidad de Granada, ha realizado bajo su dirección el trabajo titulado: “ESTUDIO DE LA SEÑAL DE LOCALIZACIÓN Y COMPORTAMIENTO DINÁMICO DEL FACTOR DE TRANSCRIPCIÓN Y *SPLICING* TCERG1”, que presenta como Memoria de Tesis para alcanzar el grado de Doctor por la Universidad de Granada y que en la realización del trabajo, se han respetado los derechos de otros autores al ser citados, cuando se han utilizado sus resultados o publicaciones.

Para que conste a efectos oportunos, expide y firma la presente en Granada a 15 de Febrero de 2013.

Fdo. Dr. Carles M<sup>a</sup> Suné Negre



# *Resumen*



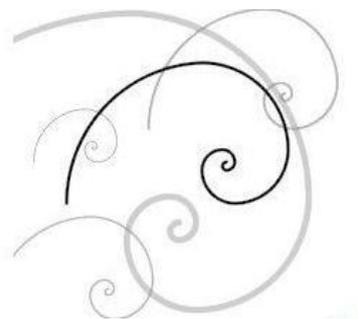


El núcleo de la célula eucariota es una estructura altamente compartimentalizada, donde el material genético se dispone a modo de territorios cromosómicos que dejan lugar a un gran número de subestructuras u orgánulos nucleares. A pesar del aparente estado estático de los orgánulos nucleares, sus proteínas están sometidas a un elevado dinamismo, probablemente debido a la falta de membranas que delimitan su estructura, característica que los diferencia de los orgánulos citoplasmáticos. En este movimiento proteico generalizado dentro del núcleo las señales de localización juegan un importante papel, pues a ellas se debe la correcta distribución de los componentes de cada uno de los diferentes orgánulos.

En esta tesis hemos identificado y caracterizado la señal responsable de la localización del factor de transcripción y *splicing* TCERG1 en la periferia de los *speckles* nucleares, orgánulos enriquecidos principalmente en factores de *splicing* como por ejemplo las proteínas SR o las partículas snRNP. Dentro de TCERG1, los dominios FF4/FF5 (aminoácidos 878 a 1021) son la señal necesaria y suficiente para la correcta localización de la proteína. Además estos dominios constituyen la señal suficiente para la localización de otras proteínas en los *speckles* nucleares. De esta manera, los dominios FF4/FF5 de TCERG1 representan una nueva señal de localización general hacia los *speckles* nucleares no identificada hasta el desarrollo de esta tesis. Nuestros estudios encaminados a conocer las bases moleculares en las que se sustenta el papel de los dominios FF4/FF5 como señal de localización hacia los *speckles* nucleares han identificado a la proteína Nopp140 como la posible responsable de esta localización. De acuerdo con el elevado dinamismo ya citado, TCERG1 presenta una alta movilidad en el núcleo, que se muestra prácticamente idéntica en los diferentes compartimentos estudiados (*speckles* nucleares y nucleoplasma). Este movimiento se muestra además independiente del estado transcripcional de la célula. Los estudios de FRAP (*fluorescence recovery after photobleaching*) realizados para determinar todos estos parámetros han sido empleados también para describir las propiedades dinámicas de TCERG1 en relación a las maquinarias de transcripción y *splicing*, aportando datos que apoyan el papel ya propuesto para TCERG1 en el acoplamiento entre ambas maquinarias.



# *Índice*





# ÍNDICE

<b>I. ABREVIATURAS</b>	<b>15</b>
<b>II. INTRODUCCIÓN</b>	<b>19</b>
<b>1. ARQUITECTURA FUNCIONAL EN EL NÚCLEO EUCARIOTA</b>	<b>21</b>
1.1 Introducción	21
1.2 Biogénesis de los compartimentos nucleares	26
1.3 Los orgánulos nucleares, estructuras estables altamente dinámicas. El principio de autoorganización	29
<b>2. LOS <i>SPECKLES</i> NUCLEARES</b>	<b>32</b>
2.1 Introducción	32
2.2 Composición de los <i>speckles</i> nucleares	33
2.3 Función de los <i>speckles</i> nucleares	37
<b>3. TCERG1, factor regulador de la elongación transcripcional (<i>Transcription elongation regulator 1</i>)</b>	<b>42</b>
3.1 Introducción	42
3.2 TCERG1 en el proceso de transcripción	45
3.3 TCERG1 en el proceso de <i>splicing</i> . Localización en los <i>speckles</i> nucleares	47
<b>4. Dominios FF y señalización hacia los <i>speckles</i> nucleares</b>	<b>48</b>
4.1 Dominios FF	48
4.1.1 Generalidades	48
4.1.2 Interacciones de los dominios FF	49
4.1.3 Estructura de los dominios FF	50
4.2 Dominios de señalización hacia los <i>speckles</i> nucleares	52
4.3 Mecanismos moleculares de localización hacia los <i>speckles</i> y otros orgánulos nucleares	55
4.3.1 Fosforilación	55
4.3.2 Otras modificaciones post-traduccionales: SUMOilación	58
<b>5. Dinámica de proteínas y ARN en el núcleo. Estudio de la expresión génica en célula viva</b>	<b>61</b>
5.1 Introducción	61
5.2 Dinámica espacial de la maquinaria de <i>splicing</i>	65
5.3 Estudio in vivo del ADN y ARN: dinámica espacial de la transcripción, el <i>splicing</i> y el acoplamiento entre ambos procesos	66
5.4 Limitaciones de los estudios in vivo; sin consenso en el modelo cinético	71
<b>III. OBJETIVOS</b>	<b>73</b>
<b>IV. MATERIALES Y MÉTODOS</b>	<b>77</b>
<b>1. PLÁSMIDOS Y ANTICUERPOS</b>	<b>79</b>
Plásmidos	79
Anticuerpos	80
<b>2. CULTIVOS CELULARES, TRANSFECCIÓN DE ADN Y TRATAMIENTOS</b>	<b>81</b>

<b>3. INMUNOFLUORESCENCIA, FISH, FRAP Y PROCEDIMIENTOS DE MICROSCOPIA</b>	<b>83</b>
Inmunofluorescencia y microscopía confocal	83
<i>In situ</i> fluorescence hybridization (FISH)	84
<i>Fluorescence recovery after photobleaching</i> (FRAP)	85
<b>4. PROCEDIMIENTOS BÁSICOS DE BIOLOGÍA MOLECULAR</b>	<b>87</b>
Extracción de ARN y análisis por RT-PCR (PCR reverse transcriptase)	87
<b>5. MANIPULACIÓN Y ANÁLISIS DE PROTEÍNAS</b>	<b>87</b>
Preparación de extractos celulares	87
Análisis por western blot	88
Producción y purificación de proteínas recombinantes en <i>E.coli</i>	88
Ensayos de <i>pull-down</i>	90
Tinción de plata	90
Tinción de coomasie	91
<b>6. OLIGONUCLEÓTIDOS</b>	<b>91</b>
<b>V. RESULTADOS</b>	<b>95</b>
<b>1. LOS DOMINIOS FF4 Y FF5 DE TCERG1 DIRIGEN PROTEÍNAS HACIA LA PERIFERIA DE LOS SPECKLES NUCLEARES</b>	<b>97</b>
1.1. <i>Identificación de la secuencia necesaria para la localización de TCERG1 en los speckles nucleares</i>	97
1.2. <i>Los dominios FF4 y FF5 de TCERG1 representan una nueva señal de localización hacia la periferia de los speckles nucleares</i>	105
1.3. <i>Implicación de los residuos altamente conservados de fenilalanina de los dominios FF4/FF5 de TCERG1 en la localización a los speckles nucleares</i>	111
1.4. <i>Papel de los dominios FF4/FF5 en el efecto de TCERG1 sobre el splicing alternativo</i>	114
1.5. <i>Estudio del mecanismo molecular implicado en la localización de TCERG1 en los speckles nucleares</i>	116
1.5.1 <i>El residuo K985, potencial aceptor de SUMO, no influye en la localización de TCERG1 a los speckles nucleares</i>	116
1.5.2 <i>La proteína nucleolar Nopp140 (NOLC1) interacciona in vitro con los dominios FF4/FF5 de TCERG1</i>	118
<b>2. DINÁMICA DE TCERG1 EN CÉLULA VIVA</b>	<b>121</b>
2.1 <i>Cinética de TCERG1 en los speckles nucleares y nucleoplasma en célula viva</i>	121
2.2 <i>El intercambio de TCERG1 entre los speckles nucleares y el nucleoplasma es independiente de la existencia de transcripción activa en la célula</i>	123
2.3 <i>TCERG1 se localiza en el sitio activo de transcripción del VIH-1 en célula viva</i>	129
2.4 <i>El reclutamiento de TCERG1 al SAT del VIH-1 es independiente del dominio CTD de la ARN polIII.</i>	131
2.5 <i>El splicing co-transcripcional modifica la dinámica de reclutamiento de TCERG1 hacia el SAT.</i>	133
<b>VI. CONCLUSIONES</b>	<b>139</b>

<b>VII. DISCUSIÓN</b>	<b>143</b>
La SUMOilación y TCERG1	146
TCERG1 podría comportarse como una proteína SR	147
Un posible papel para la unión ácidos nucleicos	148
TCERG1 podría localizar en el nucléolo antes de asociarse con los speckles nucleares	150
<b>VIII. BIBLIOGRAFÍA</b>	<b>153</b>
<b>IX. PUBLICACIONESS</b>	<b>163</b>



## ABREVIATURAS

5'ss: sitio 5' de *splicing*

### A

ADNr: ADN ribosomal

ARN poly(A): ARN poliadenilado

ARNlnc: ARN largo no codificante

ARNm: ARN mensajero

ARNnc: ARN no codificante

ARNr: ARN ribosomal

### B

bGH: hormona de crecimiento bovina

BP: branch point

BSA: albúmina de suero bovino

### C

ChIP: inmuprecipitación de cromatina

CTD: dominio carboxiterminal

CF: componente fibrilar

CFD: componente fibrilar denso

CG: componente granular

### D

DMEM: Dulbecco's Modified Eagle Medium

DRB: 5,6-dicloro-1-β-

ribofuranosilbenzimidazol

DS: desviación estándar

DTT: ditioneitol

### E

ECFP: enhanced cyan fluorescent protein

### F

F.I.: fracción inmóvil

F.M.: fracción móvil

FBS: suero bovino fetal

FCS: fluorescence correlation spectroscopy

FISH: in situ fluorescence hybridization

FLIP: fluorescence loss in photobleaching

FPS: fibras de pericromatina

FRAP: fluorescence recovery after photobleaching

FRET: fluorescence resonance energy transfer

fwd: forward

### G

GFP: green fluorescent protein

GST: glutatión-S-transferasa

### H

He-Ne: helio-neón

HLB: histone locus bodies

## I

IGCs: interchromatin granule cluster

IPTG: isopropil-1-tio- $\beta$ ,D-galactopiranosido

## K

kD: kilodalton

## L

Lac O: operon Lac

LacI: represor del Operon Lac

LTR: long terminal repeat

## M

M: peso molecular

## N

NDSM: negatively charged aminoacid-dependent sumoylation motif

NE: extracto nuclear

NLS: señal de localización nuclear

NMR: resonancia magnética nuclear

NP-40: nonidet fosfato-40

## P

PDSM: phosphorylation-dependent sumoylation motif

PFA: paraformaldehído

PMSF: fenilmetil-sulfonil-fluoruro

pre-ARNm: pre-ARN mensajero

pre-ARNr: pre-ARN ribosomal

P-TEFb: Positive Transcription Elongation Factor b

## R

rev: reverse

ROI: region of interest

RRM: RNA recognition motif

RS: dominio rico en dipéptidos de arginina y serina

RT-PCR: retro-transcription PCR

## S

SAT: sitio activo de transcripción del VIH-1

SDS: dodecil sulfato sódico

Ser2: serina 2

SFCs: *splicing* factor compartments

SIM: SUMO interaction motif

snRNPs: small nuclear ribonucleoproteins

snoRNPs: small nucleolar RNA

SUMO: Small Ubiquitin-like MOdifier

SUPERP: superposición

SSA: spliceostatina

## U

UTR: untranslated región

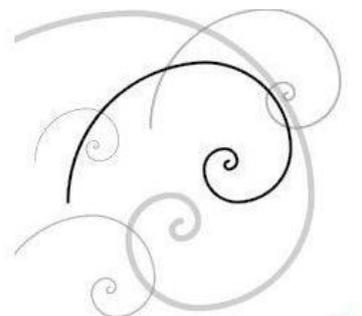
## V

VIH-1: virus de la inmunodeficiencia huma tipo 1

VRC: vanadyl ribonucleoside complexes



# *Introducción*

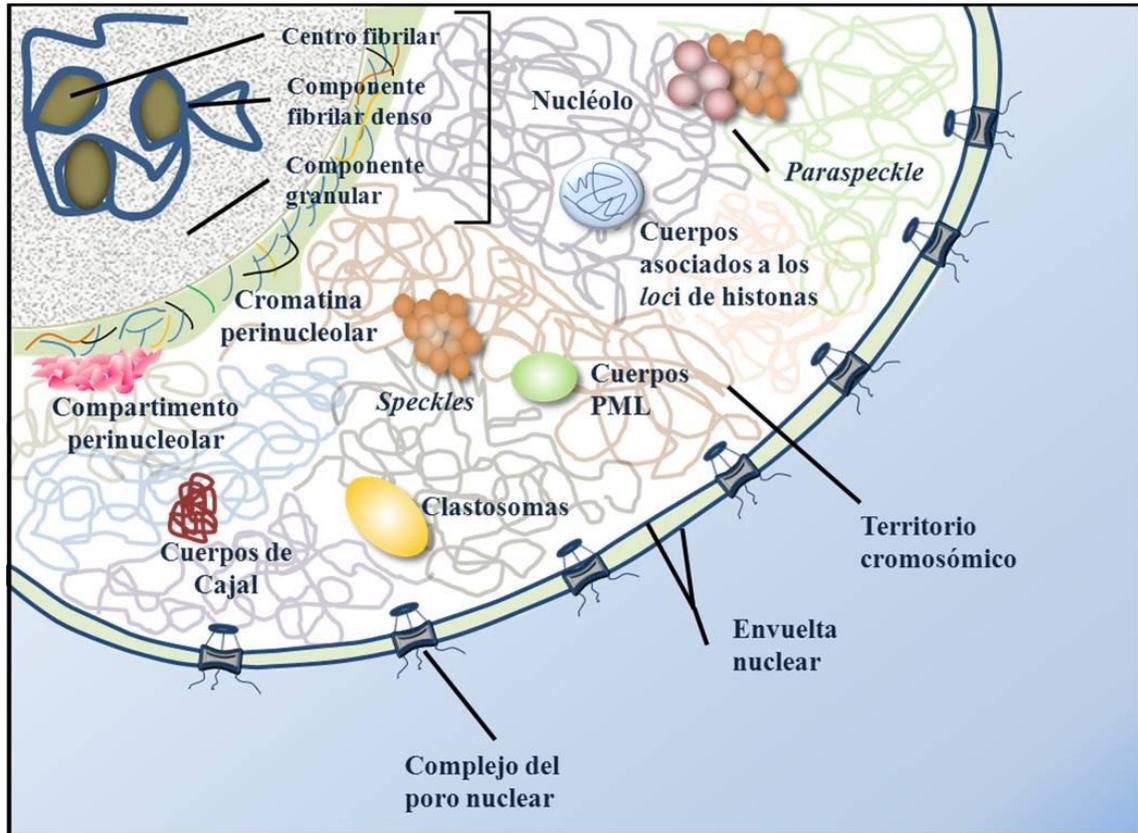




# 1. ARQUITECTURA FUNCIONAL EN EL NÚCLEO EUCARIOTA

## 1.1 Introducción

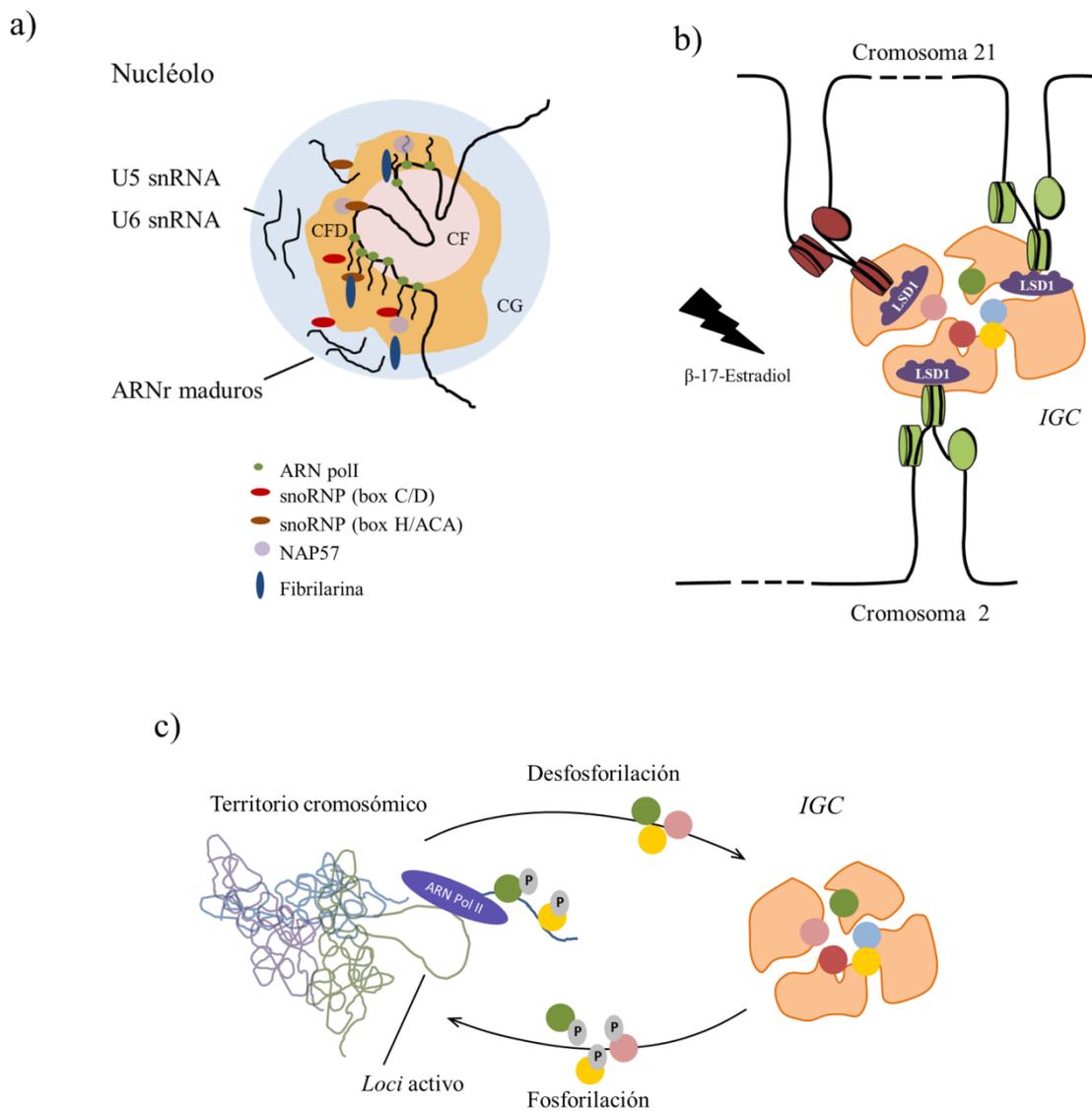
La estructura de la célula eucariota se caracteriza por la presencia de dos compartimentos claramente definidos: el citoplasma y el núcleo. Mientras que en el citoplasma los diferentes procesos metabólicos tienen lugar en dominios u orgánulos perfectamente delimitados por una membrana (p.ej. la mitocondria, el aparato de Golgi, los peroxisomas, etc.), los orgánulos nucleares no presentan ninguna estructura que separe su contenido del resto del ambiente nucleoplasmático, lo cual apunta hacia mecanismos de biogénesis y mantenimiento de estos compartimentos diferentes a aquellos que caracterizan a los de los orgánulos citoplasmáticos. A pesar de la falta de membranas que delimiten los orgánulos nucleares, el núcleo de la célula eucariota es una estructura altamente compartimentalizada, donde el material genético se dispone a modo de dominios discretos llamados territorios cromosómicos que dejan lugar a un gran número de subestructuras, orgánulos o compartimentos nucleares en el espacio intercromatínico (Figura I-1 y Tabla I-1). Entre toda esta elevada organización espacial tienen lugar procesos tan esenciales como la transcripción, el procesamiento de las moléculas de ARN, la replicación y la reparación del ADN. Los orgánulos nucleares pueden contribuir a la regulación de estas funciones de manera diferente (Figura I-2). En algunos casos pueden ser el lugar específico donde se concentran los factores necesarios para llevar a cabo una función biológica específica. Éste sería claramente el caso del nucléolo, donde se concentran todos los genes ribosomales y la mayoría de los factores de procesamiento para la transcripción y procesamiento del pre-ARN ribosomal (pre-ARNr) (Figura I-2, a). En otros casos, los orgánulos nucleares son estructuras funcionalmente inactivas que ejercen su función de una forma indirecta, actuando como sitios de almacenamiento, modificación y/o regulación de la concentración de factores disponibles en la región nucleoplasmática que los rodea (Figura I-2, c). Como veremos más adelante, esta regulación podría ser el caso de los denominados *speckles* nucleares, orgánulos enriquecidos en factores de procesamiento de las moléculas de ARN mensajero (ARNm) donde sin embargo no tiene lugar el proceso de *splicing*.



**Figura I-1. Diversidad de los orgánulos nucleares.** La imagen representa el núcleo interfásico de una célula de mamífero. El núcleo es delimitado por una doble membrana (envuelta nuclear), que se extiende hacia el citoplasma formando el retículo endoplasmático. Una de sus principales funciones es la de servir de barrera para separar el contenido nuclear del citoplasma. La envoltura nuclear es interrumpida a lo largo de su extensión por los poros nucleares encargados de controlar el transporte nucleocitoplasmático. La cara interna de la envuelta nuclear proporciona soporte mecánico para la organización de la cromatina. El núcleo contiene la gran mayoría del material génico de la célula, organizado en múltiples cromosomas. Los cromosomas en interfase se sitúan ocupando los denominados territorios cromosómicos, donde las fibras de cromatina de uno o diferentes territorios contactan y se entremezclan estableciendo múltiples uniones. La imagen representa la alta compartimentalización y organización del núcleo a modo de numerosos orgánulos o dominios nucleares. (Imagen adaptada de (Mao, Zhang et al. 2011).

Los compartimentos nucleares pueden actuar también como centros de regulación de la expresión, activando o reprimiendo determinados genes a través del reclutamiento de *loci* específicos (Figura I-2, b). Probablemente un mismo orgánulo nuclear combina varias de estas funciones. Una visión algo más simplista sobre la función de los orgánulos nucleares propone que éstos podrían ser el resultado de la tendencia de ciertas proteínas a formar agregados macromoleculares. De hecho, muchas de sus proteínas

mayoritarias tienen la capacidad de establecer uniones directas entre ellas a través de dominios de autointeracción (Hebert and Matera 2000). De acuerdo con esta idea se ha observado la formación *de novo* de algunos orgánulos, indistinguibles de los orgánulos endógenos, tras la sobre-expresión de proteínas nucleares (Shpargel, Ospina et al. 2003; Kaiser, Intine et al. 2008). Sin embargo, y en contra de la idea de que los orgánulos nucleares son simples agregados proteicos, se ha demostrado que la sobre-expresión de las proteínas que definen determinados orgánulos no induce la formación de orgánulos aberrantes ni produce un aumento en el número o tamaño del orgánulo correspondiente (Matera, Izaguirre-Sierra et al. 2009).



**Figura I-2. Posibles modos de actuación de los orgánulos nucleares.** Representación esquemática de los posibles mecanismos de actuación de los orgánulos nucleares. Ciertos orgánulos combinan varios de estos mecanismos para llevar a cabo una función determinada. **(a)** Los orgánulos nucleares pueden ser el lugar donde se lleva a cabo una función específica albergando una gran concentración de proteínas. Por ejemplo, el nucléolo funciona como una factoría donde se sintetizan y procesan los ARN ribosómicos (ARNr). La imagen muestra alguna de las proteínas implicadas en la síntesis y procesamiento de los ARNr; ARN polII encargada de la transcripción a partir de los genes que codifican para el ARN ribosomal (ARNr), localizados exclusivamente en el componente fibrilar (CF), snoRNAs (*small nucleolar RNA*) y proteínas como NAP57 y la fibrilarina encargadas del procesamiento de los pre-ARNr. La síntesis de los ARNr comienza en el CF y se extiende de forma radial por el componente fibrilar denso (CFD) y el componente granular (CG) desde donde son liberados para formar parte del proceso de síntesis ribosomal. Dentro del nucléolo también se han identificado componentes del *spliceosoma*, como los U5 y U6 snRNAs, donde se sabe que sufren procesos de modificación **(b)** Algunos orgánulos pueden funcionar como centros de regulación. El estímulo desencadenado por estrógenos (17- $\beta$ -estradiol) conduce a la interacción de los genes *TFF1* y *GREB1*, ambos regulados por los receptores de estrógenos y situados en los cromosomas 21 y 2 respectivamente, en un mismo *speckle* nuclear. Esta reorganización es dependiente del factor LDS1, una demetilasa de lisinas de histonas, ya que su silenciamiento impide esta reorganización. La convergencia de estos genes en un mismo *speckle* nuclear podría potenciar una expresión más eficiente ante determinados estímulos. Además se permitiría un acoplamiento entre los procesos de inicio de la transcripción, elongación y procesamiento de varias unidades génicas a la vez permitiendo una respuesta más rápida (Hu, Kwon et al. 2008) **(c)** Almacenamiento, modificación y/o regulación de la concentración de factores en las regiones nucleoplasmáticas cercanas. Algunos orgánulos nucleares pueden almacenar, modificar y/o reciclar proteínas y moléculas de ARN. Los *speckles* nucleares modifican, a través de las quinasas que contienen, los niveles de fosforilación de las proteínas SR (ver más adelante en apartado 4.3.1). Estos eventos de fosforilación son requeridos para el reclutamiento de los factores de *splicing* hasta el sitio donde está teniendo lugar de forma activa la transcripción (Gama-Carvalho, Krauss et al. 1997; Misteli, Caceres et al. 1998), donde se ha descrito la interacción de las proteínas SR y otros factores como TCERG1 con el dominio CTD de la ARN polII (Yuryev, Patturajan et al. 1996; Carty, Goldstrohm et al. 2000).

**Tabla I-1. Características básicas de los orgánulos nucleares**

Orgánulo	Nº por célula	Tamaño (µm)	Componentes Mayoritarios	Funciones
<b>Nucléolo</b>	1-4	0.5-8	ARN polII/Factores de procesamiento de ARNr/Factores de ensamblaje de los ribosomas	Implicado en la transcripción y procesamiento de los ARNr y en el ensamblaje de las unidades ribosomales. Participa también en la modificación y ensamblaje de otros ARNs celulares así como diferentes partículas ribonucleoproteicas. Regula la progresión del ciclo celular ejerciendo funciones de secuestro y modificación de diferentes proteínas.
<b>Compartimento perinucleolar</b>	1-4	0.2-1	PTB, CUGBP	Sus funciones concretas no se conocen hasta el momento, pero su presencia se correlaciona positivamente con la capacidad de metástasis de la célula. Se le relaciona con la regulación post-transcripcional de transcritos derivados de la ARN polIII.
<b>Speckles nucleares</b>	20-50	0.8-3	Diferentes factores de <i>splicing</i> como SRSF1 (SF2/ASF), SRSF2 (SC35). Malat1	Típicamente se les relaciona con funciones de almacenamiento, ensamblaje y modificación de factores de <i>splicing</i> , aunque como se describe en el texto sus funciones podrían ser otras.
<b>Paraspeckles</b>	2-20	0.2-1	PSP1, p54nrb, NONO, Men ε (Neat1)/ Menβ	Implicado en la retención nuclear de algunos ARNs sometidos a <i>editing</i> (cambio de adenosina por inosina).
<b>Cuerpos de Cajal</b>	0-10	0.1-2	Coilina, SMN	Implicados en el ensamblaje, modificación y tráfico de los snRNAs y snoRNAs. Pueden desempeñar un papel importante en el ensamblaje de las telomerasas y regulación del tamaño de los telómeros.
<b>Cuerpos PML</b>	10-30	0.1-1	PML, Sp100	Regulan la respuesta a diferentes formas de estrés. Implicados también en defensa viral y regulación de la estabilidad genómica mediante el secuestro, modificación y degradación de diferentes proteínas.
<b>Cuerpos de estrés nuclear</b>	2-10	0.3-3	HSF1, HAP	Su función general puede ser la regulación de la transcripción y el <i>splicing</i> bajo condiciones de estrés. Regula la expresión de ARN no codificante (ARNnc) a partir del ADN satélite III.
<b>Cuerpos asociados a los loci de histonas (HLB)</b>	2-4	0.2-1.2	NPAT, FLASH, U7snRNP	Participan en la transcripción y el procesamiento de los ARNm de los genes de histonas.
<b>Clastosomas</b>	0-3	0.2-1.2	Subunidades 19S y 20S del proteosoma	Se forman en respuesta a estímulos que activan la proteólisis dependiente del proteosoma.

## 1.2 Biogénesis de los compartimentos nucleares

Hasta el momento son dos los modelos de ensamblaje propuestos para la formación de los compartimentos nucleares; i) el modelo de ensamblaje estocástico y ii) el modelo de ensamblaje ordenado (Figura I-3). Ambos proponen el reclutamiento de subunidades o complejos macromoleculares a partir del nucleoplasma, difiriendo en su mecanismo de formación. El modelo de ensamblaje estocástico (Figura I-3, a) defiende que los compartimentos nucleares se forman a partir de las interacciones individuales aleatorias de sus componentes. Se trata de un proceso al azar en el que los diferentes componentes se ensamblan sin un orden predeterminado, existiendo múltiples posibilidades de ensamblaje para un mismo compartimento nuclear y en el que cada uno de los componentes tiene las mismas probabilidades de ser reclutado en primer lugar. El modelo de ensamblaje estocástico viene apoyado por estudios en los que el anclaje de cualquier componente de los cuerpos de Cajal (en inglés, *Cajal bodies*) en una localización específica del genoma puede dar lugar a un orgánulo de este tipo. Mediante la integración en tándem de secuencias del operon Lac (*Lac O*) (ver apartado 5.3; Sistema *LacO/LacI*) en el genoma de células HeLa, Dundr y colaboradores (Kaiser, Intine et al. 2008) demostraron que cualquier componente, proteína o ARN, de los cuerpos de Cajal puede promover la formación de uno de estos orgánulos con tamaño, forma, composición y dinámica de disociación de sus componentes similar a la de los orgánulos endógenos. Los resultados de este estudio demuestran además que los cuerpos de Cajal pueden formarse *de novo*, característica que también presentan otros orgánulos como el nucléolo, los *speckles* nucleares o los cuerpos PML.

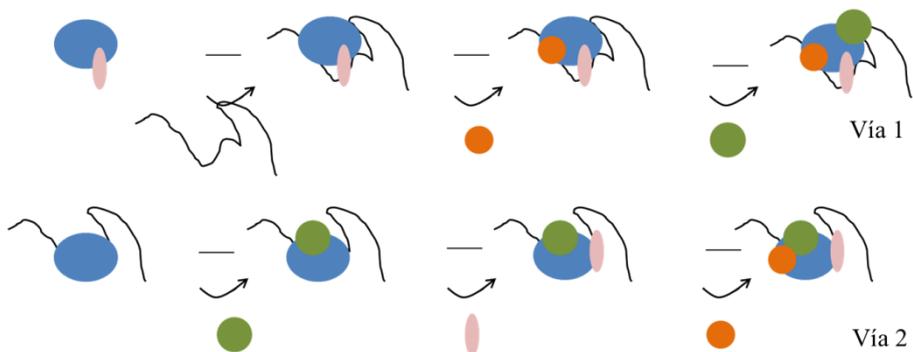
El modelo de ensamblaje ordenado (Figura I-3, b) propone un reclutamiento ordenado a partir de un determinado componente que actúa de “anclaje” para el resto del dominio. Es decir, dentro de todos los componentes que conforman el orgánulo existe uno o varios que son jerárquicamente diferentes del resto y cuya presencia es estrictamente necesaria para el posterior ensamblaje de los demás componentes. Éstos podrán ser reclutados posteriormente al azar o siguiendo un orden determinado. Usando una estrategia similar a la descrita anteriormente (sistema *LacO/LacI*), se ha demostrado que orgánulos como los *paraspeckles* no se forman siguiendo el modelo estocástico expuesto (Mao, Sunwoo et al. 2011). Tras la inmovilización de la proteína PSP1, alguno de los componentes de los *paraspeckles* son reclutados al sitio de anclaje en el genoma, aunque otros componentes proteicos así como las moléculas de ARN largo no

codificante (ARNlnc) MEN- $\epsilon$  (NEAT1) y MEN-  $\beta$  no llegan a reclutarse, haciendo que el complejo finalmente formado no sea funcional. Se demuestra, al menos para este tipo de orgánulos, que su formación no tiene lugar de manera azarosa a partir de cualquiera de sus componentes y que es estrictamente necesario un evento “nucleador” previo para la organización del orgánulo.

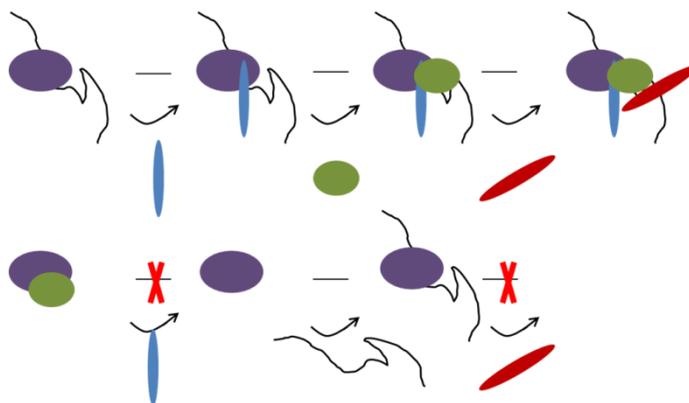
En este modelo cobra una gran importancia la idea de un elemento “semilla” sobre el que se ensamblan los demás componentes. Y dentro de este concepto, las moléculas de ARN parecen cobrar una gran importancia como elementos estructurales esenciales sobre los que se organizan los orgánulos nucleares. De hecho, durante mucho tiempo se ha sugerido un posible papel de moléculas de ARN no codificantes (ARNnc) en la arquitectura nuclear, basado en estudios pioneros que demuestran la retención de moléculas de ARN poliadeniladas (ARN poly(A)) en los *speckles* nucleares tras la inhibición del proceso de transcripción (Huang and Spector 1996). Estudios algo más recientes, acerca del modelo de ensamblaje de los *paraspeckles*, defienden también un papel esencial de ciertas moléculas de ARNnc. Los primeros estudios que evidenciaron la necesidad de este tipo de ARNs para la formación y mantenimiento de los *paraspeckles* demostraron que la permeabilización y posterior tratamiento con ARNasa provocaban la desaparición de estos orgánulos (Fox, Bond et al. 2005; Prasanth, Prasanth et al. 2005). Posteriormente, diferentes estudios han demostrado que las moléculas de ARNlnc MEN- $\epsilon$  (NEAT1) y MEN-  $\beta$  son elementos estructurales críticos para la formación y estabilidad de los *paraspeckles*, ya que su depleción provoca la desaparición de estas estructuras (Chen and Carmichael 2009; Clemson, Hutchinson et al. 2009; Sasaki, Ideue et al. 2009; Sunwoo, Dinger et al. 2009). Por último, sistemas de visualización en célula viva han demostrado que el ARNlnc es esencial para iniciar el ensamblaje de los *paraspeckles* y que su transcripción es capaz de mediar el reclutamiento de los principales componentes de estos orgánulos (PSP1, PSP2, p54nrb, PSF). Mediante estudios de FRAP (*fluorescence recovery after photobleaching*) (ver apartado 5), se ha demostrado que la dinámica de los ARNlnc (Mao, Sunwoo et al. 2011) es sustancialmente diferente a la de las proteínas principales del orgánulo. El tiempo de permanencia de MEN- $\epsilon$  (NEAT1) y MEN-  $\beta$  en los *paraspeckles*, medido mediante FRAP, es consistente con un papel estructural y organizador que permita el reclutamiento de los demás factores del orgánulo.

Además de este papel como elemento esencial en la arquitectura de un orgánulo determinado, el ARN también actúa como molde en los denominados “orgánulos dependientes de actividad”. Éstos se forman alrededor de zonas transcripcionalmente activas como son el ADN ribosomal (ADNr) dando lugar al nucléolo, los genes de las histonas dando lugar a los cuerpos asociados a las histonas, HLB (de *histone locus bodies*) o el ADN tipo satélite III, formando los orgánulos de estrés. Los transcritos generados en estos *loci* actúan como molde para atraer y retener proteínas de interacción con ARN, lo que conduce a la formación del orgánulo correspondiente.

(a) Modelo de ensamblaje estocástico



(b) Modelo de ensamblaje ordenado



**Figura I-3. Modelos de ensamblaje de los compartimentos nucleares.** (a) Modelo estocástico. La formación de los orgánulos nucleares puede tener lugar de manera estocástica, mediante la interacción de sus componentes de forma aleatoria a través de múltiples vías. (b) Modelo de ensamblaje ordenado. En este modelo, una única vía número es posible para la formación del compartimento nuclear. Una molécula determinada de ARN o proteína funciona como elemento de “anclaje” a partir del cual se va a estructurar de manera ordenada o aleatoria el resto del orgánulo. En el caso de que este elemento no esté presente algunos componentes del orgánulo podrán ser incorporados aunque el complejo final no será completamente funcional.

### **1.3 Los orgánulos nucleares, estructuras estables altamente dinámicas. El principio de autoorganización**

A pesar de su aparente estabilidad global, los orgánulos nucleares son estructuras altamente dinámicas (Phair and Misteli 2000; Handwerger and Gall 2006). La razón de esta paradoja puede encontrarse precisamente en la ausencia de membranas que, como han demostrado numerosos experimentos de FRAP, permite un intercambio libre de los componentes de los orgánulos nucleares con el medio nucleoplasmático que los rodea (Kruhlak, Lever et al. 2000; Phair and Misteli 2000; Snaar, Wiesmeijer et al. 2000; Chen and Huang 2001). La coexistencia de esta estabilidad y dinamismo representa una gran ventaja funcional para el núcleo. La naturaleza dinámica de sus interacciones permite a los orgánulos nucleares mantener una estructura estable a la vez que, mediante pequeños cambios en el comportamiento de sus constituyentes, pueden alterar su morfología cuando sea necesario. De este modo los orgánulos nucleares son capaces de dar rápidamente respuesta a cualquier tipo de señalización o condición de estrés. Esta alta movilidad proteica generalizada y la generación de estructuras estables, que resultan de las interacciones transitorias de los componentes de los orgánulos nucleares, son consistentes con la existencia del principio de autoorganización dentro del núcleo. De este modo e independientemente de la vía de ensamblaje que sigan, los orgánulos nucleares se mantienen en base al denominado principio de autoorganización (Misteli 2001), diferente del concepto de autoensamblaje. Éstos son los dos mecanismos descritos para el mantenimiento de las estructuras macromoleculares. Mientras que el autoensamblaje implica la asociación física permanente de las diferentes moléculas para formar un complejo estático que permanece en equilibrio (Kushner 1969), la autoorganización se basa en las interacciones transitorias de los componentes del complejo para dar lugar a una aparente estructura estática que sin embargo sufre un elevado grado de dinamismo (Nicolis 1977). En un contexto biológico, la autoorganización se define como la capacidad de los complejos macromoleculares u orgánulos de definir su propia estructura en base a las interacciones funcionales de sus componentes al ser sometidos a diversas condiciones (Misteli 2001). Como se ha comentado anteriormente, es precisamente la alta movilidad a la que se ven sometidos los orgánulos nucleares uno de los principales indicadores de la existencia de la autoorganización. El dinamismo que sufren los componentes de los orgánulos les permite establecer rápidas y múltiples interacciones hasta alcanzar la estructura más

estable. Otros indicadores de la existencia de autoorganización es la presencia de síntesis *de novo* de manera estocástica así como un tamaño limitado independiente de la concentración de los componentes del orgánulo (Dundr and Misteli 2001). El nucléolo representa quizás uno de los ejemplos más claros de orgánulo nuclear gobernado por el principio de autoorganización. Sus proteínas constitutivas se mueven continuamente desde y hacia el nucleoplasma generando un compartimento estable a pesar de la alta movilidad de sus componentes. Al inicio de la mitosis, la transcripción del ADNr es inhibida por la fosforilación, dependiente de mitosis, de la ARN polimerasa I y otros componentes de procesamiento de las moléculas de ARNr. Experimentos en célula viva demuestran que estos factores difunden libremente por el nucleoplasma de células en mitosis en busca de sitios de unión mientras la síntesis ribosomal se encuentra inhibida. Una vez que la célula se divide y la transcripción ribosomal se reanuda, los componentes nucleolares se acumulan progresivamente, y de manera aleatoria, en los sitios de transcripción de los ARNr, volviendo a establecer uniones transitorias entre ellos y determinando de nuevo la morfología del nucléolo (Dundr, Misteli et al. 2000). De igual modo, el comportamiento de los *speckles* nucleares es consistente con el principio de autoorganización. En concreto, tras la inhibición de la transcripción, los *speckles* nucleares ven reducido su número y además, su morfología sufre un cambio muy característico, pasando a ser más redondeados, más expandidos (O'Keefe, Mayeda et al. 1994; Sanchez-Alvarez, Goldstrohm et al. 2006). En cambio, ante un aumento de la expresión de genes que contienen un elevado número de intrones o ante un mayor nivel de transcripción ocasionado por una infección viral, la concentración de factores de *splicing* en los *speckles* se ve reducida (Lamond and Spector 2003). Estos cambios pueden ser explicados en base al principio de autoorganización. Los factores de *splicing*, dependiendo de la disponibilidad de sitios específicos de unión en los transcritos recién sintetizados, interaccionan de manera más o menos frecuente con los demás componentes de los *speckles* nucleares provocando así la restructuración del orgánulo.

Las bases moleculares sobre las que se sustenta el principio de autoorganización pueden encontrarse en los ya citados dominios de autointeracción entre las proteínas constitutivas de los orgánulos nucleares. Por ejemplo, la proteína mayoritaria de los cuerpos de Cajal, la coilina, contiene un dominio de autointeracción en su región aminoterminal (Hebert and Matera 2000). De igual modo, este tipo de dominios se pueden encontrar en las proteínas PML y SMN, constituyentes de los cuerpos PML y cuerpos de Cajal respectivamente. En los *paraspeckles*, sus proteínas constitutivas son capaces de formar homo y heteroligómeros y asociarse con moléculas de ARN. La eliminación de estos motivos proteicos de autointeracción da lugar a la desaparición del dominio (Bond and Fox 2009). En los *speckles* nucleares, las proteínas de la familia SR, factores de *splicing* altamente enriquecidos en estos dominios, se caracterizan por contener un número variable de dominios RRM (*RNA recognition motif*) en su región aminoterminal y un dominio tipo RS (dominio rico en dipéptidos de arginina y serina) en la región carboxiterminal. Éste último promueve estas interacciones proteína-proteína. Por su parte, el nucléolo contiene un elevado número de proteínas altamente cargadas, lo cual facilita la autointeracción entre ellas (Misteli 2001). Todas las uniones detectadas entre los componentes de los orgánulos nucleares a través de los dominios proteicos descritos son de tipo no covalente y débiles, característica totalmente necesaria para permitir el rápido movimiento de estas proteínas (Dundr and Misteli 2010).

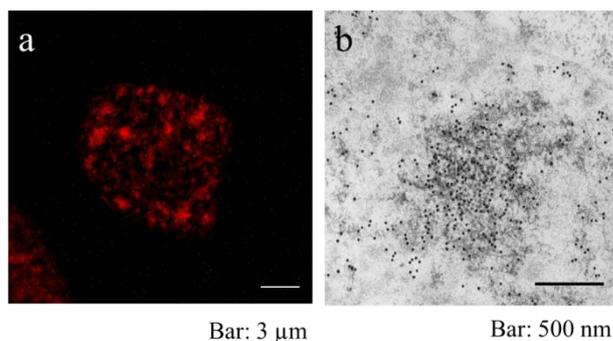
## 2. LOS *SPECKLES* NUCLEARES

### 2.1 Introducción

La visualización mediante microscopía óptica o microscopía de fluorescencia de los factores de procesamiento del pre-ARN mensajero (pre-ARNm) en el núcleo de la célula eucariota muestra una distribución no homogénea, con un patrón punteado de acumulación a modo de 10-30 dominios irregulares denominados *speckles* nucleares, compartimentos de factores de *splicing* (SFCs, de *splicing factor compartments*) o dominios SC35. Este último término se utiliza para hacer referencia al marcaje casi exclusivo de estructuras coincidentes con estos dominios que se obtiene al utilizar un anticuerpo monoclonal que reconoce específicamente un fosfoepítipo del factor esencial de *splicing* SC35 (Spector, Fu et al. 1991) (Figura I-4, a). Con un nivel mayor de resolución mediante microscopía electrónica (Figura I-4 b), se pueden observar dentro de cada uno de estos dominios dos estructuras bien diferenciadas: i) acúmulos de gránulos intercromatínicos (en adelante IGCs, de *interchromatin granule cluster*), estructuras más o menos esféricas, de naturaleza ribonucleoprotéica, que presentan un diámetro aproximado de 20-30 nm y se localizan en regiones desprovistas de cromatina (espacio intercromatínico) en el núcleo interfásico de células de organismos eucariotas superiores, y ii) fibras de pericromatina (FPs), con un tamaño medio de 3-5 nm y localizadas tanto en la periferia de los IGCs como en otras regiones nucleoplasmáticas, correspondientes a zonas donde localizan moléculas de pre-ARNm recién sintetizadas (Fakan 1994; Cmarko, Verschure et al. 1999; Fakan and van Driel 2007).

Los *speckles* nucleares fueron identificados por primera vez usando anticuerpos específicos frente a las partículas ribonucleoproteicas snRNPs (*small nuclear ribonucleoproteins*) (Lerner, Boyle et al. 1981; Spector, Schrier et al. 1983), principales componentes de la maquinaria de *splicing* o *spliceosoma*. Esta observación sentó las bases de una posible implicación de los *speckles* nucleares en el proceso de *splicing* de los pre-ARNm. El enriquecimiento en los *speckles* nucleares de la práctica totalidad de los factores de *splicing*, que a su vez demuestran una distribución secundaria difusa por el nucleoplasma (Lamond and Spector 2003), junto con la presencia de moléculas de ARNm poly(A) (Carter, Taneja et al. 1991; Carter, Bowman et al. 1993; Visa, Puvion-Dutilleul et al. 1993; Huang, Deerinck et al. 1994), corroboran la relación entre estos compartimentos y el metabolismo de las moléculas de pre-ARNm. Estudios de

proteómica (Mintz, Patterson et al. 1999; Saitoh, Spahr et al. 2004) y de visualización de célula viva (Phair and Misteli 2000) reafirman una marcada concentración en estos compartimentos en factores implicados en los procesos de transcripción y las diferentes etapas del procesamiento de las moléculas de ARNm, especialmente en eventos de *splicing*.



**Figura I-4. *Speckles* nucleares.** (a) Imagen obtenida mediante inmunofluorescencia en células HEK293T empleando anticuerpos específicos frente al factor esencial de *splicing* SC35. La distribución de este factor queda casi exclusivamente restringida a los *speckles* nucleares con una escasa tinción en el nucleoplasma. (b) Los *speckles* nucleares son equivalente a los *interchromatin granule cluster (IGC)*. Mediante microscopía electrónica se puede observar con detalle la acumulación de los gránulos intercromatínicos (20-30 nm) en los espacios desprovistos de cromatina. (Imagen tomada de (Spector and Lamond 2011)).

## 2.2 Composición de los *speckles* nucleares

La población más prominente dentro de los *speckles* nucleares se corresponde a los diferentes factores de *splicing*, incluyendo las partículas snRNPs y proteínas SR, llegando a constituir aproximadamente un 50% de la composición proteica de estos dominios (Saitoh, Spahr et al. 2004). Además, dentro de los *speckles* nucleares se ha descrito la presencia de diferentes quinasas, como CrkRS, Clk/Sty, SRPK1, DYRK1, CDKL5 y caseína I (Gui, Lane et al. 1994; Colwill, Pawson et al. 1996; Gross, Loijens et al. 1999; Alvarez, Estivill et al. 2003; Ricciardi, Kilstrup-Nielsen et al. 2009), y fosfatasas, como PP1 (Trinkle-Mulcahy, Ajuh et al. 1999), capaces de fosforilar y desfosforilar a los componentes de la maquinaria de *splicing*. La presencia de estas quinasas y fosfatasas en los *speckles* nucleares apoyaría un posible papel de estos

dominios en la regulación de la disponibilidad de los factores de *splicing* para las maquinarias de transcripción y procesamiento de los ARNm en el nucleoplasma circundante (ver apartado 4.3.1). En relación a esta disponibilidad de factores de *splicing* en el sitio de transcripción activa cabe destacar la presencia y papel regulador del complejo formado por el proteosoma 20S y el regulador PA28 $\gamma$  (Baldin, Militello et al. 2008).

Aunque diferentes estudios basados en la incorporación de nucleótidos marcados (uridinatriada o bromo-UTP) demuestran la ausencia de actividad transcripcional dentro de los *speckles* nucleares (Pombo and Cook 1996; Cmarko, Verschure et al. 1999; Xie, Martin et al. 2006), se ha podido demostrar que estos dominios albergan una pequeña población estable de ARN polimerasa II (ARNpolII) fosforilada en el residuo de serina 2 (Ser2) de su dominio carboxiterminal (CTD). La fosforilación de la Ser2 es la responsable de la conversión de la ARN polII en una forma productiva capaz de mediar una elongación altamente procesiva (Marshall and Price 1995; Marshall, Peng et al. 1996; Peterlin and Price 2006). La fosforilación en Ser2 se correlaciona, por tanto, con la etapa de la elongación transcripcional. A pesar de la fosforilación en Ser2, estas polimerasas de los *speckles* se consideran transcripcionalmente inactivas y para ellas se propone un papel en un potencial pre-ensamblaje del *spliceosoma* y eventos post-transcripcionales en el interior de los *speckles* nucleares (ver a continuación, apartado 2.3) (Xie, Martin et al. 2006). Otro factor implicado en transcripción y presente en los *speckles* nucleares es el complejo P-TEFb (Positive Transcription Elongation Factor b) (Herrmann and Mancini 2001; Dow, Liu et al. 2010), formado por la quinasa Cdk9 y la ciclina T1, implicado en la etapa de elongación transcripcional a través de la fosforilación de la Ser2 del dominio CTD de la ARNpolII. Aunque este complejo se distribuye de manera uniforme en el nucleoplasma, la ciclina T1 muestra una colocalización parcial muy significativa con los *speckles* nucleares. La quinasa Cdk9, aunque se localiza en regiones cercanas a estos dominios, muestra sin embargo un grado de colocalización muy reducido con ellos (Matera and Ward 1993; Herrmann and Mancini 2001). Los factores de elongación Brd4 y 7SK/HEXIM1, que acompañan al complejo P-TEFb, también se encuentran presentes en los *speckles* nucleares (Prasanth, Prasanth et al. 2005; Dow, Liu et al. 2010). Estos dominios muestran también un marcado enriquecimiento en factores implicados en otras etapas del procesamiento de los ARNm como proteínas de empaquetamiento, transporte, corte y poliadenilación

(Krause, Fakan et al. 1994; Schul, van Driel et al. 1998; Herrmann and Mancini 2001; Saitoh, Spahr et al. 2004; Schmidt, Richter et al. 2006).

Además de los factores de transcripción y procesamiento de los pre-ARNm, los *speckles* nucleares muestran un enriquecimiento de moléculas de ARNm poly(A), a pesar de la ya mencionada ausencia de actividad transcripcional dentro de ellos (Carter, Taneja et al. 1991; Carter, Bowman et al. 1993; Visa, Puvion-Dutilleul et al. 1993; Huang, Deerinck et al. 1994). Esta población de ARN no es transportada al citoplasma cuando la transcripción global de la célula es bloqueada, lo cual sugiere que se trata de una población estable de ARN que podría participar en procesos estructurales o reguladores (Huang, Deerinck et al. 1994). Apoyando esta hipótesis, estudios adicionales muestran que tras la sobre-expresión de la quinasa Clk/Sty y posterior desensamblaje de los *speckles* nucleares, las moléculas de ARNm poly(A) presentes en ellos se distribuyen por el nucleoplasma sin llegar a abandonar el núcleo para su posterior traducción (Sacco-Bubulya and Spector 2002). La idea de un papel estructural para alguna molécula de ARN es muy atractiva y análoga a lo que ocurre en los *paraspeckles* donde las moléculas de ARNinc MEN- $\epsilon$  NEAT1 y MEN-  $\beta$  proporcionan una clara función organizadora (Bond and Fox 2009; Chen and Carmichael 2009; Clemson, Hutchinson et al. 2009; Sasaki, Ideue et al. 2009; Sunwoo, Dinger et al. 2009). En este sentido el ARNinc MALAT1 (*metastasis-associated lung adenocarcinoma transcript 1*) se encuentra altamente enriquecido en los *speckles* nucleares (Hutchinson, Ensminger et al. 2007), aunque no se ha descrito para él por el momento una función estructural en estos dominios.

Contradiendo este posible papel estructural, estudios en célula viva han demostrado que la amplia mayoría de la población de ARNm poly(A) se internaliza y difunde libremente a través de los *speckles* nucleares sin acumularse significativamente dentro de ellos, algo que se esperaría si estas moléculas estuviesen actuando como andamiaje del orgánulo (Politz, Tuft et al. 1999; Molenaar, Abdulle et al. 2004; Politz, Tuft et al. 2006). En este contexto hay que destacar que las moléculas de ARNm poly(A) no son necesarias para el correcto ensamblaje de los *speckles* que tiene lugar en fase G1, tras la salida de la fase de mitosis (Ferreira, Carmo-Fonseca et al. 1994; Gama-Carvalho, Krauss et al. 1997). Como veremos más adelante, se ha sugerido un cierto papel de estos dominios en procesos de control de calidad de ARNm específicos y en eventos

post-transcripcionales, papel que sería consistente con el comportamiento dinámico de las moléculas de ARNm poly(A) en el núcleo.

Por último y aunque no se ha identificado hasta el momento una estructura proteica sobre la que se organiza el resto de los componentes, los *speckles* nucleares presentan proteínas con una clara función estructural: laminina A/C (Jagatheesan, Thanumalayan et al. 1999), actina asociada a los snRNPs (Nakayasu and Ueda 1984) y miosina V y VI. Como veremos más adelante en las posibles funciones de los *speckles* nucleares (apartado 2.3), a la actina y miosina se les atribuye un hipotético papel en el movimiento de genes que convergen en estos orgánulos, por lo que se sugiere que, además de su función estructural podrían tener un importante papel en la regulación transcripcional (Pestic-Dragovich, Stojiljkovic et al. 2000; Grummt 2006; Pederson 2008). En los *speckles* nucleares se encuentran también diferentes proteínas de carácter regulador, como el fosfatidil-inositol-(4,5)-bifosfato que es un lípido que actúa sobre las proteínas de unión a actina (Zhao et al 1998), y diferentes isoformas de la quinasa fosfatidil-inositol fosfato (PIPK) (Boronenkov, Loijens et al. 1998). Recientemente, la proteína Son, cuya secuencia primaria presenta múltiples serinas y un motivo tipo RS ha sido implicada en esta, hasta ahora hipotética, estructura organizativa dentro de los *speckles* nucleares (Sharma, Takata et al. 2010). En base a experimentos de interferencia de su expresión, Son ha sido propuesta como una posible proteína de andamiaje de los *speckles* nucleares. A pesar de este amplio espectro de proteínas con función estructural son numerosos los estudios que ponen en duda la presencia de una verdadera estructura subyacente en estos dominios. A manera de ejemplo, se ha demostrado que la laminina A/C no es esencial para la formación de los *speckles* nucleares (Vecerova, Koberna et al. 2004).

En líneas generales, los *speckles* nucleares se caracterizan por una composición proteica de carácter básico, que podría ser el reflejo de la relación funcional de estas compartimentos con el metabolismo del ARN (Bickmore and Sutherland 2002). A nivel de composición aminoacídica, los aminoácidos con mayor presencia en los *speckles* nucleares son la arginina y la serina, debido a la elevada presencia de dominios tipo RS en numerosos factores de *splicing* (Birney, Kumar et al. 1993). Otro aminoácido relativamente abundante es la prolina, aunque en general los aminoácidos de carácter no polar se encuentran ausentes en los *speckles* nucleares, especialmente la leucina (Bickmore and Sutherland 2002).

### 2.3 Función de los *speckles* nucleares

La función más clara y tradicionalmente aceptada para los *speckles* nucleares es la de ser compartimentos transcripcionalmente inactivos donde se almacenan, ensamblan y modifican los factores de *splicing* y del metabolismo del pre-ARNm (Figura I-5, (a)). Estos factores son reclutados desde los *speckles* nucleares hasta zonas cercanas a los mismos donde la transcripción está teniendo lugar de manera activa (en las FPs) (Huang and Spector 1996; Huang and Spector 1996; Gama-Carvalho, Krauss et al. 1997; Misteli, Caceres et al. 1997; Mabon and Misteli 2005). Este reclutamiento es dependiente de un CTD totalmente íntegro (Misteli and Spector 1999), lo que indica una conexión entre la fase de elongación transcripcional y el reclutamiento de los factores de *splicing* desde los *speckles* nucleares hasta los sitios de transcripción activa.

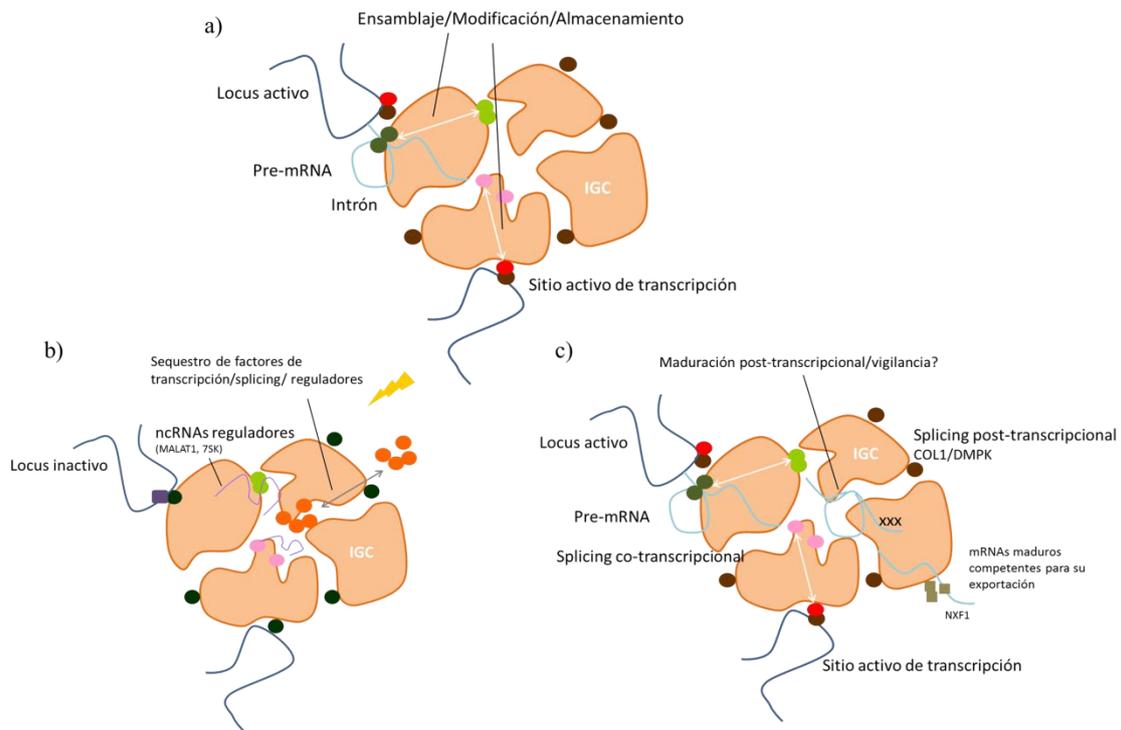
Existen numerosos estudios que apoyan el modelo propuesto: i) los estudios en célula viva ya mencionados demuestran que los factores de *splicing* son reclutados preferentemente desde y no hacia los *speckles* nucleares (Misteli, Caceres et al. 1997; Misteli and Spector 1999; Mabon and Misteli 2005), ii) experimentos de visualización por microscopía electrónica de la incorporación de nucleótidos marcados demuestran que la síntesis de pre-ARNm se localiza preferentemente en la periferia de los *speckles* nucleares (en las FPs), sugiriendo que el interior de estos dominios no desempeña ningún papel directo en los procesos de transcripción, *splicing* u otros procesos del metabolismo del ARNm; iii) por último, estudios que emplean sondas específicas para diferenciar las formas de pre-ARNm y las formas ya maduras de ARNm de un mismo gen han permitido observar que aquellos transcritos que se localizan en los *speckles* nucleares en algún momento de su síntesis y maduración lo hacen una vez que ya han sido en su mayor parte procesados (Hall, Smith et al. 2006). Por último, el dramático cambio que sufren los *speckles* nucleares ante condiciones tales como la inhibición de la transcripción apoya la idea de que estos compartimentos funcionen como lugares de almacenamiento, ensamblaje o modificación de los factores de *splicing* que son allí reclutados.

El papel de los *speckles* nucleares como sitios de almacenamiento y ensamblaje de factores es compatible con una potencial función reguladora inhibitoria mediante el secuestro de proteínas en el interior de estos dominios (Figura I-5, (b)). Así por ejemplo, tras la inducción de estrés, la quinasa SRPK1, es liberada del complejo formado por las

chaperonas Hsp70 y Hsp90 en el citoplasma. Esto provoca su mayor translocación desde este compartimento hacia el núcleo donde hiperfosforila de forma específica al factor esencial de *splicing* SRSF1 (anteriormente conocido como SF2/ASF) que en este estado permanece secuestrado en el interior de estas regiones (Zhong, Ding et al. 2009). Al ARNlnc MALAT1 también se le asigna un papel regulador en la dinámica de reclutamiento de las proteínas SR desde los *speckles* nucleares hasta los sitios de transcripción (Bernard, Prasanth et al.). Su depleción provoca un aumento en las formas hipofosforiladas de estos factores en el nucleoplasma, alterando no solo el *splicing* alternativo de diferentes pre-ARNm sino también otros eventos reguladores post-transcripcionales mediados por las proteínas SR (Tripathi, Ellis et al.). El mecanismo por el cual MALAT1 altera el ratio de proteínas SR hipo/hiperfosforiladas disponibles en el nucleoplasma no se conoce, aunque se propone un modelo en el que MALAT1 podría modificar la actividad de ciertas quinasas (SRPK1 y/o Clk/Sty) y fosfatasa (PP1 y/o PP2A). Sin embargo y contradiciendo un posible papel general de MALAT1 en la regulación de la expresión génica y el *splicing* de los pre-ARNm, los datos obtenidos por Spector y *col.* demuestran que este ARNlnc no es esencial para el ensamblaje y mantenimiento de los *speckles* nucleares, los niveles de fosforilación de las proteínas SR o para la viabilidad y proliferación celular (Zhang, Arun et al. 2012). Por último, y en línea con un posible papel de los *speckles* en el secuestro de factores, se ha propuesto que el ARNnc 7SK secuestra al complejo P-TEFb inactivo, del que forma parte junto con la proteína HEXIM, en los *speckles* nucleares (Prasanth, Camiolo et al. 2010).

A pesar de que los *speckles* nucleares son zonas transcripcionalmente inactivas, estos orgánulos se sitúan preferentemente cerca de regiones cromatínicas con una elevada densidad de genes transcripcionalmente activos (Niedojadlo, Perret-Vivancos et al. ; Iborra, Pombo et al. 1996; Pombo and Cook 1996; Wei, Somanathan et al. 1999; Xie, Martin et al. 2006). Así, otra atractiva e interesante visión de la función de los *speckles* nucleares se obtiene al observar que en ellos convergen genes codificantes de proteínas funcionalmente relacionadas que se expresan de manera coordinada de forma que sus transcritos, una vez procesadas gran parte de sus secuencias intrónicas, migran desde una posición periférica hasta el interior de los *speckles* (Shopland, Johnson et al. 2002; Shopland, Johnson et al. 2003; Moen, Johnson et al. 2004; Hu, Kwon et al. 2008). Estos datos abren la puerta a un posible papel de estos dominios como centros organizadores

de expresión génica y quizás un papel de control y vigilancia de los transcritos que se internalizan en ellos (Figura I-5, (c)).



**Figura I-5. Posibles modos de actuación de los *speckles* nucleares en la regulación de la transcripción y maquinaria de procesamiento de los pre-ARNm.** (a) Los *speckles* nucleares pueden actuar como sitios especializados en el almacenamiento, ensamblaje y reciclaje de complejos de transcripción y procesamiento de los pre-ARNm a través de múltiples ciclos de modificaciones post-traduccionales. (b) Determinados factores nucleares como ARNnc (MALAT1, 7SK) presentes en los *speckles* nucleares pueden actuar como “secuestradores” de factores de transcripción y *splicing*, bloqueando el reclutamiento de estos complejos desde los *speckles* nucleares hasta los sitios activos de transcripción cercanos. (c) En los *speckles* nucleares pueden tener lugar determinados eventos de procesamiento post-transcripcional y potenciales actividades de vigilancia de transcritos alterados. Además en ellos puede tener un posible acoplamiento entre las funciones de transcripción y ensamblaje de partículas ribonucleoproteicas necesarias para el transporte hasta el citoplasma de los ARNm.

Uno de los ejemplos más claros del papel de los *speckles* como centros de organización lo encontramos para los genes implicados en el programa de diferenciación muscular, COL1A1 y COL1A2, y en modelos de eritropoyesis (Moen, Johnson et al. 2004; Brown, Green et al. 2008). El papel propuesto para los *speckles* nucleares en estas asociaciones génicas sería el de mediar el reclutamiento de factores de *splicing* hacia

determinados genes activos en fase temprana de G1, cuando los *speckles* nucleares comienzan a reorganizarse tras la mitosis, lo cual supondría la señal necesaria para la asociación genes relacionados, mediante movimientos dinámicos de la cromatina y el proceso de nucleación final de los *speckles* nucleares, en fases posteriores del ciclo celular (Shopland, Johnson et al. 2003; Brown, Green et al. 2008). Otros ejemplos de asociaciones inter cromosómicas en los *speckles* nucleares vienen dados por los genes TFF1 y GREB1, genes inducibles por el receptor  $\alpha$  de estrógenos (ER $\alpha$ ). En células de cáncer de mama (MCF7) y células epiteliales humanas primarias (HMEC), la inducción con estradiol provoca el movimiento e interacción en los *speckles* nucleares de los cromosomas 21 y 7 que albergan, respectivamente, los locus de ambos genes (Hu, Kwon et al. 2008). Los autores de este trabajo proponen que este reordenamiento, rápido y dependiente de actina, miosina y dineína, permite y potencia la transcripción coordinada de genes regulados por esta señalización dependiente de hormonas. Es interesante señalar que la correlación espacial observada en los *speckles* nucleares es dependiente de promotor y de una transcripción activa, al menos en el caso del gen Hsp70, cuya asociación con estos dominios está regulada a nivel de la fase de elongación transcripcional (Hu, Kireev et al. 2009). Todas estas observaciones, y como ya habíamos intuido con el requerimiento de un CDT íntegro para el reclutamiento de los factores de *splicing* hasta sitios transcripcionalmente activos (Misteli and Spector 1999), apoyan la conexión funcional entre la fase de elongación de la transcripción y la maquinaria de procesamiento de los pre-ARNm.

Queda por resolver si realmente los *speckles* nucleares ya existentes juegan un papel directo en este orden superior de organización y coordinación de la expresión génica o sí por el contrario los *speckles* nucleares surgen como consecuencia de la concentración de factores requeridos en áreas próximas a genes activos que se expresan y mueven de forma coordinada. Probablemente, futuros experimentos de visualización en célula viva aportarán la información necesaria para discernir ambas posibilidades.

Son numerosas las preguntas que se abren a partir de la localización específica de determinados genes. ¿Cuál es la razón de esta localización? ¿Están regulados principalmente a nivel de la fase de elongación? ¿La transcripción y el procesamiento de los pre-ARNm resultantes se ve favorecida tras este reclutamiento coordinado hacia estos compartimentos? E igualmente interesante, ¿existe alguna ventaja post-transcripcional para estos genes que convergen en un mismo *speckle*? Aquí habría que

hablar de un hipotético papel de los *speckles* nucleares en funciones de control de calidad post-transcripcional.

De la información de que disponemos actualmente, la mayor parte de los eventos de *splicing* que tienen lugar en el núcleo de la célula eucariota ocurren de manera co-transcripcional en la periferia de los *speckles* nucleares (Johnson, Primorac et al. 2000; Shopland, Johnson et al. 2002; Hall, Smith et al. 2006). Ésta podría ser considerada como una interfase donde tiene lugar el *splicing* co-transcripcional de las moléculas de pre-ARNm. Sin embargo, algunos intrones son eliminados de manera post-transcripcional pudiéndose observar una acumulación relativa de ciertos transcritos en el interior de los *speckles* nucleares. Se ha demostrado que moléculas de pre-ARNm con sitios de *splicing* totalmente funcionales se acumulan en el interior de estos orgánulos, donde posteriormente tiene lugar su procesamiento antes de abandonar el núcleo (Dias, Dufu et al. 2010). Al interferir en la expresión de UAP65 o Aly, ambos componentes del complejo TREX encargado del transporte al citoplasma de las moléculas de ARNm procesadas, se produce una acumulación significativa de las moléculas de ARNm expresadas de forma exógena así como de las moléculas de ARNm poly(A) totales. Asimismo, el intrón 26 del gen COL1A endógeno es procesado de manera post-transcripcional, observándose su acumulación en el interior de los *speckles* nucleares. Mutaciones que afectan al *splicing* de este intrón inducen un aumento de su concentración en el interior de estos dominios (Johnson, Primorac et al. 2000), lo que evidencia un posible papel de control de transcritos defectuosos de los *speckles* nucleares. Datos adicionales de los *speckles* nucleares como ejecutores de funciones de vigilancia y control de ARNm defectuosos proceden del gen de la proteín-quinasa DMPK que, en la distrofia muscular de tipo 1, presenta una expansión del triplete (CTG) en el extremo 3'UTR (*untranslated region*). Los transcritos generados de esta mutación quedan incluidos en agregados ribonucleoproteicos que se localizan en la periferia de los *speckles* nucleares, no llegando a alcanzar el citoplasma (Holt, Mittal et al. 2007). Todos estos datos apoyan un más que interesante papel de los *speckles* nucleares en eventos post-transcripcionales tales como un *splicing* tardío o eventos de control de calidad de transcritos defectuosos.

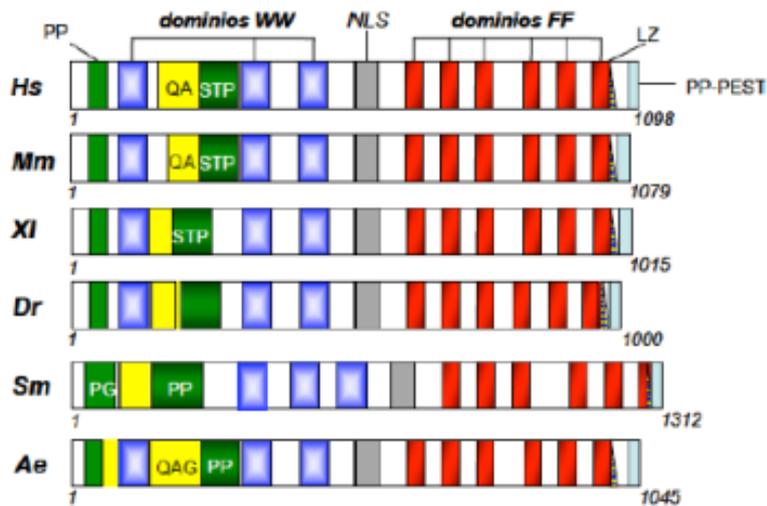
### **3. TCERG1, factor regulador de la elongación transcripcional** ***(Transcription elongation regulator 1)***

#### **3.1 Introducción**

TCERG1, previamente conocido como CA150, es una proteína de localización nuclear de 1098 aminoácidos y aproximadamente 120 kDa (Sune, Hayashi et al. 1997). Su secuencia está compuesta por múltiples motivos y dominios implicados en interacciones proteína-proteína, muchos de los cuales se encuentran presentes en numerosos factores de transcripción (Sune, Hayashi et al. 1997; Bohne, Cole et al. 2000). Los elementos más destacados de la secuencia de la proteína son los siguientes (Figura I-6):

1. En su porción aminoterminal destaca la presencia de tres dominios WW, una secuencia rica en residuos de prolina (P), una secuencia rica en residuos de alanina y glutamina (QA) y una secuencia rica en residuos de serina, treonina y prolina (STP).

Los dominios WW son los módulos proteicos más pequeños descritos en la naturaleza, compuestos aproximadamente de 40 aminoácidos. Median interacciones tipo proteína-proteína a través del reconocimiento de motivos peptídicos ricos en prolina y sitios de serina/treonina fosforilados. Según la naturaleza básica de las proteínas a las que se unen los dominios WW se clasifican en cuatro grupos diferentes (I a IV). Su nombre hace referencia a dos residuos de triptófano altamente conservados situados entre los aminoácidos 20 y 22. Se pliegan de forma estable formando una triple hélice antiparalela con estructura de lámina  $\beta$  en ausencia de ligando (Macias, Wiesner et al. 2002). Estos dominios se encuentran presentes en una gran variedad de proteínas envueltas en procesos tales como la transcripción y el procesamiento del ARN, la estabilidad y tráfico proteico, la señalización a través de receptores o la estabilidad del citoesqueleto. Estudios sistemáticos de proteómica han demostrado que dominios WW de una misma proteína pueden presentar diferentes especificidades de unión. De igual modo, un único polipéptido puede unirse a múltiples dominios WW a través de diferentes secuencias ricas en prolina (Ingham, Colwill et al. 2005). Estas características hacen de los dominios WW unas excelentes plataformas para el ensamblaje de complejos macromoleculares.



**Figura I-6. Estructura de TCERG1.** Representación esquemática de la estructura de TCERG1 y su conservación evolutiva en distintas especies eucariotas. *Hs*: *Homo sapiens*, *Mm*: *Mus musculus*, *Xl*: *Xenopus laevis*, *Dr*: *Daniorerio*, *Sm*: *Schistosomamansonii*, *Ae*: *Aedes aegyptii*. PP; secuencia rica en prolina. QA; secuencia rica en glutaminas y alaninas. STP; región rica en serinas, treoninas y prolina. WW; dominios con dos residuos altamente conservados de triptófano. NLS; señal de localización nuclear. FF; dominio con dos fenilalaninas altamente conservadas. LZ; secuencia de cremallera de leucinas (*leucine zipper*). PP-PEST; secuencia peptídica rica en residuos de prolina, ácido glutámico, serina, y treonina. PG; secuencia rica en residuos de prolina y glicinas. QAG; secuencia rica en residuos de glutamina, alanina y glicina. Junto a cada una de las proteínas se indica su tamaño en aminoácidos.

TCERG1 presenta tres dominios tipo WW:

-El dominio WW1 pertenece al llamado grupo fosfoserina/fosfotreonina (poS/poT), lo que explicaría las interacciones directas observadas con el factor de *splicing* SF1 (Goldstrohm, Albrecht et al. 2001; Goldstrohm, Greenleaf et al. 2001).

-El dominio WW2 pertenece a los grupos denominados L o poli-P, encargados de reconocer motivos cuya secuencia es PPPLIPPPP y PPLIPPPP. Al igual que el dominio WW1, interviene en el reconocimiento de factores de *splicing* ricos en motivos de prolina tales como SF1, U2snRNP y U2AF (Goldstrohm, Albrecht et al. 2001; Lin, Lu et al. 2004).

-El dominio WW3 ha sido caracterizado más recientemente y se ha descrito que para su correcto plegamiento, a diferencia de los otros dos dominios WW, es necesario la región  $\alpha$ -hélice situada justo delante del dominio (aminoácidos 567-573) (Fidan, Younis et al. 2011). Este dominio interacciona con el factor de *splicing* U2AF<sup>65</sup> (Sanchez-Alvarez, Goldstrohm et al. 2006).

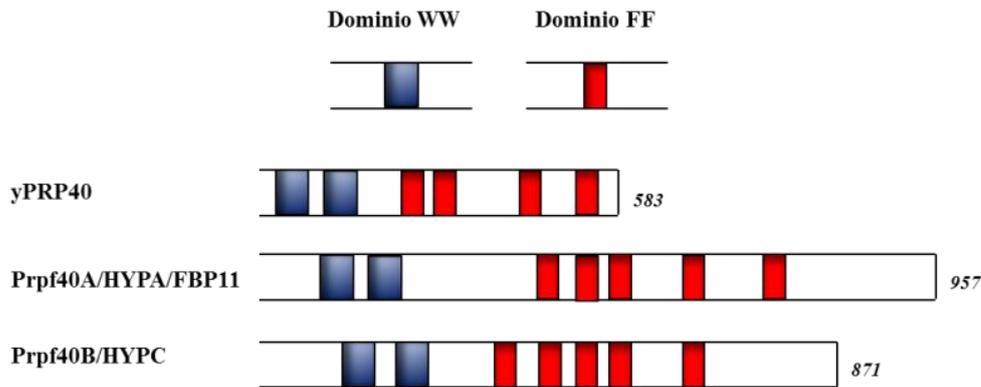
La región PP, situada entre los aminoácidos 32 y 132, probablemente esté implicada en mediar interacciones con dominios WW o dominios SH3 de otras proteínas. Una interacción intramolecular o bimolecular de TCERG1 es, asimismo, potencialmente posible. La región QA abarca la zona comprendida entre los aminoácidos 178 y 257. Se trata de una región de 38 repeticiones del dipéptido glutamina-alanina (QAQV(QA)<sub>29</sub>QVQAQVQAQV(QA)<sub>2</sub>) (Sune, Hayashi et al. 1997). Alteraciones en la longitud de esta secuencia podrían estar asociadas a diferencias en la edad de inicio de la enfermedad de Huntington (Holbert, Denghien et al. 2001; Chattopadhyay, Ghosh et al. 2003; Arango, Holbert et al. 2006; Andresen, Gayan et al. 2007). La región STP abarca la zona comprendida entre los aminoácidos 258 y 430. Se trata de una región rica en residuos de serina, treonina y prolina alternada con motivos cortos altamente hidrofóbicos.

2. En la porción carboxiterminal destacan seis dominios FF, que por la importancia que suponen para este trabajo, serán tratados más extensivamente en un apartado posterior (apartado 4). Además en esta región encontramos una secuencia de cremallera de leucinas (*leucine zipper*), parcialmente solapante con el dominio FF6, una secuencia de poliprolinas semejante a la descrita anteriormente y una secuencia reguladora tipo PEST (secuencia peptídica rica en residuos de prolina (P), ácido glutámico (E), serina (S), y treonina (T)) probablemente responsable de la regulación de la estabilidad de la proteína.

La organización estructural de TCERG1 se encuentra muy conservada en organismos eucariotas. El mayor grado de conservación entre los homólogos descritos de TCERG1 (desde humanos hasta platelmintos) se encuentra en los dominios WW y en toda la región carboxiterminal (Figura I-6). Los demás fragmentos de la secuencia están sometidos a un mayor grado de divergencia.

Esta organización basada en múltiples repeticiones de dominios WW en la región aminoterminal y dominios FF en la región carboxiterminal de TCERG1 está presente en otros factores, como el factor esencial de *splicing* de levaduras Prp40 (yPrp40) y sus ortólogos en mamífero Prpf40A/HYPA/FBP11 (de ahora en adelante FBP11) y Prpf40B/HYPC (de ahora en adelante HYPC) (Figura I-7). Diferentes estudios han revelado la asociación de estas proteínas con las maquinarias de transcripción y

procesamiento de los ARN de manera similar a como lo hace TCERG1 (ver a continuación) (Bedford and Leder 1999; Morris and Greenleaf 2000; Allen, Friedler et al. 2002) Goldstrohm et al 2001. Estos datos, junto con la arquitectura modular de estos factores apoya un modelo en el que esta familia de proteínas tendría un papel comunicador entre ambas maquinarias.



**Figura I-7. Los dominios WW y FF están presentes en otras proteínas.** Las proteínas representadas en la imagen están relacionadas, al igual que TCERG1, con las maquinarias de transcripción y *splicing* y comparten además con ella la disposición de dominios WW y dominios FF en las regiones amino y carboxiterminal respectivamente. Se representan en azul los dominios WW y en rojo los dominios FF. Junto a cada una de las proteínas se indica su tamaño en aminoácidos.

### 3.2 TCERG1 en el proceso de transcripción

La proteína TCERG1 fue inicialmente purificada a partir de fracciones de extractos nucleares de células HeLa mediante columnas de afinidad de la proteína Tat del virus de la inmunodeficiencia humana de tipo 1 (VIH-1) (Sune, Hayashi et al. 1997). La inmunodepleción de TCERG1 y factores asociados reduce la trans-activación transcripcional mediada por Tat *in vitro*, por lo que se considera que TCERG1 es necesaria para el progreso hacia un estado de elongación progresiva. La sobre-expresión de TCERG1 en células resulta en una disminución específica de la actividad basal del promotor del VIH-1 a través de la modulación de la elongación transcripcional. Otros promotores virales como CMV, HSV-TK o SV-40 son refractarios al efecto de la sobre-expresión de TCERG1 (Sune and Garcia-Blanco 1999). La actividad transcripcional del promotor celular de la  $\alpha 4$ -intergrina también es modulada por TCERG1. El efecto de TCERG1 es dependiente de una caja TATA totalmente funcional, al menos en el caso

del VIH-1 (Sune and Garcia-Blanco 1999). En conjunto, esos datos sugieren un papel de TCERG1 en la regulación del proceso de síntesis del pre-ARNm, en particular en el proceso de la elongación de la transcripción mediada por la ARNpolIII.

TCERG1 puede ejercer un efecto tanto positivo como negativo sobre la transcripción dependiendo del contexto (Pearson, Robinson et al. 2008; Sanchez-Alvarez, Montes et al. 2010), habiéndose descrito que modificaciones post-traduccionales como la SUMOilación (adición del grupo SUMO, ver más adelante en apartado 4.3.2) inhiben la activación transcripcional mediada por TCERG1 en un modelo de plásmido reportero (Sanchez-Alvarez, Montes et al. 2010).

Consistente con este papel en transcripción, TCERG1 co-inmunoprecipita en extractos nucleares con la ARN polIII e interacciona *in vitro* con su CTD hiperfosforilado a través de los dominios FF de la proteína (Sune, Hayashi et al. 1997; Carty, Goldstrohm et al. 2000; Goldstrohm, Albrecht et al. 2001). Construcciones que incluyen los dominios WW2 y WW3 también se unen al CTD fosforilado aunque con una afinidad notablemente menor (Carty, Goldstrohm et al. 2000). Otras interacciones descritas con la maquinaria de elongación transcripcional implica a componentes del complejo P-TEFb (ciclina T1 y quinasa CDK9) y a los factores TFIIH y Tat-SF1 (*Tat specific factor 1*) (Smith, Kulkarni et al. 2004; Sanchez-Alvarez, Goldstrohm et al. 2006). Estudios de nuestro laboratorio ha puesto de manifiesto la presencia de TCERG1 en los complejos de pre-iniciación del VIH-1 (Montanuy, Torremocha et al. 2008), siendo reclutado probablemente a través del promotor (McFie, Wang et al. 2006; Montanuy, Torremocha et al. 2008; Jablonski, Amelio et al. 2010) en un complejo que contendría diversas proteínas SR y quizás promovido por la unión de Tat a la estructura viral TAR (Jablonski, Amelio et al. 2010). Sin embargo, la influencia que TCERG1 ejerce sobre la transcripción recae preferentemente sobre la etapa de la elongación antes que sobre la iniciación de la misma (Sune and Garcia-Blanco 1999). De acuerdo con este modelo, TCERG1 ha sido localizado a lo largo de regiones codificantes de diferentes genes (Montes, Cloutier et al. 2012) (y datos de laboratorio sin publicar). En el caso del gen de apoptosis *Bcl-x*, TCERG1 es capaz de modificar la distribución de la polimerasa sobre la unidad génica, de forma que el silenciamiento génico de TCERG1 favorece la acumulación de moléculas de ARN polIII en elementos reguladores. La sobre-expresión de TCERG1 aumenta la concentración de moléculas de polimerasa en regiones más

distales del gen. Consistente con estos datos, TCERG1 aumenta la procesividad de la ARNpolIII para transcribir el gen *Bcl-xin vivo* (Montes, Cloutier et al. 2012).

### 3.3 TCERG1 en el proceso de *splicing*. Localización en los *speckles* nucleares

Numerosos datos apoyan un papel de TCERG1 en el proceso de *splicing*:

i) TCERG1 interacciona con diferentes factores de *splicing* como SF1 o U2AF<sup>65</sup> (Goldstrohm, Albrecht et al. 2001; Sanchez-Alvarez, Goldstrohm et al. 2006).

ii) TCERG1 ha sido co-purificado en complejos del *spliceosoma* (Neubauer, King et al. 1998; Jurica, Licklider et al. 2002; Makarov, Makarova et al. 2002; Rappsilber, Ryder et al. 2002; Zhou, Licklider et al. 2002; Deckert, Hartmuth et al. 2006).

iii) Consistente con un papel en el procesamiento del pre-ARNm, TCERG1 colocaliza en los *speckles* nucleares. Un análisis en detalle de la distribución espacial de TCERG1 dentro de estos dominios muestran como este factor se distribuye a lo largo de su periferia e internalizado, pero quedando excluido de la parte interna, *core* o núcleo de los mismos. Esta distribución coincide con sitios de transcripción activa cercanos a los *speckles* nucleares. La localización de TCERG1 en estos lugares es independiente de la presencia de ADN y de transcripción activa dentro de la célula, lo que sugiere que esta localización tan específica no es el resultado de la asociación de TCERG1 con genes que están siendo transcritos en la zona de las FPs (Sanchez-Alvarez, Goldstrohm et al. 2006). Los dominios FF de TCERG1 son esenciales para su localización en la periferia de los *speckles* nucleares (Sanchez-Alvarez, Goldstrohm et al. 2006) sin embargo, y hasta el desarrollo de esta tesis, los requerimientos proteicos exactos para la correcta localización de la proteína no habían sido descritos.

iv) Por último, diversos datos funcionales apoya un papel de TCERG1 en el procesamiento del pre-ARNm. Así, TCERG1 modula el *splicing* alternativo de minigenes reporteros derivados de los genes de la  $\beta$ -globina,  $\beta$ -tropomiosina, *CD44*, *fibronectina* y *Bcl-x* (Lin, Lu et al. 2004; Pearson, Robinson et al. 2008; Sanchez-Alvarez, Montes et al. 2010; Montes, Cloutier et al. 2012). Para este último gen, estudios recientes de nuestro laboratorio han demostrado también el papel de TCERG1

sobre el gen endógeno, favoreciendo la expresión de su isoforma pro-apoptótica (Montes, Cloutier et al. 2012).

## **4. Dominios FF y señalización hacia los *speckles* nucleares**

### **4.1 Dominios FF**

#### **4.1.1 Generalidades**

Los dominios FF son módulos proteicos de aproximadamente 60 aminoácidos de longitud descritos por primera vez en 1999 (Bedford and Leder 1999). Deben su nombre a dos residuos de fenilalanina altamente conservados situados cerca de los extremos amino y carboxiterminal del dominio. Su presencia se limita a un número restringido de proteínas eucariotas que pueden ser divididas en dos grupos:

i) Proteínas nucleares, relacionadas principalmente con las maquinarias de transcripción y *splicing* del pre-ARNm y que contienen a su vez dominios tipo WW en su región aminoterminal. Dentro de este grupo encontramos a TCERG1, al factor esencial de *splicing* de levaduras y Prp40 y sus ortólogos en mamífero FBP11 y HYPC.

ii) La familia p190 de las RhoGAP/GTPasas que, a diferencia de las anteriores proteínas, son de localización citoplasmática y en su región aminoterminal no presentan dominios WW sino un dominio catalítico de tipo GTPasa. Estas proteínas están implicadas en las vías de señalización reguladoras de la organización del citoesqueleto, aunque la presencia de dominios FF en esta familia de proteínas sugiere una relación con el anterior grupo de proteínas relacionadas con el metabolismo del ARN. Acorde con un papel común entre ambos grupos de proteínas portadoras de dominios FF, el factor p190A antagoniza la actividad del factor de transcripción TFII-I mediante su secuestro en el citoplasma. La fosforilación de la tirosina situada en la posición 308 (Tyr308) de p190 (primer dominio FF de la proteína) inducida por factores de crecimiento bloquea su interacción con TFII-I, haciendo posible que éste se desplace hasta el núcleo donde regula la transcripción de genes inducibles incluyendo a *c-fos* (Jiang, Sordella et al. 2005).

Una de las características de los dominios FF es que se presentan en tándem de 2 a 6 repeticiones separados por secuencias de unión de longitud variable que oscilan entre 64 y 187 residuos (Bedford and Leder 1999). Dentro de esta disposición en tándem

generalizada encontramos una excepción, el factor URN1 de levaduras, que contiene un único dominio FF. Este factor ha sido recientemente identificado en un análisis masivo de proteínas que interaccionan con el factor esencial de *splicing* Prp19 (Ren, McLean et al. 2011).

#### **4.1.2 Interacciones de los dominios FF**

Es probable que la organización en tándem de estos dominios sea importante para la función de las proteínas que los contienen. Diversos estudios indican que los diferentes dominios FF de una misma proteína son capaces de interaccionar de forma individual con un determinado factor a través de múltiples y débiles interacciones, y es precisamente la disposición en tándem de estos dominios la que incrementa la especificidad de unión (Carty, Goldstrohm et al. 2000; Morris and Greenleaf 2000; Smith, Kulkarni et al. 2004). En el caso de TCERG1, dominios FF específicos se unen de forma individual a proteínas como Tat-SF1 (Smith, Kulkarni et al. 2004) o al dominio CTD de la ARN polIII provocando la suma de estas múltiples interacciones una mayor afinidad por el ligando (Carty, Goldstrohm et al. 2000). Sin embargo, también cabe la posibilidad de que varios dominios FF actúen de manera conjunta en la mediación de interacciones proteicas. De este modo, y para la ya citada proteína p190A, se ha sugerido que sus dominios FF podrían formar una estructura de orden superior que permitiría la interacción óptima con su ligando TFII-I (Jiang, Sordella et al. 2005).

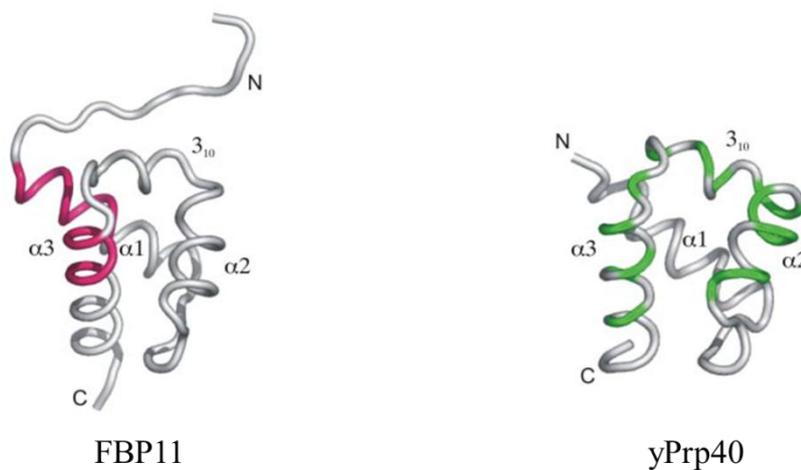
De manera análoga a lo que sucede con TCERG1, se ha demostrado la interacción de yPrp40 y FBP11 con el dominio CTD fosforilado de la ARN polIII (Morris and Greenleaf 2000; Allen, Friedler et al. 2002), lo que hizo proponer a los dominios FF como una nueva clase de motivos proteicos de unión a fosfopéptidos. La interacción de yPrp40 al CTD fosforilado se realiza a través de múltiples interacciones con la región que contiene a los dominios FF (Gasch, Wiesner et al. 2006). Sin embargo, dentro de ésta el dominio FF1 de yPrp40 no interacciona con el CTD de la ARN polIII en ninguno de sus estados de fosforilación, aunque sí lo hace con el factor Clf/Syf3p a través de dominios tipo crn-TRP (*crooked neck like tetratricopeptid repeats*) presentes en éste, los cuales no son fosforilados (Gasch, Wiesner et al. 2006). Estas observaciones demuestran que no todos los dominios FF son motivos de unión a fosfopéptidos y que además son funcionalmente diferentes.

A nivel de secuencia aminoacídica, en el factor Tat-SF1 se ha descrito la presencia de varias secuencias consenso tipo ((D/E)<sub>2/5</sub>-F/W/Y-(D/E)<sub>2/5</sub>) a través de las cuales los dominios FF de TCERG1 establecen múltiples interacciones (Smith, Kulkarni et al. 2004). La fosforilación de las Ser 2 y 5 del CTD de la ARN polII crea un motivo similar al anterior, en el que residuos aromáticos quedan rodeados por aminoácidos cargados negativamente (Gasch, Wiesner et al. 2006). Esta similitud en las secuencias de unión a dominios FF presentes en factores de las maquinarias de transcripción y *splicing* sugieren la participación de las proteínas que contiene dominios FF en el acoplamiento funcional entre ambos procesos. Las interacciones que establecen los dominios FF podrían, al menos en parte, deberse a sus valores de pKa. Por ejemplo, el dominio FF1 de yPrp40 presenta una carga neta negativa con un pKa aproximado de 4.7, lo que podría explicar la falta de interacción con el dominio CTD fosforilado de la ARN polII (Gasch, Wiesner et al. 2006). En el caso del dominio FF1 de FBP11, cuya interacción con el CTD fosforilado ha sido demostrada (Allen, Friedler et al. 2002), el pKa es de 9.6. A pesar de la utilidad de los valores de pKa como predictores de las posibles interacciones de los dominios FF, este parámetro debe ser utilizado con precaución. Por ejemplo, el dominio FF1 de yPrp40 interacciona con los dominios tipo *cnr*-TPR, mientras que el dominio FF de URN1, a pesar de tener el mismo valor pKa, no establece este tipo de interacción (Bonet, Ramirez-Espain et al. 2008). Como se describe a continuación, la estructura de los dominios FF se encuentra altamente conservada aunque existe un elevado grado de divergencia en el modo en que se distribuyen las cargas negativas a lo largo de sus secuencias (lo que se ha denominado “parches cargados negativamente”). Ésta distribución de cargas puede ser más relevante en las interacciones proteína-proteína que el valor de pKa (Bonet, Ramirez-Espain et al. 2008).

#### **4.1.3 Estructura de los dominios FF**

Hasta el momento, se ha descrito en detalle la estructura secundaria de los dominios FF1 y FF4 del factor yPrp40 (Gasch, Wiesner et al. 2006; Bonet, Ramirez-Espain et al. 2008), del único dominio del factor URN1 (Bonet, Ramirez-Espain et al. 2008), de los dominios FF1 y FF5 de la proteína FBP11 (Allen, Friedler et al. 2002) y de los dominios FF1, FF2, FF3 y FF4 de TCERG1. Aunque las secuencias resultan ser

divergentes, la estructura general del dominio se basa en tres hélices  $\alpha$  y una hélice tipo  $3_{10}$  que se posicionan en una conformación tipo  $\alpha 1-\alpha 2-3_{10}-\alpha 3$  (Figura I-8). Dentro de este alto grado de conservación en el plegamiento de los dominios FF, encontramos una variación a esta conformación típica en el primer dominio FF del factor p190A (Bonet, Ruiz et al. 2009), donde la hélice  $3_{10}$  es remplazada por una hélice  $\alpha$  más larga. El plegamiento de los dominios FF encierra un núcleo hidrofóbico que contiene la secuencia FXXLL, situada en la primera hélice  $\alpha$ , donde encontramos uno de los dos residuos de fenilalanina que dan nombre al dominio y dos residuos de leucina, todos ellos altamente conservados. Otro residuo fuertemente conservado del dominio corresponde al triptófano que ocupa la posición 411 de la secuencia del factor FBP11 (Trp411). Este residuo se sitúa al comienzo de la segunda hélice  $\alpha$  y queda igualmente encerrado en el núcleo hidrofóbico. En general, la secuencia aminoacídica de los dominios FF está compuesta mayoritariamente por residuos de carácter hidrofóbico o aromático formando parte de la hélices  $\alpha$ , haciendo posible el empaquetamiento de las mismas y el establecimiento de su estructura secundaria.



**Figura I-8. Estructura secundaria de los dominios FF.** La imagen representa los dominios FF1 de los factores FBP11 y yPrp40. Se representan los extremos amino y carboxiterminal así como los elementos de la estructura secundaria (hélice  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$  y  $3_{10}$ ). En rojo y verde se indican los sitios de unión a través de los cuales ambos dominios establecen interacciones con otras proteínas (Imagen tomada de (Gasch, Wiesner et al. 2006)).

Fuera de las hélices  $\alpha$  encontramos de forma conservada dos residuos, tirosina y leucina, situadas en la hélice tipo  $3_{10}$  y encargados de anclar esta hélice al núcleo hidrofóbico. En cuanto a los residuos situados en la superficie del dominio, éstos son los menos conservados. Dentro de ellos los menos divergentes son dos dominios de aspártico y arginina, presentes en al menos el 50% de los dominios FF descritos y situados entre la segunda hélice  $\alpha$  y la hélice  $3_{10}$ .

A pesar de que la estructura secundaria de los dominios FF de forma individual está bien descrita, la presencia o formación de estructuras de orden superior dentro del tándem formado por estos dominios no se encuentra tan estudiada. Los estudios más recientes que abordan esta posibilidad usan precisamente como modelo los dominios FF1 y FF2 de TCERG1 (Murphy, Hansen et al. 2009). Experimentos de resonancia magnética nuclear (NMR) demuestran que aunque ambos dominios no se encuentran fijamente orientados no se pliegan de manera totalmente independiente. La explicación a esta organización se debe a la interacción que la secuencia de unión entre los dominios FF1 y FF2 establece con ciertos residuos de FF1. La organización que alcanzan estos dos dominios dentro del tándem en el que se encuentran les permite un reconocimiento óptimo de diferentes factores gracias a la movilidad que presentan unos dominios respecto a otros.

#### **4.2 Dominios de señalización hacia los speckles nucleares**

Dentro de los dominios proteicos definidos como señales de localización hacia los *speckles* nucleares, los mejor estudiados y caracterizados son los dominios RS, los dominios RRM y las repeticiones de histidinas.

El papel de los dominios RS en la señalización hacia los *speckles* nucleares fue descrito en primer lugar en las proteínas tipo-SR Tra y SWAP, ambas reguladoras del proceso de *splicing* en *Drosophila* (Li and Bingham 1991; Hedley, Amrein et al. 1995). En ellas, este dominio es necesario y suficiente para una correcta localización en los *speckles* nucleares. Este papel como señal de localización de los dominios RS ha sido corroborado en diferentes proteínas SR como SRp20, SC35 y 9G8, todas con un único motivo RRM en su región aminoterminal. En ellas el dominio RS es también necesario para su localización y suficiente para dirigir a una proteína citoplasmática reportera

hasta los *speckles* nucleares (Caceres, Misteli et al. 1997). Otro factor en el que se ha descrito un requerimiento absoluto del dominio RS para una correcta acumulación en los *speckles* nucleares es la quinasa CrkRS (Ko, Kelly et al. 2001). En el caso de proteínas SR con dos motivos RRM (SRSF1 o SRp40) la localización hacia los *speckles* nucleares no es dependiente del dominio RS, ya que su delección no impide su correcta localización en estos orgánulos (Caceres, Misteli et al. 1997), siendo además este motivo incapaz de dirigir a una proteína heteróloga hasta los *speckles*. Estas diferencias en los requerimientos del dominio RS sugirieron la existencia de dos mecanismos en la localización hacia los *speckles* nucleares; un mecanismo directo, en el que la presencia de una señal de localización determina la correcta distribución nuclear de la proteína gracias a las interacciones que se establecen a través de ella y, un mecanismo indirecto en el que proteínas que carecen de una señal de localización propiamente dicha interactúan con otras proteínas que sí presentan dicha señal (Caceres, Misteli et al. 1997). En el caso de los dominios RS, el mecanismo de localización directo estaría basado en las interacciones proteína-proteína que se establecen a través de ellos, las cuales a su vez, como veremos a continuación, están reguladas por ciclos de fosforilación y desfosforilación. Este modelo regula de forma mayoritaria la localización de proteínas SR con un solo dominio RRM, como SC35, que de forma consistente con esta forma de reclutamiento se localiza en los *speckles* nucleares incluso cuando la célula es sometida a tratamientos con ARNasa (Spector, Fu et al. 1991). Además se ha descrito, tanto para SC35 como para otras proteínas SR con un solo dominio RRM (SRp20 y 9G8), la necesidad de un pequeño motivo peptídico consistente en tres o cuatro aminoácidos de carácter básico junto con el dominio RS para la correcta localización de la proteína (Hedley, Amrein et al. 1995). Este motivo no se encuentra en proteínas con dos dominios RRM (SRSF1, SRp40 y SRp30c). En cuanto a la localización indirecta, un ejemplo claro de este modelo es la localización de un mutante de la proteína Tra sin dominio RS. Este mutante es aún capaz de concentrarse en los *speckles* nucleares, probablemente a través de su interacción con proteínas como Tra2 que sí presenta un dominio RS (Hedley, Amrein et al. 1995). La localización en los *speckles* nucleares de los mutantes de SRSF1 y SRp40 delecionados de sus dominios RS hace pensar que su distribución subnuclear está también determinada de manera indirecta, mediante interacciones proteína-proteína o proteína-ARN a través de cualquiera de sus tres dominios. En estas proteínas, los dominios RRM y el dominio RS presentarían señales de localización débiles que podrían considerarse redundantes. El

requerimiento absoluto de un dominio RRM para una correcta localización en los *speckles* nucleares ha sido demostrado en el factor PSF (*protein-associated splicing factor*) (Dye and Patton 2001). El dominio RRM2 de este factor es necesario para su acumulación en los *speckles*, sin embargo no es suficiente para la localización de otras proteínas heterólogas en estos orgánulos. Al igual que en las proteínas SR con dos motivos RRM, la localización subnuclear de este factor probablemente se realice de manera indirecta, mediante interacciones proteína-proteína o proteína-ARN.

El extremo carboxiterminal de la quinasa DYRK1A presenta una región rica en histidinas necesaria para su correcta localización, y a la vez suficiente para la acumulación de otras proteínas en los *speckles* nucleares (Alvarez, Estivill et al. 2003). Esta misma señal también ha sido identificada en la ciclina T1 lo que sugiere un papel más generalizado para las repeticiones de histidinas en el proceso de localización en los *speckles* nucleares. Esta hipótesis ha sido recientemente corroborada mediante una búsqueda genómica a gran escala de proteínas que contienen repeticiones de histidinas (Salichs, Ledda et al. 2009). El estudio realizado por Salichs y colaboradores ha identificado 86 proteínas humanas que contienen extensiones de 5 o más histidinas, estando la mayoría de estas proteínas implicadas en el metabolismo del ADN y el ARN. De ellas 15 se localizan en los *speckles* nucleares, lo que evidencia que las repeticiones de histidinas son una señal de localización general hacia esos orgánulos nucleares. La base molecular de la localización mediada por las repeticiones de histidinas puede deberse a las propiedades de este aminoácido, cuya cadena lateral puede variar en su carga dependiendo del pH. El gradiente de carga que provoca la presencia de estos tramos de histidinas en las proteínas que las contienen favorece las uniones proteína-proteína o proteína-ADN a través de interacciones electroestáticas.

Otras señales de localización a *speckles* nucleares descritas son el dominio “forkhead-associated” en NPP1 (*protein phosphatase-1 regularor*) (Jagiello, Van Eynde et al. 2000), las repeticiones de treoninas y prolinas (dominios TP) de SF3b1/SF3b<sup>155</sup> (Eilbracht and Schmidt-Zachmann 2001) y la región rica en arginina, prolina y serina del factor SRm160 (Wagner, Chiosea et al. 2003).

### 4.3 Mecanismos moleculares de localización hacia los *speckles* y otros orgánulos nucleares

#### 4.3.1 Fosforilación

Los ciclos de fosforilación y desfosforilación a los que se ven sometidas las proteínas SR son el mejor ejemplo de mecanismo molecular regulador del continuo intercambio proteico que tiene lugar entre los *speckles* nucleares y los sitios de transcripción activa. La mayoría de las proteínas SR se encuentran en un estado hipofosforilado en el interior de los *speckles* nucleares donde, como ya se ha comentado, existen diferentes quinasas capaces de fosforilar de forma específica a los dominios RS. Entre éstas, las familias de las SRPK (SRPK1 y SRPK2) y las Clk (Cdc2-like kinases; Clk/Sty-1, Clk2-4) son las mejor caracterizadas. Las SRPK fosforilan de forma específica residuos de serina, mientras que las quinasas Clk presentan una especificidad dual pudiendo fosforilar residuos de serina y treonina así como residuos de tirosina (Colwill, Feng et al. 1996). Análisis enzimáticos han demostrado que SRPK utiliza un mecanismo altamente procesivo para la fosforilación de una región determinada dentro del dominio RS de cada proteína SR (aproximadamente la mitad de las serinas disponibles) (Aubol, Chakrabarti et al. 2003; Ngo, Giang et al. 2008). Para la fosforilación mediada por SRPK es necesaria una secuencia de anclaje (*docking motif*) para la quinasa en la proteína SR a fosforilar (Ngo, Chakrabarti et al. 2005). Posteriormente, Clk finaliza la fosforilación del dominio sugiriendo así la posibilidad de que ambas quinasas actúan de manera coordinada (Ngo, Chakrabarti et al. 2005; Hagopian, Ma et al. 2008). La idea de una cooperación entre SRPK y Clk es consistente con la distribución espacial de ambas quinasas. Mientras que la familia Clk se localiza preferentemente en el núcleo (Colwill, Pawson et al. 1996; Nayler, Schnorrer et al. 1998), las proteínas SRPK se sitúan tanto en el núcleo como en el citoplasma (Wang, Lin et al. 1998; Ding, Zhong et al. 2006), donde son necesarias para iniciar la fosforilación de las proteínas SR y mediar así su transporte hacia el núcleo y su localización en los *speckles* nucleares.

Una vez que tiene lugar la hiperfosforilación, llevada a cabo principalmente por la quinasa Clk, las proteínas SR abandonan los *speckles* nucleares para poder así asociarse al *spliceosoma* que se está formando en el sitio de transcripción activo. Aquí son desfosforiladas lo que les permite participar en la progresión del proceso de *splicing* y además restablecer su afinidad por los *speckles* (Tazi, Kornstadt et al. 1993; Cao,

Jamison et al. 1997; Xiao and Manley 1997; Misteli, Caceres et al. 1998). El intercambio de proteínas SR entre los *speckles* y los sitios de transcripción se ve interrumpido con el uso de inhibidores específicos de quinasas, lo que apoya el papel de la fosforilación en la dinámica existente entre ambos lugares (Misteli, Caceres et al. 1997).

El efecto de la fosforilación sobre el intercambio basal de proteínas entre los *speckles* y el nucleoplasma se evidencia de forma muy clara tras la sobre-expresión de Clk (Colwill, Pawson et al. 1996) o la adición de SRPK1 a células permeabilizadas (Gui, Lane et al. 1994), generando una completa redistribución de los factores de *splicing* y la desorganización total de los *speckles*. Además, la sobre-expresión de un mutante de Clk carente de su actividad quinasa provoca una elevada concentración de factores de *splicing* en los *speckles*, probablemente por la incapacidad de éstos de alcanzar los sitios donde está teniendo lugar la transcripción (Sacco-Bubulya and Spector 2002). Para las quinasas DYRK1A y CDKL5 también se ha demostrado esta capacidad de desorganización de los *speckles*, apoyando el papel regulador de las quinasas presentes en los *speckles* nucleares en la disponibilidad de factores de *splicing* en los sitios de transcripción activa y por tanto, en el ensamblaje del *spliceosoma* y la elección del sitio de *splicing* de ciertos pre-ARNm (Alvarez, Estivill et al. 2003; Ricciardi, Kilstrup-Nielsen et al. 2009). Estas dos proteínas se unirían al pequeño grupo de quinasas para las que se ha demostrado un papel modulador sobre el *splicing*, entre ellas Clk (Duncan et al 1997; Xiao and Manley 1998; Prasad et al 1999; Wang et al 1999; Hartmann AM et al, 2001; Allemand, E. et al 2007; Muraki, M et al 2004). Un estudio reciente demuestra que SRPK1 es capaz de modificar el patrón de *splicing* de un minigen reportero a través de la hiperfosforilación de diferentes proteínas SR y su redistribución a los *speckles* nucleares ante condiciones de estrés osmótico (Zhong, Ding et al. 2009). Se establece por tanto una relación entre los cambios en los eventos de *splicing* que tienen lugar en la célula tras diferentes señalizaciones celulares y la acción de las quinasas encargadas de regular el estado de fosforilación de las proteínas SR. Otro ejemplo incluye la señalización mediada por insulina que activa a la quinasa PKB/Akt provocando la fosforilación del factor de *splicing* SRp40 dando así lugar a un cambio en el patrón de *splicing* alternativo de las moléculas de pre-ARNm de PKC (Patel, Chalfant et al. 2001; Patel, Kaneko et al. 2005). Este cambio tiene como consecuencia favorecer la formación de una isoforma encargada de la introducción de glucosa al interior de la célula. Otra vía de regulación de la actividad de estas quinasas ha sido propuesta a

través de la acción de las moléculas del proteosoma presentes en los *speckles* nucleares. Experimentos de silenciamiento de las subunidades 20S-PA28 $\gamma$  del complejo del proteosoma provocan una reestructuración de estos orgánulos y un menor reclutamiento de SRSF1 hasta el sitio de transcripción activa, probablemente a través de una alteración de la actividad de quinasa presente en los *speckles* nucleares (Baldin, Militello et al. 2008).

Relacionando aún más los cambios en los patrones de fosforilación y el resultado del proceso de *splicing*, cabe destacar que las proteínas con acción fosfatasa ejercen una función importante en este contexto. SRp38 (SRSF10) es extensivamente desfosforilada mediante la acción de la fosfatasa PP1 en respuesta a estrés térmico (Shi and Manley 2007), ejerciendo así su papel como represor general de *splicing* ya que impide la correcta formación del *spliceosoma*. Junto con PP1, PP2A, PP2B y PP2C conforman el conjunto de fosfatasas específicas de serinas y treoninas, aunque solo PP1 y PP2A han sido implicadas en la regulación del estado de fosforilación de las proteínas SR (Misteli 1999). Un exceso de PP1 recombinante provoca la desfosforilación de proteínas SR y la inhibición del *splicing in vitro* (Mermoud, Cohen et al. 1994). De igual modo, la microinyección de PP1 recombinante provoca la desfosforilación de proteínas SR y su localización en los *speckles* nucleares (Misteli and Spector 1996). La restauración de los niveles de fosforilación es esencial para la reanudación de la actividad de *splicing*. Recientemente se ha descrito la acumulación de precursores inmaduros de las quinasas Clk1 y Clk4 y su rápido procesamiento ante condiciones de estrés, cuando los procesos de transcripción y *splicing* se encuentran inhibidos. De esta manera, la rápida expresión de estas quinasas asegura una correcta refosforilación de las proteínas SR una vez que las condiciones de estrés han finalizado (Ninomiya, Kataoka et al.).

El estado de fosforilación regula también la localización de proteínas situadas en otros orgánulos nucleares. Por ejemplo, la coilina es fosforilada durante la fase de mitosis en al menos dos residuos de serina (Carmo-Fonseca, Ferreira et al. 1993), provocando una disminución en la capacidad de autointeracción de la proteína (Hebert and Matera 2000). Este cambio en el estado de fosforilación coincide con la desorganización de los cuerpos de Cajal durante esta fase del ciclo celular, a lo largo del cual los niveles de coilina permanecen constantes (Andrade, Tan et al. 1993). Todos estos datos sugieren que el estado de fosforilación de la coilina juega un papel fundamental para su correcta localización en los cuerpos de Cajal así como para la propia organización de estos

orgánulos. De igual modo, el factor de *splicing* YT521-B, que se localiza en la región del nucleoplasma denominada YT, ve alterada su localización nuclear tras la fosforilación en diferentes residuos de tirosina, lo que además provoca un cambio en los patrones de *splicing* de diferentes minigenes reporteros (Rafalska, Zhang et al. 2004).

#### **4.3.2 Otras modificaciones post-traduccionales: SUMOilación**

Otra modificación post-traducciona con un papel relevante en la localización de proteínas nucleares es la conjugación del grupo SUMO (del inglés *Small Ubiquitin-like Modifier*) a residuos de lisina mediante un enlace de tipo covalente (Hay RT 2005). La modificación mediante SUMO (SUMOilación) juega además un papel importante en procesos celulares como el transporte nuclear, el metabolismo del ADN, la regulación de la transcripción, la estructura de la cromatina o la estabilidad genómica (Hay RT, 2005; Klein HL, 2006; Watts FZ, 2007). Esta gran versatilidad de funciones difiere en gran medida de la conjugación a ubiquitina, con la que SUMO comparte una gran similitud estructural, pero que sin embargo y en la mayoría de los casos provoca la degradación de la proteína diana mediante la acción del proteosoma.

La familia de proteínas SUMO en humanos comprende tres parálogos bien caracterizados, SUMO-1, -2, -3 y una cuarta isoforma, SUMO-4 que ha sido únicamente caracterizada a nivel de secuencia de ADN y de la que se desconoce todavía si puede ser eficientemente conjugada (Hay RT, 2005; Wilkinson KA, 2010). SUMO-2 y SUMO-3 comparten una identidad de secuencia del 97%, lo que les hace ser consideradas una única familia (SUMO- 2/3). Las diferentes funciones biológicas de estos tres parálogos se basan en diferencias de secuencia en la región aminoterminal, que permite a SUMO-2/3 formar cadenas covalentes de poliSUMOilación, mientras que SUMO-1 se encuentra únicamente en forma monomérica (Geoffroy, MC et al 2009; Walker D, 2010).

La ruta de conjugación de SUMO se encuentra altamente conservada en eucariotas y se caracteriza por tres pasos bien definidos: i) la activación de SUMO, mediada por una enzima de activación (E1), ii) la transferencia de SUMO a la enzima que cataliza su conjugación al sustrato (enzima de conjugación, E2) y, iii) la modificación del sustrato, llevada a cabo por una enzima con actividad ligasa (E3). De forma diferente al sistema

de ubiquitinación, donde existen numerosas enzimas de conjugación (Hershko, A and Ciechanover, A, 1998), la maquinaria de SUMOilación depende de una única enzima de este tipo, Ubc9, lo que podría indicar un menor número de sustratos potenciales para la SUMOilación que para la ubiquitinación. Además y mientras que las vías de ubiquitinación utilizan ligasas específicas de sustrato, en la SUMOilación no existe tal especificidad, por lo que Ubc9 cobra aún una mayor importancia dentro de esta ruta de modificación post-transcripcional. Ubc9 interactúa directamente con el sustrato uniéndose a la secuencia consenso  $\psi$ KXD/E, donde  $\psi$  es un residuo hidrofóbico de gran tamaño, X puede ser cualquier aminoácido y K la lisina a modificar. Existen otras dos secuencias consenso, algo más extensas, basadas en la presencia de una serie de cargas electrostáticas negativas debido a la acumulación de residuos ácidos (NDSM, *negatively charged aminoacid-dependent sumoylation motif*) (Yang et al 2006) o eventos de fosforilación (PDSM, *phosphorylation-dependent sumoylation motif*) (Hietakangas et al 2006) tras el motivo mínimo de SUMOilación.

La SUMOilación es un proceso reversible, estableciéndose un equilibrio entre sustratos modificados y no modificados en condiciones fisiológicas. La ruptura de la unión a SUMO es llevada a cabo por diferentes proteasas: la familia de las proteasas SUSP (*SUMO-specific protease*) a nivel citoplasmático y las proteasas de la familia SENP (*sentrin protease*) a nivel nuclear. Éstas últimas presentan diferentes localizaciones dentro del núcleo, lo que podría estar relacionado con la especificidad de algunas SENPs para los diferentes parálogos de SUMO (Di Bacco A et al, 2006; Gong, L, Yeh ET, 2006).

La SUMOilación tiene un papel relevante en la correcta localización de proteínas nucleares. Así por ejemplo, la conjugación covalente de SUMO a la proteína PML parece ser crucial para la correcta formación de los cuerpos PML, ya que mutantes de PML carentes de señales de SUMOilación son incapaces de ensamblarse en estos orgánulos. El mecanismo por el que esta modificación es capaz de regular la organización de los cuerpos PML depende de un motivo de interacción con SUMO o SIM (del inglés *SUMO interaction motif*). Estos motivos consisten típicamente en un centro hidrofóbico (V/I-x-V/I-V/I) flanqueado por residuos ácidos que permiten la unión no covalente a las diferentes formas de SUMO actuando como motivo de anclaje para potenciar la unión a otras proteínas SUMOiladas (Hecker, CM et al 2006; Song J et al 2004). De este modo, en las proteínas PML los SIMs median la unión no covalente

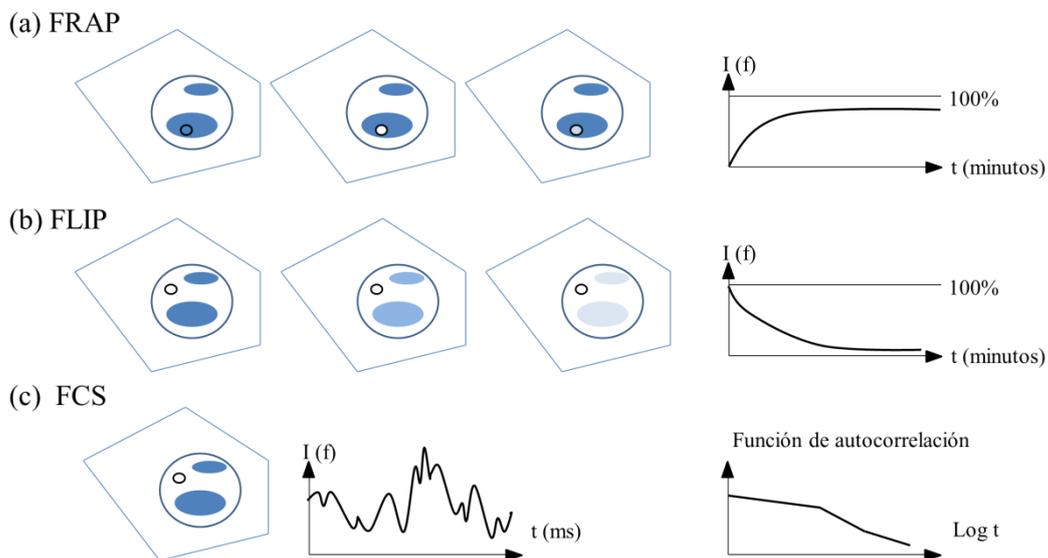
con otras proteínas SUMOiladas incluida ella misma formándose una red macromolecular a la que se pueden unir otras moléculas dependiendo de la presencia de SUMO o un motivo tipo SIM en ellas (Shen TH et al 2006; Takahashi H, et al., 2005).

Otros ejemplos de localización nuclear dependiente del estado de SUMOilación son la quinasa HIPK2 y los factores transcripcionales Daxx y CREB. HIPK2 y Daxx presentan en su secuencia motivos SIM de los que depende su localización en los cuerpos PML (de la Vega et al., 2011; Lin, DY et al 2006). La unión covalente a SUMO modifica de igual modo la localización de CREB en los cuerpos PML al mismo tiempo que regula su dinámica en el nucleoplasma (Ryan, CM et al 2010).

## 5. Dinámica de proteínas y ARN en el núcleo. Estudio de la expresión génica en célula viva

### 5.1 Introducción

El uso conjunto de proteínas fluorescentes, principalmente la proteína GFP (del inglés *green fluorescent protein*), técnicas basadas en la eliminación de fluorescencia y análisis de su posterior recuperación (en inglés y a partir de ahora técnicas de *photobleaching*: *FRAP*, *FLIP*, y *FCS*) (Figura I-9 y Tabla I-2) y técnicas de transmisión de fluorescencia (*FRET*, *FRET-FLIM*) (Tabla I-2) ha permitido en los últimos años el estudio en célula viva de la dinámica y movimiento de las proteínas y ácidos nucleicos en el interior del núcleo. Ambas metodologías ofrecen grandes ventajas respecto a las técnicas *in vitro*, ya que estas últimas no tienen en cuenta el alto contenido en ADN, ARN y proteínas que existe en el ambiente nuclear. Es precisamente esta elevada densidad molecular la que hace pensar *a priori* en el núcleo como un escenario altamente viscoso, masificado, en el que el movimiento de los diferentes componentes se debería principalmente a mecanismos activos y específicos de transporte.



**Figura I-9. Representación esquemática de los métodos usados para el estudio del movimiento de moléculas fluorescentes.** (a) **FRAP**, *fluorescence recovery after photobleaching*, (b) **FLIP**, *fluorescence loss in photobleaching*, (c) **FCS**, *fluorescence correlation spectroscopy*. Los detalles de cada una de estas técnicas se explican brevemente en la Tabla I-2.

## Tabla I-2. Técnicas de *photobleaching* - Definición y aplicaciones

Las diferentes técnicas que se explican a continuación se basan en la expresión transitoria o estable de las diferentes proteínas de estudio marcadas de forma fluorescente. En la mayoría de los casos, y debido a su estabilidad, la proteína fluorescente usada es la GFP.

---

<p><b>FRAP</b> (Fluorescence recovery after photobleaching)</p>	<p>En un experimento típico de FRAP la fluorescencia en una región de interés de la célula es eliminada mediante un breve pulso de luz láser de alta intensidad. A continuación, se realiza la medida de la recuperación de la fluorescencia a lo largo del tiempo en esa región de interés.</p> <p>La recuperación de la fluorescencia es debida al intercambio de moléculas de la zona de estudio (que han perdido su fluorescencia) por nuevas moléculas fluorescentes que proceden de otras regiones de la célula. Las tasas de recuperación de fluorescencia indican cómo de rápido se mueve la proteína de estudio. Esta movilidad depende de su capacidad de difusión y principalmente de las interacciones que establezca con otros componentes.</p> <p>Uno de los datos obtenidos a partir de las curvas de recuperación de los experimentos de FRAP es el “tiempo medio de recuperación”, tiempo en el que se recupera la mitad de la fluorescencia inicial. Para la mayoría de las proteínas nucleares (de especial interés en esta tesis) se observan dos poblaciones distintas; una “población móvil” que es reemplazada por nuevas moléculas fluorescentes, y una “población inmóvil” de moléculas fuertemente asociadas a otros componentes y que no son reemplazadas a lo largo del experimento.</p>
<p><b>FLIP</b> (Fluorescence loss in photobleaching)</p>	<p>Los experimentos de FLIP se basan en la eliminación de la fluorescencia de manera continuada en una zona concreta de la célula mientras se miden los niveles de fluorescencia en compartimentos cercanos a la zona de estudio. Si entre ambas zonas existe intercambio proteico el análisis reflejará pérdida de fluorescencia. El análisis de los datos obtenidos mediante FLIP proporciona gran información acerca de cómo las moléculas abandonan una determinada región celular y las posibles conexiones entre compartimentos celulares.</p>
<p><b>FCS</b> (Fluorescence correlation spectroscopy)</p>	<p>En este caso una pequeña área de la célula es iluminada con luz de baja intensidad. Los cambios en los niveles de fluorescencia de este punto son analizados a lo largo del tiempo con una gran rapidez y sensibilidad. Las fluctuaciones en la intensidad de la fluorescencia contienen información de la movilidad de las moléculas en el área de estudio. Cuanto más móviles sean dichas moléculas más frecuentes y cortas serán las fluctuaciones.</p> <p>Los tiempos de observación en los experimentos de FCS varían entre 200 ns y 3500 ms, tiempo suficiente para provocar estas pequeñas variaciones en los cambios de intensidad de fluorescencia y proporcionar curvas de autocorrelación.</p>
<p><b>FRET</b> (Fluorescence resonance energy transfer)</p>	<p>FRET es un proceso que tiene lugar cuando dos proteínas marcadas con dos fluoróforos diferentes con espectros de emisión y absorción solapantes se encuentran extremadamente cerca (entre 1 y 10 nm) y en una orientación apropiada. Una aproximación para conocer si existe FRET, lo cual es indicativo de que existe interacción entre dos proteínas, es eliminar de forma permanente la fluorescencia de la molécula que está actuando como aceptor y comparando la emisión del donante antes y después de esta eliminación. Si existe FRET entre ambas moléculas se produce un aumento de la emisión de fluorescencia en la región de estudio.</p>
<p><b>FRET/FLIM</b> (Fluorescence resonance energy transfer/Fluorescence life time imaging microscopy)</p>	<p>Uno de los inconvenientes de la técnica FRET es que tiene que ser llevada a cabo en célula fijada si las proteínas de estudio son móviles, ya que la reorganización de la proteína que actúa como aceptor puede interferir en la intensidad de emisión de fluorescencia del emisor. Para solventar este problema, y poder estudiar una posible interacción de dos proteínas <i>in vivo</i>, la técnica FRET/FLIM estudia la cinética de decaimiento de la emisión de fluorescencia del donante, en lugar de su intensidad absoluta, sin necesidad de una eliminación irreversible de la misma.</p>

---

Un gran número de proteínas nucleares, funcionalmente no relacionadas, exhiben un rápido movimiento generalizado dentro del núcleo que podría estar basado en procesos de difusión (Phair and Misteli, 2000). El análisis de proteínas como la GFP o sustancias orgánicas como el dextrano, que no establecen interacciones específicas con el ADN ni con proteínas, ha demostrado que el nucleoplasma es permisivo a un rápido movimiento por difusión, que se ve únicamente limitado por el impedimento estérico que representan la cromatina y los orgánulos nucleares (Görisch et al 2005; Misteli, 2001; Seksek et al 1997). Este mecanismo, caracterizado por ser independiente de energía y no direccional (Phair and Misteli 2000; Pederson T, 2000; Shopland and Lawrence, 2000), es suficiente para asegurar la correcta distribución de las proteínas a través del núcleo. De esta manera, los coeficientes de difusión medidos para diferentes factores de transcripción varían desde  $0.5 \mu\text{m}^2\text{s}^{-1}$  a  $5 \mu\text{m}^2\text{s}^{-1}$ , velocidades que permiten a una molécula recorrer por completo el volumen nuclear en pocos minutos sin ningún gasto de energía. No hay que descartar, sin embargo, que las proteínas puedan ser vehiculizadas de manera activa mediante mecanismos directos de transporte.

El movimiento de ciertos ARNs, como por ejemplo moléculas de ARN poly(A), está basado de igual modo en procesos de difusión no dependientes de energía que permiten una rápida distribución de estas moléculas sin una dirección predeterminada dentro del núcleo (Ritland-Politz, JC et al 2006). Ejemplos claros de este movimiento proceden de estudios que demuestran cómo moléculas de ARN recién sintetizadas a partir de genes localizados cerca de la envuelta nuclear difunden hacia todas las direcciones posibles y no solo hacia los poros nucleares (Ritland-Politz, JC et al 2006; Shav-Tal et al 2004). Sin embargo, existen estudios que sugieren que este movimiento es más complejo. Así, la dinámica espacial de las proteínas implicadas en la poliadenilación y transporte de los ARNm podría ser una combinación de un movimiento por difusión pasiva y procesos dependientes de ATP (Calapez, A et al. 2002).

A pesar de la capacidad para difundir libremente por el nucleoplasma, la mayoría de las proteínas nucleares presentan un movimiento menor de aquel que tendrían si sobre ellas actuase únicamente el proceso de difusión. La razón para este retardo en el movimiento viene dada por la incorporación de las proteínas en complejos macromoleculares y por las interacciones que establecen con componentes inmóviles (Misteli T, 2001). De este modo, en el interior del núcleo las proteínas estarían escaneando continuamente el nucleoplasma en búsqueda de sitios de unión en el genoma o en componentes de las

diferentes maquinarias del metabolismo nuclear. Puesto que se trata de un movimiento estocástico (no determinista, involucrando aleatoriedad), las proteínas pueden establecer en primer lugar interacciones débiles, transitorias, hasta que finalmente se asocian de manera estable en sus sitios de unión específicos (Misteli T, 2001). Este modelo, denominado en inglés “*stop and go*”, se encuentra apoyado por la dinámica espacial observada para diferentes factores de transcripción mediante estudios de *photobleaching*. Así, se ha demostrado que el tiempo que residen los factores de transcripción en sus sitios de unión específicos en el ADN es de unos pocos segundos, mientras que la unión con sitios no específicos es mucho menor (Gorski et al. 2006; Phair et al. 2004; Sprague et al 2006). Estas interacciones altamente transitorias y dinámicas son las responsables de que los coeficientes de difusión de la mayoría de estos factores sean de 10 a 100 veces menores de lo esperado en base a su tamaño y estructura (Sprague et al. 2004; Mueller et al 2008). En principio este sistema de búsqueda aleatorio parece ser bastante ineficiente, pero teniendo en cuenta que la mayoría de los factores de transcripción presentan varios miles de copias es posible generar las interacciones suficientes como para establecer y mantener el control transcripcional (Hager GL et al 2009).

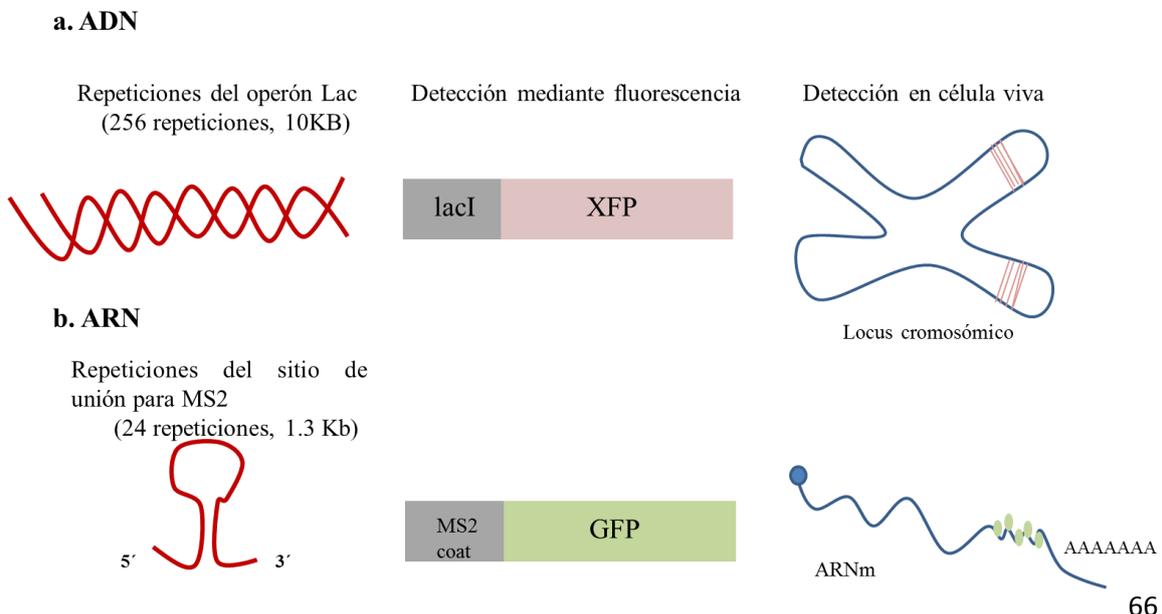
## 5.2 Dinámica espacial de la maquinaria de *splicing*

El movimiento aleatorio mencionado anteriormente en el interior del núcleo ha sido corroborado para numerosos componentes de la maquinaria de *splicing* (Rino, J et al 2007). Estudios de visualización en célula viva demuestran que el intercambio de los componentes del *spliceosoma* entre los *speckles* nucleares y el nucleoplasma es continuo, siendo incluso mayor en condiciones de inhibición de los procesos de transcripción y *splicing*. Bajo estas condiciones, los factores de *splicing* presentan una mayor movilidad debido probablemente a la ausencia de moléculas de pre-ARNm sobre las que establecer nuevas interacciones. Rino y colaboradores proponen un modelo en el que los factores de procesamiento del ARN difunden constantemente por el núcleo e interaccionan de manera aleatoria tanto con los componentes de los *speckles* nucleares como con el pre-ARNm recién sintetizados sin necesidad de una señal específica como la existencia de transcripción activa. Ejemplos de estas interacciones y de la ya mencionada disminución que éstas ejercen en el proceso de difusión han sido demostrados para componentes específicos del *spliceosoma* (Rino J, et al 2008). Experimentos de FRAP revelan que SF1 y ambas subunidades de U2AF (U2AF<sup>65</sup> y U2AF<sup>35</sup>) se mueven por el núcleo de forma más lenta a la esperada a como lo harían de manera individual, estando su movimiento determinado por la unión directa entre las proteínas. Mediante experimentos de FRET se ha demostrado *in vivo* la interacción de U2AF<sup>65</sup> con SF1 y U2AF<sup>35</sup>, tanto en el nucleoplasma como los *speckles*. Esta interacción, que persiste en condiciones de inhibición transcripcional, sugiere la presencia de complejos preformados entre diferentes factores de *splicing* antes de la formación del complejo activo del *spliceosoma* (Rino J, et al 2008). De hecho, se ha demostrado también que diferentes proteínas SR interaccionan con los factores de *splicing* U1-70K y U2AF<sup>35</sup> en diferentes compartimentos nucleares y en ausencia de transcripción activa, apoyando fuertemente la idea de una asociación previa de los factores de *splicing* antes de formar parte de la maquinaria co-transcripcional (Ellis, JD et al 2008). Más ejemplos de retardo en el movimiento provocado por interacciones en el nucleoplasma se observan para las diferentes proteínas mayoritarias de los snRNP (Huranová, M et al 2010). El estudio de la dinámica de estos factores ha demostrado que los componentes de estas partículas ribonucleoproteicas muestran dos fracciones con movilidad claramente diferenciada; una fracción lenta y una fracción rápida,

aparentemente no incorporada al complejo snRNP, característica observada también en otros complejos macromoleculares (Schmidt et al 2009). Estudios de inhibición de los procesos de transcripción y *splicing* han demostrado que las partículas snRNP interaccionan preferentemente con moléculas de ARNm en el nucleoplasma, siendo éste el principal responsable del enlentecimiento de su movilidad (Huranová, M et al 2010).

### 5.3 Estudio in vivo del ADN y ARN: dinámica espacial de la transcripción, el *splicing* y el acoplamiento entre ambos procesos

El estudio de la dinámica de regiones cromosómicas concretas, del movimiento de las moléculas de ARN por el nucleoplasma o de interacciones específicas entre proteínas y secuencias de ADN y ARN ha tenido un gran avance en los últimos años gracias al desarrollo de técnicas de marcaje de ácidos nucleicos. El uso de proteínas fluorescentes que interactúan con gran afinidad con secuencias específicas de ácidos nucleicos permite la visualización de éstas últimas mediante microscopía óptica convencional en célula viva. En el caso de regiones cromosómicas, un sistema ampliamente usado es la integración de numerosas repeticiones en tándem del operon Lac (*lacO*) dentro de la región a estudiar. La secuencia integrada es revelada mediante la expresión de una versión fluorescente del represor Lac (*lacI*) que se une con gran afinidad a estas repeticiones (Figura I-10). Gracias a esta estrategia se he podido estudiar la dinámica y posicionamiento de determinadas cromosomas en célula viva (Janicki, SM et al, 2004; Tsukamoto, T et al 2000; Chubb, JR et al 2002).



**Figura I-10 (página anterior). Técnicas de marcaje para la detección de ácidos nucleicos.** (a) Detección de ADN mediante la inserción de múltiples repeticiones de la secuencia del operon Lac (LacO) en una región cromosómica determinada. Una proteína fluorescente es entonces anclada artificialmente a esta secuencia cromosómica mediante su fusión al represor Lac (LacI), que se une de forma altamente específica a las repeticiones insertadas, permitiendo de esta manera la detección en célula viva del locus que ha sido modificado. XFP; cualquier proteína fluorescente. (b) Detección de ARN usando una proteína de fusión que contiene la proteína verde fluorescente (GFP) y la proteínas de la cápside viral del fago MS2, la cual presenta una elevada especificidad de unión por un secuencia de unión a ARN derivada del propio genoma del fago. De esta manera, moléculas de ARN que contienen repeticiones de estos motivos de unión pueden ser detectadas en célula viva.

Para la visualización y estudio del comportamiento de las moléculas de ARN se han empleado diversas estrategias basadas en el uso de sondas o proteínas fluorescentes capaces de unirse a una secuencia diana concreta dentro del ARN (Politz, JC et al 1998; Carmo-Fonseca et al 1991; Molenaar, C. et al 2004). Uno de los grandes inconvenientes de estas estrategias se encuentra en la estabilidad de los complejos formados y que el comportamiento de las moléculas de ARN estudiadas se podría ver alterado debido a la posible formación de estructuras de doble cadena. Además, la detección de transcritos específicos se ve dificultada con el uso de proteínas y sondas que detectan poblaciones completas de ARN (por ej., sondas oligo-dT frente a ARN poly(A)). Estas limitaciones han sido superadas mediante el desarrollo de un sistema altamente específico y estable que permite el reconocimiento de la molécula de ARN de interés sin alterar la secuencia codificante de la misma. Este sistema, que además puede ofrecer una resolución de una única molécula (Fusco, D. et al 2003), se basa en la integración de una serie de aptámeros de ARN en la secuencia de interés que permiten su reconocimiento mediante una proteína de unión a dicha secuencia marcada de forma fluorescente (Figura I- 10). El marcaje, por ejemplo, de moléculas de ARN con 24 sitios de unión para la proteína de la cápside del fago MS2 ha demostrado que este sistema ofrece una excelente sensibilidad en célula viva (Fusco, D et al. 2003; Shav-Tal et al 2004; Boireau, S. et al. 2007; Darzacq, X et al 2007).

La interacción que se establece entre la secuencia insertada y las proteínas marcadas de forma fluorescente es altamente estable y permite a la molécula de ARN tener el mismo comportamiento respecto a la molécula endógena (Shav-Tal et al 2004). La combinación de los sistemas *lacO-lacI* y MS2 integrados en células U2OS permitió la visualización

conjunta de los cambios que tienen lugar en la cromatina tras el inicio de la transcripción y los transcritos recién sintetizados a partir de ella (Janicki, SM et al 2004). Este sistema permite la visualización de un locus concreto antes y después de la activación de la transcripción, pudiéndose observar la apertura de la cromatina y el reclutamiento de las maquinarias de remodelación y transcripción.

La aplicación de las técnicas de *photobleaching* a estos sistemas de marcaje ha hecho posible el estudio de la dinámica de reclutamiento de la ARN polIII y diferentes factores de transcripción en los *loci* o transcritos marcados. Además y gracias a estas aproximaciones es posible el análisis de los diferentes pasos de la transcripción (inicio, elongación, terminación) o del *splicing*, ya que la movilidad de los diferentes complejos formados a lo largo de estos procesos se relaciona con la cinética de sus diferentes etapas. En este contexto, los estudios de FRAP han revelado tasas de elongación para la ARN polIII que varían desde 0.4 kb/min a 4.3 kb/min (Kimura et al. 2002; Boireau et al. 2007; Yao et al. 2007; Darzacq et al. 2007). Estas tasas dan una idea del tiempo que la ARN polIII permanece a lo largo del gen que está siendo transcrito (tiempo de residencia) más que del tiempo de la fase de elongación. La posibilidad de una fase de inicio excesivamente prolongada o la existencia de señales de parada (*pausing*) a lo largo del gen, aumenta el tiempo de residencia de la polimerasa dando lugar a tasas aparentes de elongación algo más lentas (Kimura et al 2002; Dundr et al. 2002; Boireau et al. 2007; Darzacq et al 2007). Por ejemplo, Singer y colaboradores concluyen que en su sistema de estudio la tasa de elongación real es de 4.3 kb/min, aunque ésta llega a ser de 0.4 kb/min (tasa de elongación aparente) si se tienen en cuenta los eventos de parada transcripcional a los que se ve sometida la polimerasa a lo largo del gen (Darzacq et al. 2007). La forma en la que la ARN polIII es reclutada hasta el promotor de los genes también ha podido ser estudiada gracia a estos sistemas. En este sentido, el inicio de la transcripción ha sido definido por algunos autores como un proceso altamente ineficiente, en el que aproximadamente sólo un 1% de las moléculas de ARN polIII que son reclutadas al promotor continúan en la etapa de elongación, mientras que la población restante sufre ciclos continuos de asociación y disociación al promotor (Darzacq et al 2007). Sin embargo, otros modelos basados en el gen del VIH-1 y que de igual modo utilizan el sistema basado en las repeticiones de MS2 han demostrado que la mayoría de las moléculas de polimerasa se encuentran implicadas en la fase de elongación de la transcripción, siendo ésta definida como un proceso altamente eficiente

(Boireau et al. 2007). De forma consistente con esta observación, estudios bioquímicos demuestran que el promotor del VIH-1 se encuentra constantemente ocupado por la polimerasa (Jeang et al. 1999; Marcello et al. 2001). Datos contradictorios con lo anterior proceden de experimentos de RT-PCR que establecen una tasa de elongación para la ARN polIII de 3.8 kb/min para genes endógenos humanos (Singh and Padgett, 2009) sin la existencia de fenómenos de parada transcripcional. Además, se establece que estas tasas de elongación se mantienen constantes a lo largo de genes ricos en secuencias intrónicas, donde el acoplamiento entre los procesos de transcripción y *splicing* no parece tener un efecto de enlentecimiento de la velocidad de la polimerasa. (Estas divergencias en los datos obtenidos mediante experimentos de FRAP serán tratadas más en detalle en el apartado 5.4).

En el mismo contexto que los estudios de FRAP anteriormente citados, se ha llevado a cabo el análisis de la dinámica del complejo de elongación de la transcripción P-TEFb en la trans-activación mediada por Tat del gen del VIH-1 (Molle D et al 2007). Estos estudios sugieren que el mecanismo por el que Tat contribuye a la trans-activación mediada por P-TEFb es a través de la estabilización de la quinasa CDK9 en el sitio de transcripción, aumentando su tiempo de residencia en casi 10 veces.

En cuanto a la aplicación de las técnicas de *photobleaching* en el estudio del proceso de *splicing*, recientes experimentos en célula viva han demostrado que los diferentes snRNPs presentan tiempos de residencia variables. Así, las partículas U1 y U4/U6 snRNP interaccionan con el pre-ARNm de manera transitoria (<1 s), mientras que las partículas U2 y U5 presentan tiempos de residencia mucho mayores (15-30s), lo que apoya un modelo de ensamblaje secuencial del *spliceosoma* (Huranová M et al. 2010). De estos datos se puede concluir que a diferencia de estudios previos que proponían un tiempo de 0.5 a 10 minutos para la ejecución del proceso de *splicing* (Audibert et al. 2002; Singh and Padgett, 2009), éste tiene lugar en tan solo 30 segundos en la mayoría de los casos. Esta elevada eficiencia en la reacción de *splicing* está apoyada por experimentos adicionales en los que se observa la acumulación de todos los snRNAs en el sitio de transcripción activa aunque a niveles muy débiles (Schmidt U. et al. 2011).

La co-transcripcionalidad del proceso de *splicing*, es decir la eliminación de los intrones en el transcrito naciente antes de que el proceso de transcripción finalice, ha sido

propuesta y demostrada en numerosas ocasiones en genes de mamíferos (Neugebauer, 2002; Kornblihtt, 2004; Bentley, 2005; Moore and Proudfoot, 2009). La predominancia del *splicing* co-transcripcional se ha confirmado recientemente tanto *in vitro* como *in vivo* (Schmidt U. et al. 2011; Vargas et al 2011). Sin embargo, se sabe que el *splicing* puede ocurrir también de manera independiente a la transcripción (Wetterberg et al 1996), es decir, de forma post-transcripcional. El estudio realizado por Vargas y colaboradores, mediante la visualización de una única molécula de ARN en célula viva, pone de manifiesto que a pesar de que el *splicing* constitutivo es mayoritariamente co-transcripcional, existen determinados casos en los que este proceso ocurre una vez la transcripción ha finalizado, especialmente en genes sometidos a un alto grado de regulación (Vargas et al 2011).

Gracias al desarrollo de las técnicas *in vivo* y a los sistemas de marcaje de ácidos nucleicos de los que venimos hablando, se está consiguiendo un mayor conocimiento de las características de este potencial acoplamiento. Así por ejemplo, se ha confirmado *in vivo* el reclutamiento de componentes iniciales del *spliceosoma* (U1snRNA, U1-70K y SF2ASF) a genes que están siendo activamente transcritos y en los que no tiene lugar el proceso de *splicing* (Spiluttini, B. et al 2010). Estos datos contrastan con experimentos previos de inmunoprecipitación de cromatina (ChIP) en los que no se detecta asociación entre los componentes de la partícula U1 snRNP y un gen reportero que no contiene intrones (Listerman I, Sapra AK and Neugebauer KM, 2006). Las diferencias observadas podrían deberse a cuestiones técnicas, poniendo de manifiesto que los métodos de imagen son mucho menos sensibles a la distancia entre la proteína o sonda de ARN empleada y la secuencia génica con la que se pretende estudiar una posible asociación. La asociación de U1 snRNP de manera independiente al proceso de *splicing*, junto con las múltiples interacciones que establece con la maquinaria de transcripción y procesamiento apoyan un importante papel de este factor en el acoplamiento entre ambos procesos favoreciendo la actividad de los complejos de transcripción y maduración de los pre-ARNm. La presencia de U1 snRNP pre-ensamblado en el complejo de transcripción podría además facilitar el reconocimiento de los sitios 5' de *splicing*.

Junto con U1 snRNP, otros factores de la maquinaria de *splicing* influyen notablemente en el proceso de transcripción. Por ejemplo, SC35 juega un importante papel en la elongación transcripcional incluso antes de formar parte del *spliceosoma* (Lin et al.

2008). El aumento de la eficiencia de la transcripción gracias a la existencia de secuencias intrónicas (Brinster RL et al. 1988; Palmiter RD, et al. 1991; Lacy-Hulbert, A. et al 2001; Lu S and Cullen BR, 2003) y al efecto que ejercen los sitios de *splicing* 5'proximales en el reclutamiento de factores de transcripción demuestran además un claro efecto del proceso de *splicing* sobre la transcripción. Esta cuestión ha sido abordada de forma muy elegante en recientes estudios realizados en célula viva (BentoMartins S et al. 2011; Brody Y et al 2011) que demuestran que el ensamblaje co-transcripcional de la maquinaria de *splicing* no produce cambios en las tasas de elongación de la ARN polII. La inhibición general del procesamiento de los pre-ARNm no altera la tasa de transcripción de los genes incluidos en estos estudios, si bien el uso del inhibidor general de *splicing* SSA (*spliceostatina*) provoca una rápida disociación de la ARN polII del sitio de transcripción activa impidiendo el mecanismo de control que ésta ejerce sobre las moléculas de pre-ARNm que no han sido procesadas mediante su parada en el extremo 3' de los transcritos recién sintetizados (BentoMartins S et al. 2011).

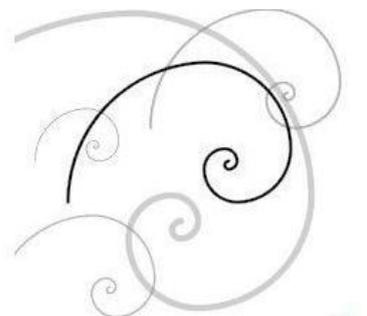
#### **5.4 Limitaciones de los estudios in vivo; sin consenso en el modelo cinético**

Las divergencias descritas anteriormente en los datos obtenidos a partir de los estudios de FRAP no se limitan a la ARN polII. Así por ejemplo, el tiempo de residencia estimado para la unión a la cromatina de diferentes factores de transcripción puede variar desde milisegundos hasta varios segundos dependiendo del estudio. De la misma manera, el estudio de la unión a la cromatina de la histona H1 ofrece datos contradictorios (Mueller F et al. 2010). Esta variabilidad podría deberse a diferencias biológicas, como el empleo de diferentes transgenes o líneas celulares, así como a variaciones técnicas experimentales. Sin embargo, parece ser que la principal razón se encuentra en los parámetros incluidos en los modelos cinéticos de análisis (Mueller F et al. 2010). Los datos experimentales obtenidos a partir de la técnica de FRAP son analizados a través de un modelo cinético determinado en el que se incluyen diferentes parámetros relativos a la forma en el que se ha llevado a cabo la eliminación de la fluorescencia, a la existencia o no de difusión y a las características de unión de la proteína en el área de estudio donde se ha llevado a cabo la eliminación de la fluorescencia. A partir del modelo empleado se genera una curva de FRAP que debe

ajustarse lo máximo posible a los datos experimentales obtenidos y que ofrece valores cuantitativos como el coeficiente de difusión, el tiempo de residencia o el tiempo transcurrido entre dos eventos de unión diferentes. La inclusión de diferentes combinaciones de estos parámetros en los modelos cinéticos puede dar lugar a curvas que se ajusten correctamente a los mismos datos experimentales, pero que sin embargo conducen a interpretaciones biológicas muy diferentes.

La “validación cruzada” de los resultados obtenidos en diversos experimentos de FRAP ha demostrado que las diferencias observadas en el comportamiento de los factores incluidos en el estudio se debe únicamente al modelo cinético empleado para cada uno de ellos. Además, la aplicación de tres modelos cinéticos diferentes a un mismo factor de transcripción en una misma línea celular da lugar a diferentes predicciones (Mueller F. et al. 2008). Por todo esto y hasta la estandarización de los criterios de inclusión de los diferentes parámetros en los modelos cinéticos se ha propuesto la aplicación de varios modelos cinéticos a un mismo factor. Junto a esta estrategia se propone además el uso de metodologías alternativas como FCS o el estudio en molécula única que sean capaces de cubrir las limitaciones técnicas de los experimentos de FRAP (Mueller F. et al. 2008).

## *Objetivos*





Los objetivos generales de esta tesis doctoral son:

1. La identificación y caracterización de la señal de localización hacia los *speckles* nucleares de TCERG1.
2. El estudio del comportamiento dinámico de TCERG1 en célula viva.

Para ello se plantearon los siguientes objetivos concretos:

1.1 Identificar y caracterizar la secuencia necesaria para la localización de TCERG1 en los *speckles* nucleares.

1.2 Estudiar el posible papel de la señal de localización hacia los *speckles* nucleares de TCERG1 como una señal general de localización a estos dominios.

1.3 Determinar el efecto de la señal de localización de TCERG1 hacia los *speckles* nucleares en la regulación del *splicing* alternativo mediado por esta proteína.

1.4 Estudio del mecanismo molecular implicado en la localización de TCERG1 en los *speckles* nucleares.

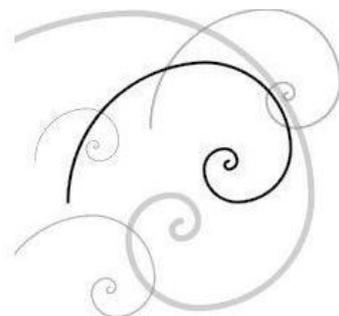
2.1 Determinar las características cinéticas de TCERG1 en los *speckles* nucleares y nucleoplasma.

2.2 Estudiar la localización de TCERG1 en sistemas de transcripción y *splicing* co-transcripcional. Determinar las características funcionales de ambas localizaciones en relación con las maquinarias de transcripción y *splicing*.

2.3 Determinar el comportamiento dinámico de TCERG1 en sistemas de transcripción y *splicing* co-transcripcional.



# *Materiales y métodos*





# 1. PLÁSMIDOS Y ANTICUERPOS

## Plásmidos

El vector pEF-ECFP y el plásmido de expresión en eucariotas pEFBOS/ECFP/T7-TCERG1[1-1098] han sido descritos en un trabajo anterior (Sánchez-Álvarez, M 2010). El vector pEFBOS/ECFP/T7-TCERG1[1-662] fue generado, mediante procedimientos estándares de clonaje, usando fragmentos con extremos EcoRI obtenidos a partir del vector original pEFBOST7-TCERG1 descrito previamente (Suñé C, 1999) e insertándolos en el vector de clonaje intermedio pEF-ECFP.

Todos los mutantes pEFBOS/ECFP/T7-TCERG1[1-662]-FF fueron construidos insertando la secuencia del correspondiente fragmento FF con extremos BstBI en el vector pEFBOS/ECFP/T7-TCERG1[1-662]. Los vectores de fusión pEGFP-FF5 y pEGFP-FF4/FF5 fueron construidos introduciendo el producto de PCR apropiado en sitios XhoI y XhoI/BamHI del vector comercial pEGFP-C1 (Invitrogen), respectivamente. El vector pEFBOS/ECFP/T7-TCERG1 $\Delta$ FF4/FF5 fue generado mediante dos pasos consecutivos de clonaje que incluyeron la introducción de la secuencia correspondiente desde FF1 hasta FF3 en un sitio AvrII y posteriormente la secuencia de FF6 con extremos MluI y XbaI en el vector pEFBOS/ECFP/T7-TCERG1[1-662]. El vector pEFBOS/GFP/T7-TCERG1 $\Delta$ FF5 fue cedido por el Dr. Mariano García Blanco (Universidad de Duke, Durham, NC). Los vectores que contiene mutaciones puntuales para dianas de SUMOilación y mutaciones puntuales en residuos de fenilalaninas fueron construidos siguiendo las instrucciones proporcionadas por el kit de mutagénesis dirigida *Quickchange II XL-Site-Directed* (Stratagene). Para ello se diseñaron parejas de oligonucleótidos específicos purificados por HPLC para obtener los siguientes cambios puntuales en el vector pEFBOS/ECFP/T7-TCERG1[1-662]-FF4/FF5:K985R(possible diana de SUMOilación en el dominio FF5); F903A, F961A, F903A/F961A, F946A y F903A/F946A, todas ellos residuos de fenilalaninas en los dominios FF4 y FF5. Las mismas parejas de oligonucleótidos se usaron para crear las mismas mutaciones puntuales en el vector pEFBOS/ECFP/T7-TCERG1[1-1098].

El vector de expresión SRSF1 y sus mutantes (RRM1 y RRM2) fueron cedidos por el Dr. Javier Cáceres (MRC, Edimburgo) y han sido descritos previamente (Cáceres and

Krainer 1993). Para la generación de los vectores pCGT7-RRM1 y pCGT7-RRM2, las secuencias de los dominios RRM1 y RRM2 fueron amplificadas mediante PCR con extremos XbaI y BamHI/XhoI e insertados en el vector de expresión de la proteína *wild type* SRSF1 pCGT7. Los mutantes de los dominios RRM y dominios FF fueron creados usando las secuencias correspondientes flanqueadas con extremos BglII amplificadas por PCR a partir del vector pEFBOS/ECFP/T7-TCERG1[1-1098] e insertados en el correspondiente plásmido pCGT7-RRM digerido con BamHI.

El minigen reportero del *splicing del gen Bcl-X* (HIV2-X2) fue cedido por el Dr. Benoit Chabot, (Universidad de Sherbrooke, Canadá).

Los vectores de expresión en bacterias pGEX2TK-FF4/FF5 y pGEX2TK-FF4/FF5-F903A/F961A se generaron amplificándolos fragmentos correspondientes a la pareja FF4/FF5 y FF4/FF5-F903/961A a partir de los vectores pEFBOS/ECFP/T7-TCERG1[1-662]-FF4/FF5 y pEFBOS/ECFP/T7-TCERG1[1-662]-FF4/FF5-F903A/961A, respectivamente. Dichos fragmentos fueron insertados en sitios BamHI y EcoRI del vector comercial pGEX2TK (Amersham).

Los vectores de expresión de las versiones de la ARN polIII resistentes a  $\alpha$ -amanitina  $wt^{RES}$  y  $\Delta CTD^{RES}$  fueron cedidos por el Dr. Edouard Bertrand y han sido descritos previamente (de la Mata et al. 2003; Fong et al. 2003).

En el apartado 6 de esta sección se describen las secuencias de todos los oligonucleótidos usados durante este trabajo de tesis.

## **Anticuerpos**

Para la detección de las proteínas expresadas con epítipo T7 se utilizó anticuerpo anti-T7 (Bethyl, A190-117A) a una dilución 1:20.000 y 1:1.000 para western blot e inmunofluorescencia, respectivamente. Se usaron también anticuerpos anti-CDK9 (Santa Cruz, sc-484) a una dilución 1:500, anti-U2AF<sup>65</sup> (proporcionado por el Dr. Juan Valcárcel, CRG, Barcelona) a una dilución 1:500, anti SC35 (Sigma, S4045) a una dilución de 1: 4.000, anti GST (Bethyl, A190-121) a una dilución 1:500 y anti NOLC1 (Abcam, ab106324). Para los análisis de western blot, los anticuerpos primarios fueron detectados usando anticuerpos secundarios frente a conejo, ratón o cabra conjugados

con HRP (todos ellos de Perkin Elmer Life Science). La dilución estándar empleada para estos anticuerpos fue de 1:5.000 y su detección se realizó mediante quimioluminiscencia (Western Lightning Chemiluminiscence Reagent Plus, Perkin Elmer). Para los estudios de inmunofluorescencia se usaron anticuerpos secundarios de cabra conjugados a los fluorocromos Alexa Fluor 488 y Alexa Fluor 647 frente a ratón y conejo respectivamente (Molecular Probes). La dilución usada para estos anticuerpos secundarios fue de 1:500 en todos los casos.

## **2. CULTIVOS CELULARES, TRANSFECCIÓN DE ADN Y TRATAMIENTOS**

Las líneas celulares HEK293T y HeLa fueron cultivadas en idénticas condiciones: 37°C en atmósfera de CO<sub>2</sub> al 5% y humedad del 95%. En ambos casos el medio utilizado fue DMEM *Low Glucose (Dulbecco's Modified Eagle Medium, Gibco)* suplementado con un 10% de suero bovino fetal (FBS, Gibco), penicilina y estreptomina como agentes antimicrobianos (100 U y 0.1 mg/ml respectivamente, (Sigma)), aminoácidos no esenciales (Gibco) y L-Glutamina a una concentración final de 4 mM (Gibco).

Para los experimentos de transfección de ADN y posterior análisis mediante *western blot* o RT-PCR (ver a continuación), se sembraron aproximadamente  $2 \times 10^6$  células en placas de 6 pocillos (Falcon) unas 20 horas antes de la transfección en un volumen de 2 ml de medio de cultivo. Con una confluencia cercana al 60-70%, el medio es reemplazado por 1.5 ml de medio fresco y las células son transfectadas con las cantidades de ADN correspondiente siguiendo el método de precipitación de Ca<sub>2</sub>PO<sub>4</sub>. Brevemente, 200 µl de Solución A (250 mM CaCl<sub>2</sub>) son mezclados con la cantidad de ADN adecuado, agitados mediante vortex y mezclados con 200 µl de la Solución B (50 mM HEPES, 1.5 mM NaHPO<sub>4</sub>, 140 mM NaCl; pH ajustado al 7.05) durante 1 minuto para ser luego añadidos al cultivo. Aproximadamente 48 horas después de la transfección las células son recogidas y procesadas para su análisis como se describe a continuación.

En los experimentos de inmunofluorescencia (ver a continuación, apartado 3), se siguió el mismo procedimiento de transfección. Sin embargo, las células fueron sembradas sobre cubreobjetos previamente depositados en las placas de 6 pocillos (Falcon). Al alcanzar una confluencia de aproximadamente el 50-60% las células son transfectadas para ser procesadas posteriormente a las 20 horas de la transfección.

Para la inhibición de la transcripción mediada por la ARN polII se utilizó  $\alpha$ -amanitina (Sigma) a una concentración de 25  $\mu\text{g/ml}$  durante las 6 horas previas al procesamiento del cultivo.

En los experimentos de hibridación *in situ* fluorescente (FISH) y estudio de la recuperación de fluorescencia (FRAP) se utilizaron células U2OS, que contienen establemente transfectado un vector reportero de la transcripción del VIH-1 con todos los elementos necesarios para la producción del ARN viral. Este reportero contiene además 24 sitios de unión para la proteína de la cápside del fago MS2, lo que permite su detección en célula viva (reportero: pExo-MS2x24; línea celular: pU2OS\_Exo1) (ver imagen de la figura R-23, A del apartado de “Resultados” para una mayor explicación de la construcción) (Boireau S et al 2007). Se utilizaron también células U2OS establemente transfectadas con un gen reportero que contiene la secuencia LTR del VIH-1 y la secuencia intrónica MINX (*wild type* o con mutaciones en el sitio *branch point*, BP) fusionada al gen *LacZ* y a la señal 3'UTR del gen de la hormona de crecimiento bovina (bGH). Además, este reportero contiene 4 sitios de unión para la proteína MS2 dentro de la secuencia intrónica MINX que permiten su detección en célula viva (reportero: MINX\_MS2in/ líneas celulares MINX, BPI y BPII, para las secuencias *wild type*, y mutaciones del sitio *branch point* I y II respectivamente) (ver imágenes de las figuras R-25 y R-26 de “Resultados” para una mayor explicación de las construcciones). Estas células fueron crecidas en medio DMEM *Low Glucose* (Gibco) suplementado con un 10% de FBS (Gibco), penicilina y estreptomina como agentes antimicrobianos (100 U y 0.1 mg/ml respectivamente, (Sigma) e higromicina 100  $\mu\text{g}/\mu\text{l}$  como agente de selección. Las condiciones de cultivo fueron las mismas que anteriormente: 37°C en atmósfera de CO<sub>2</sub> al 5% y humedad del 95%. Unas 20 horas antes de la transfección las células son sembradas sobre cubreobjetos previamente depositados en placas de 12 pocillos (Falcon) en un volumen de 1 ml de medio de cultivo para los experimentos de FISH y en placas de 35 mm de diámetro (Nunc) en un volumen de 5 ml de medio de cultivo para los experimentos de FRAP. Al alcanzar una

confluencia aproximada del 50% las células son transfectadas con las cantidades de ADN correspondiente utilizando los reactivos comerciales *GeneJuice* (Novagene) y *Lipofectamine 2000* (Invitrogen), siguiendo las instrucciones del fabricante. Aproximadamente 24 horas después de la transfección las células son procesadas para los experimentos correspondientes. Para la inhibición de la transcripción en estas células se utilizó  $\alpha$ -amanitina a 0.1  $\mu\text{g}/\mu\text{l}$  y DRB (5,6-dicloro-1- $\beta$ -ribofuranosilbenzimidazol) a 100 $\mu\text{M}$ , durante 2 horas en ambos casos, antes de su procesamiento. Para la inhibición del proceso de *splicing* las células fueron tratadas con *spliceostatina* (SSA) a 30 ng/ml (Kaida et al. 2007) durante un mínimo de 3 horas antes de su procesamiento.

### **3. INMUNOFLUORESCENCIA, FISH, FRAP Y PROCEDIMIENTOS DE MICROSCOPIA**

#### **Inmunofluorescencia y microscopía confocal**

Para los estudios de inmunofluorescencia, las líneas celulares empleadas son cultivadas en cubreobjetos y transfectadas con las cantidades de ADN correspondiente al alcanzar una confluencia aproximada del 50-60%. Aproximadamente 20 horas después de la transfección, las células son fijadas con 3.5 % de paraformaldehído (PFA) en 20 mM HEPES (pH 7.4) durante 45 minutos en hielo. Tras 3 lavados con PBS-1X, las células son permeabilizadas en una solución de PBS-1X que contiene 0.5% de TritónX-100 durante 5 minutos a temperatura ambiente y finalmente bloqueadas en una solución de 2.5% (p/v) de albúmina de suero bovino (BSA) en PBS-1X a 4°C durante al menos 16 horas. Se utiliza una misma solución de incubación y lavados de anticuerpos primarios y secundarios (0.1% de BSA en PBS-1X) con la dilución correspondiente de anticuerpo. La secuencia de incubación y lavados, todos realizados a temperatura ambiente, en cámara de humedad y con volúmenes totales de 100  $\mu\text{l}$ , es de 1 hora de incubación con el anticuerpo primario, 5 lavados de 5 minutos, 1 hora de incubación en oscuridad con el anticuerpo secundario y 3 lavados de 5 minutos. Por último se realizan 3 lavados con PBS-1X, para posteriormente montar las preparaciones utilizando como medio de montaje *Prolong Goldantifade* (Molecular Probes). Las imágenes fueron adquiridas con un microscopio confocal invertido (Leica SP2) usando el objetivo HCX PL APO CS

40.0 X 1.25 OIL UV. Las imágenes son tomadas simultáneamente en todos aquellos casos en los que se realiza un doble marcaje de inmunofluorescencia. GFP, CFP y el fluorocromo Alexa 488 son excitados con el láser de argón de 488nm, mientras que el fluorocromo Alexa 647 es excitado con el láser de helio-neón (He-Ne) HeNe de 633nm. El diámetro de la ranura confocal (*pinhole*) fue constante y en todos los casos de 1µm.

Para el estudio de la localización específica de TCERG1 en la periferia de los *speckles* (figura R-4) se realizaron 3 secciones ópticas seriadas a lo largo del plano Z de 0.3 µm de grosor cada una de ellas. La adquisición de estas imágenes fue llevada a cabo con una intensidad baja de señal.

La cuantificación mostrada en la Figura R-5 fue realizada midiendo la intensidad media por pixel de la proteína ECFP en la región de los *speckles* nucleares y en las regiones de nucleoplasma adyacentes y comparando la diferencia entre ambas intensidades. Para ello se calculó la proporción de señal presente en *speckles* y en nucleoplasma respecto al valor total de intensidad en cada una de las regiones de interés (ROI, *region of interest*) y se calculó la diferencia entre ambas proporciones. A mayor diferencia, mayor localización de la proteína en los *speckles* nucleares. El estudio de estas intensidades incluye una media de 12 células por construcción lo que equivale a 72 medidas para los *speckles* nucleares y 108 para el nucleoplasma. El análisis estadístico de los datos se llevó a cabo usando el test estándar *t-Student*. La adquisición y análisis de las imágenes para los estudios de cuantificación fue realizada mediante el software LAS AF v2.3.6 *Build* 5381sps. El procesamiento digital fue realizado usando la versión CS3 extended v10.0 del software *Adobe Photoshop*.

### ***In situ* fluorescence hybridization (FISH)**

Células U2OS son sembradas en cubreobjetos depositados previamente en placas de 12 pocillos. Al alcanzar una confluencia aproximada del 50% son transfectadas con 20 ng de un plásmido de expresión de la proteína viral Tat (CMV-Tat) y 0,5 µg del vector de expresión de EGFP-TCERG1 con el reactivo de transfección *GeneJuice* (Novagen) siguiendo las instrucciones del fabricante en medio Optimen (Gibco). A las 24 horas tras la transfección, las células son lavadas con PBS-1X (tres lavados consecutivos) y fijadas durante 30 minutos a temperatura ambiente en PFA al 4%. Tras la fijación y el

lavado con PBS-1X (tres lavados consecutivos) las células son permeabilizadas con etanol al 70% a 4°C durante al menos 16 horas. Una vez finalizada la permeabilización las células son rehidratadas con PBS-1X (tres lavados consecutivos). Durante la preparación de la solución de hibridación, cuya composición se detalla a continuación, las células son incubadas en una solución que contiene un 20% de formamida en SSC-1X (solución de hibridación: formamida al 4%, tRNA 1µg/µl, 20 ng de la sonda MS2b-Cy3 [ver secuencia de la sonda al final de la sección de “Materiales y métodos”, junto con las secuencias de los oligonucleótidos] y agua libre de ARNasa en SSC-1X). Ésta solución se calienta durante 1 minuto a 90°C y a continuación se le añade una mezcla que contiene BSA libre de ARNasa al 0,01 µg/µl, *vanadyl ribonucleoside complexes* (VRC) (Sigma) 20 mM, y dextrán sulfato al 10%. Las células adheridas a los cubreobjetos son incubadas con un volumen de 20 µl de la solución de hibridación en cámara húmeda durante 16 horas. Tras esta incubación, se realizan dos lavados de 30 minutos cada uno con una solución al 20X% de formamida en SSC-1X a 37°C. Al terminar, las células son lavadas en PBS-1X durante 2 minutos y preparadas para su observación mediante microscopía de fluorescencia en un medio que contiene 90% de glicerol, p-fenilendiamina a 1 mg/ml y DAPI a 0,1 µg/ml.

Las imágenes fueron adquiridas en un microscopio vertical de campo amplio (Leica DM6000) (objetivo 100X NA 1.4) acoplado al software *Metamorph* (Universal Imaging). El procesamiento de las imágenes fue realizado usando la versión CS3 extended v10.0 del software *Adobe Photoshop*.

### ***Fluorescence recovery after photobleaching (FRAP)***

Para los experimentos de FRAP, células U2OS son sembradas en placas de 35 mm de diámetro. Al alcanzar un 50% aproximado de confluencia se lleva a cabo la transfección con las cantidades correspondientes de los plásmidos de expresión de la proteína viral Tat (20 ng), EGFP-TCERG1 (0.5 µg) y cherry-MS2 (0.2 µg) que permite la visualización del sitio activo de transcripción (ST) del VIH-1 en estas células. En el momento de la adquisición de imágenes, unas 20 horas tras la transfección, el medio es reemplazado por medio sin rojo fenol ni riboflavina para permitir la visualización de la fluorescencia.

Durante el transcurso de los experimentos las células son mantenidas a 37°C en la cámara de célula viva incorporada a un microscopio de fluorescencia Nikon TE200. La fluorescencia de EGFP-TCERG1 en el ST activa del VIH-1, en los *speckles* y en el nucleoplasma es eliminada mediante la línea 488 del láser de argón (al 56% de intensidad) en áreas de  $0.96 \mu\text{m}^2$ . El tiempo de eliminación de la fluorescencia es de 300 milisegundos (ms). La recuperación de la fluorescencia es captada mediante la toma sucesiva de imágenes (1 imagen cada 100 ms) durante 30 segundos (s) tras la eliminación de la fluorescencia (a partir de ahora *bleaching*) con el objetivo 100X NA 1.45 y el empleo del software *Metamorph* (Universal Imaging Corp.) acoplado al microscopio. El número de imágenes captadas antes del *bleaching* es de un total de 20. El análisis de las imágenes obtenidas se realizó con el mismo *software*. A cada una de las intensidades obtenidas para cada una de las regiones en las que se ha eliminado la fluorescencia (ST, *speckles* nucleares y/o nucleoplasma) se le resta el valor de la fluorescencia de fondo y se corrige el efecto del *bleaching* dividiendo cada valor entre la intensidad total de la célula, a la que también se le sustrae el valor de fondo ( $I_x - BG / WC - BG$ ; donde  $I_x$  es la intensidad obtenida para cada región en cada imagen, BG es el valor de intensidad de fondo y WC es la intensidad total de fluorescencia de la célula en cada imagen). A partir de estos valores se calcularon las curvas de recuperación de la fluorescencia en cada uno de los compartimentos. Estas curvas fueron normalizadas estableciendo un valor 0 de intensidad a la primera imagen *post-bleach* y calculando el valor relativo de cada intensidad respecto a las 10 últimas imágenes antes del *bleaching*. El valor de la fracción móvil (F.M.) de cada curva se corresponde al promedio de los 10 últimos valores de intensidad de las curvas normalizadas a 0. La fracción inmóvil (F.I.) es igual a 1-F.M. El tiempo medio de recuperación fue leído en cada caso directamente en la curva normalizada a 0.

## 4. PROCEDIMIENTOS BÁSICOS DE BIOLOGÍA MOLECULAR

### Extracción de ARN y análisis por RT-PCR (PCR reverse transcriptase)

Para los estudios de RT-PCR, aproximadamente  $2 \times 10^6$  células son sembradas en placas de 6 pocillos 24 horas antes de la transfección. Las células son transfectadas con 0.6  $\mu\text{g}$  del minigen reportero HIV2-X2 y 1  $\mu\text{g}$  del plásmido pEFBOST7-TCERG1.

El ARN total fue extraído mediante TRIzol (Invitrogen) según las indicaciones de la casa comercial. Aproximadamente 1  $\mu\text{g}$  de ARN total fue digerido con 10 U de DNAsa (*DNase I*, Roche) durante 30 minutos a 37°C en un volumen de 10  $\mu\text{l}$ . La mitad del volumen de esta digestión se retro-transcribe con M-MLVreverse-transcriptase (Invitrogen) durante 1 hora a 37°C, según indicaciones de la casa comercial y utilizando el oligonucleótido *RT-Sveda* para el minigen HIV2-X2. Un décimo de la reacción se amplificó mediante PCR convencional usando los oligonucleótidos *X34* y *XAgeI* (ver apartado 6, “Oligonucleótidos”, al final de “Materiales y Métodos”). Tras 30 ciclos de PCR ( $T_m$  60°C), los productos fueron resueltos en un gel de agarosa al 2%. La intensidad de las bandas fue cuantificada usando el software *Quantity One* (BioRad). Todos los experimentos presentados se realizaron un mínimo de tres veces. El análisis estadístico se realizó con el programa *Prism 5.0 software* (GraphPad). El test Two-tailed Student se utilizó para comparar medias entre muestras. Los valores P estadísticamente significativos se representan en las gráficas mediante asteriscos (\*,  $p < 0.05$ ).

## 5. MANIPULACIÓN Y ANÁLISIS DE PROTEÍNAS

### Preparación de extractos celulares

Transcurrido el tiempo correspondiente tras la transfección del cultivo celular las células son recogidas de la placa de cultivo y centrifugadas durante 5 minutos a 1200 rpm a 4°C. El precipitado obtenido se lisa con tampón de lisis T7 frío (20 mM HEPES, pH 7.9, 150 mM NaCl, 5 mM EDTA, 1% NP-40, 1 mM ditioneitol (DTT) y 1 mM

fenilmetil-sulfonil-fluoruro (PMSF)) durante 30 minutos a 4 °C. Tras este tiempo, el volumen de lisado es centrifugado a 13.200 rpm en microfuga durante 5 minutos y el sobrenadante es utilizado como extracto total. La concentración de proteína en el extracto se mide mediante el ensayo de Bradford (BioRad) a 595 nm.

### **Análisis por western blot**

Las proteínas de los extractos fueron separadas en geles al 10% SDS-PAGE. Las diferentes muestras a analizar se diluyen en tampón de carga Laemmli 2X (0.2% azul de bromofenol, 10% de glicerol, 0.1 M TrisHCl pH 6.8, 2% dodecil sulfato sódico (SDS), 50 mM DTT) y se calientan durante 5 minutos a 95°C. Tras su carga en el gel, la electroforesis se efectúa en buffer Laemmli (25 mM Tris, 192 mM de glicina y 0.1% SDS) a un voltaje continuo aproximado de 150 V. Desde los geles, las proteínas son transferidas a membranas de nitrocelulosa (Amersham Biosciences) mediante un sistema húmedo de transferencia proteica en una solución que contiene 25 mM Tris, 192 mM de glicina y 20% (v/v) de metanol a 300 mA de intensidad durante 1 hora. El análisis del western blot se realiza en una misma solución de bloqueo (durante 1 hora), incubación y lavado (5% (p/v) leche desnatada, PBS-1X, 0.1% (v/v) Tween-20). Las condiciones, diluciones y tiempos de incubación se ajustaron a cada anticuerpo.

Para el análisis de la expresión de los mutantes de SRSF1 (Figura R-9) las proteínas fueron separadas mediante geles al 15% SDS-PAGE y transferidas a membranas de polivinilo (PVDF, BioRad). El voltaje aplicado para la separación de las proteínas fue de 75 V.

Como marcadores del peso molecular se usó una preparación comercial pre-teñida con un rango de 175 KD a 6.25 KD (New England Biolabs).

### **Producción y purificación de proteínas recombinantes en *E.coli***

En la expresión heteróloga de proteínas recombinantes para su posterior purificación se utilizaron bacterias *E. coli* deficientes en proteasas (BL21 pLys).

En la producción de las proteínas recombinantes usadas en los experimentos de *pull-down* (GST, GST-FF4/FF5 y GST-FF4/FF5 F903A, F961A) se ha usado un sistema de cromatografía por afinidad en condiciones nativas de glutatión-S-transferasa (GST)/glutathione. Un volumen de precultivo en medio LB de 1/20 del volumen de cultivo final es inoculado en presencia de ampicilina y cloramfenicol durante aproximadamente 16 horas a 37°C. El precultivo es entonces diluido en un volumen final de 250 ml de LB fresco (con ampicilina) e incubado a 37°C hasta alcanzar una densidad óptica de 0.6. En este momento se induce la producción de proteína añadiendo isopropil-1-tio-β,D-galactopiranosido (IPTG, USB biochemicals) a una concentración final en el cultivo de 0.4 mM. El tiempo de inducción para las diferentes construcciones es de 2 horas. Transcurrido este tiempo se recogen las bacterias mediante centrifugación a 6000 rpm y se procede a su congelación o a su procesamiento de manera directa. Los pellets de bacterias son resuspendidos en 1/100 volúmenes, con respecto al cultivo de origen, de PBS-1X suplementado con un cóctel de inhibidores de proteasas (*Complete*, Roche). La pared bacteriana es lisada mediante la adición de lisozima a una concentración final de 1.5 mg/ml, agitación mediante vórtex e incubación a temperatura ambiente durante 5 minutos. Seguidamente, el ADN genómico se fragmenta mediante la adición de 130 U de DNAsI (Sigma), suplementada con MgCl<sub>2</sub> y MnCl<sub>2</sub> a unas concentraciones finales de 10 mM y 1 mM respectivamente, e incubando a 37°C en agitación durante 30 minutos. Por último se añade Tritón X-100, Tween-20, DTT y PMSF a unas concentraciones finales de 1% (v/v), 1% (v/v), 1 mM y 0.5 mM respectivamente, suplementando con PBS-10X para mantener una concentración final de PBS-1X. El posterior aclarado del lisado se realiza mediante centrifugación a 14000 rpm durante 30 minutos.

La resina liofilizada de agarosa acoplada a glutatión (Glutathione-Agarose, Sigma) se prepara, según las especificaciones de la casa comercial, en un volumen correspondiente a 1/1000 respecto al volumen de cultivo de origen. Los lisados obtenidos anteriormente son incubados con la resina en rotación durante un mínimo de 2 horas a 4°C. Transcurrido este tiempo se procede al lavado de la resina con un tampón que contiene PBS-1X, 1% Tritón X-100, 1% Tween-20, 0.5 mM PMSF, 0.5 mM DTT) 3 veces, 5 minutos en rotación a 4 °C.

La elución de las proteínas unidas a la resina se hace por competición en un volumen equivalente al volumen de la resina, en un tampón de elución que contiene 100 mM Tris

HCl, pH 8, 150 mM NaCl; 0.5% (v/v) Tritón X-100, 1 mM PMSF, 20 mM glutatión reducido (Sigma) . Para su correcto almacenamiento (a -80°C) y posterior uso, las proteínas recombinantes obtenidas son dializadas en 2 cambios consecutivos, durante al menos 16 horas, frente a un volumen de PBS-1X 1000 veces mayor al de la muestra a 4°C.

### **Ensayos de *pull-down***

Para los ensayos de *pull-down*, 200 µl de extracto nuclear (NE) de células HeLa diluidos en 800 µl de buffer de unión GST (20 mM Tris HCl, pH 7.9, 180 mM KCl, 0.2 mM EDTA, 0.05% Nonidet P-40, 0.5 mM PMSF, 1 mM DTT) son pre-aclarados en rotación en 40 µl de resina de agarosa acoplada a glutatión (Glutathione-Agarose, Sigma). Paralelamente, en otros 40 µl de resina, 6 µg de la proteína de fusión correspondiente es inmovilizada junto con 800 µl de buffer de unión GST al que se le añade BSA a 1mg/ml. Tras las 2 horas de incubación, la proteína unida a la resina es incubada con los extractos nucleares pre-aclarados durante 16 horas y en rotación a 4°C. La resina es extensamente lavada (5 lavados de 5 minutos en rotación a 4°C) con buffer de lavado GST conteniendo 0.1% Nonidet P-40. Las proteínas y complejos unidos son eluidos mediante la adición de 30 µl de buffer de carga SDS-2X(0.2% de azul de bromofenol, 100 mM DTT, SDS 4%, 20% glicerol y Tris HCl 0.1 M, pH 6.8) durante 10 minutos a 99°C. Los eluidos recogidos de esta manera son almacenados a -80°C o directamente resueltos en geles SDS-poliacrilamida del 10% y posteriormente analizados mediante tinción de plata.

### **Tinción de plata**

Los geles a analizar mediante esta tinción se sumergen en agitación suave durante 30 minutos en una solución de fijado (40% metanol (v/v), 10% (v/v) ácido acético). A continuación el gel es lavado durante 5 minutos en Solución A (10% (v/v) etanol, 5% (v/v) ácido acético) y sensibilizado mediante una solución al 0.01% de tiosulfato sódico (Na<sub>2</sub>O<sub>3</sub>S<sub>2</sub>) durante 5 minutos. Tras varios pasos de lavado con Solución A (durante 2 minutos), etanol al 10% (durante 5 minutos) y agua destilada (durante 5 minutos) se procede a la tinción del gel. Para ello se incuba en una solución al 0.1% de nitrato de plata (AgNO<sub>3</sub>) en agitación y oscuridad durante 5 minutos. El revelado se realiza en dos

pasos. En primer lugar, durante 5 minutos, el gel se incuba en una solución carbonato sódico ( $\text{Na}_2\text{CO}_3$ ) al 10%. Por último, el gel se sumerge en una solución de carbonato sódico al 2% y un 0.01% de formaldehído durante 20 segundos. Transcurrido este tiempo se realiza un breve lavado con agua destilada y se vuelve a añadir la misma solución hasta la aparición de bandas. La detención de la tinción se realiza lavando con agua destilada. El gel puede ser almacenado en una solución con un 40% de metanol y un 10% de ácido acético hasta su uso.

### **Tinción de coomasie**

Los geles a analizar mediante este tipo de tinción son sumergidos en agitación suave durante un mínimo de 20 minutos a temperatura ambiente en una solución de fijación/tinción que contiene 0.0025% (p/v) de Coomasie Blue R250, 45% (v/v) de metanol y 10% (v/v) de ácido acético. Transcurrido este tiempo se lavan abundantemente con una solución que contiene 25% (v/v) de metanol y un 7% (v/v) de ácido acético. El secado de los geles para su análisis y posterior almacenamiento se realiza mediante secado progresivo en papel celofán (BioRad) o secado convencional en aparato de vacío/calor.

## **6. OLIGONUCLEÓTIDOS**

A continuación se detallan los oligonucleótidos usados en este trabajo. Para los oligonucleótidos usados en procesos de clonaje se indica entre paréntesis el primer aminoácido codificante en el caso de los oligonucleótidos *forward* (*fwd*) y el último aminoácido codificante en el caso de los oligonucleótidos *reverse* (*rev*). En los oligonucleótidos utilizados en los procesos de mutagénesis dirigida se señalan los nucleótidos que fueron sustituidos en la secuencia original. Todos estos oligonucleótidos fueron sintetizados en el Servicio de “Síntesis de Oligos” del Instituto “López Neyra”, (CSIC). Se indican también los oligonucleótidos empleados en los experimentos de RT-PCR así como la secuencia de la sonda MS2b-Cy3 (descrita en Boireau S. et al 2007) usada en los experimentos de FISH.

### **Mutantes pEFBOS/ECFP/T7-TCERG1 [1-662]- FF**

*FF4 fwdBstBI* 5'-GGG TTC GAA AGG CCC GTT CAG AAC A-3' (Lys 878)  
*FF4 revBstBI* 5'-GGG TTC GAA TGT GTT CAT TAA AAA G-3' (His 949)  
*FF5 fwdBstBI* 5'-GGG TTC GAA ATG AAC ACA TTG AAG-3' (Asn 947)  
*FF5 revBstBI* 5'-GGG TTC GAA GCG TCC TGA AGT C-3' (Thr 1021)  
*FF6 rev BstBI* 5'-GGG TTC GAA TGG ACC CCG GCG ATC-3' (Pro 1083)

### **Mutantes pEGFP-FF5 y pEGFP-FF45**

*FF4 fwdXhoI* 5'-GGG CTC GAG GAA AGG CCC GTT CAG AAC A-3' (Lys 878)  
*FF5 fwd XhoI* 5'-GGG CTC GAG GAA ATG AAC ACA TTG AAG-3' (Asn 947)  
*FF5 revXhoI* 5'-GGG CTC GAG GCG TCC TGA AGT C-3' (Thr 1021)  
*FF5 revBglII* 5'-GGG AGA TCT GCG TCC TGA AGT C-3' (Thr 1021)

### **Mutante pEFBOS/ECFP/T7-TCERG1 ΔFF4/FF5**

*MS5 (multicloning site)* 5'-CGA AAC GCG TAC TAG TTC TAG ATT-3'  
*MS3 (multicloning site)* 5'-TTT GCG CAT GAT CAA GAT CTA AGC-3'  
*Amino- FF3 fwdAvrII* 5'-GGG CCT AGG ATG GCC CAA CAG CAG GCC TTG-3' (Met 21)  
*Amino-FF3 rev XbaI/MluI* 5'-GGG TCT AGA CTA ACG CGT GGC CTT TTG AAC CTC CCT TTC TCG-3' (Ala 879)  
*FF6 fwdMluI* 5'-GGG ACG CGT AAA TAT ATC ACA GCC AAA GCT-3' (Lys 1011)  
*FF6 rev XbaI* 5'-GGG TCT AGA CTA GTC ATC AAC ATA TGC CAC-3' (Asp 1077)

### **Mutantes pCGT7-RRM1- FF y pCGT7-RRM2-FF**

#### **RRM1 nonstop signal**

*RRM1 fwdXbaI* 5'-GGG TCT AGA TCG GGA GGT GGT GT-3'  
*RRM1 rev BamHI/XhoI* 5'-GGG GGA TCC CTC GAG CTC TCG GCC TGT TCC ACG GCC-3'

#### **RRM2 nonstop signal**

*RRM2 fwdXbaI* 5'-GGG TCT AGA GCT CCC CGA GGT CGG TAT- 3'  
*RRM2 rev BamHI/XhoI* 5'-GGG GGA TCC CTC GAG CTC TCC ATA ACT TGG ACT TCT-3'  
*FF4 fwd BglII* 5'-GGG AGA TCT AAG GCC CGT TCA GAA CA-3' (Lys 878)  
*FF5 fwd BglII* 5'-GGG AGA TCT AAT GAA CAC ATT GAA G-3' (Asn 947)  
*FF5 rev BglII* 5'-GGG AGA TCT GCG TCC TGA AGT C-3' (Thr 1021)

**pGEX2TK-FF4/FF5 y pGEX2TK-FF4/FF5-F903A/ F961A**

*FF4 fwd BglII* 5'-GGG AGA TCT AAG GCC CGT TCA GAA CA-3' (Lys 878)  
*FF5 rev EcoRI* 5'-GGG GAA TTC CGT CCT GAA GTC AGC TTT GGC TG-3' (Thr 1021)

**pEFBOS/ECFP/T7-TCERG1 [1-662]- FF4/FF5 y pEFBOS/ECFP/T7-TCERG1 [1-1098]  
con la mutación K985R**

*K985R fwd* 5'-GTA AAA AAA ATC ATT **AGG** GAA GAT CCT CGA TG-3'  
*K985R rev* 5'-CAT CGA GGA TCT TCC CTA ATG ATT TTT TTT AC-3'

**pEFBOS/ECFP/T7-TCERG1 [1-662]- FF4/FF5 y pEFBOS/ECFP/T7-TCERG1 [1-1098]  
con las mutaciones F903A, F961A, F903A/F961A, F946A y F903A/F946A**

*F903A fwd* 5'-GAA GCT ATC CAG AAT **GCC** AAA GCT CTT CTG TCT GAC-3'  
*F903A rev* 5'-GTC AGA CAG AAG AGC TTT GGC ATT CTG GAT AGC TTC-3'  
*F946A fwd* 5'-GAG GAG AAA GAG AAG CTT **GCT** AAT GAA CAC ATT GAA GCA C-3'  
*F946A rev* 5'-GTG CTT CAA TGT GTT CAT TAG CAA GCT TCT CTT TCT CCT C-3'  
*F961A fwd* 5'-CCA AAA AAA AGA GAG AGC **ACGCTA** GGC AAC TTC TGG ATG-3'  
*F961A rev* 5'-CAT CCA GAA GTT GCC TAG CGT GCT CTC TCT TTT TTT TGG-3'

**Retrotranscripción de ARN para el minigen HIV2-X2**

*RT-Sveda* 5'-GGGAAGCTAGAGTAAGTAG-3'

**PCR convencional para el minigen HIV-X2**

*X34 (fwd)* 5'-AGGGAGGCAGGCGACGGCGACGAGTTT-3'  
*XAgeIR (rev)* 5'-GTGGATCCCCCGGGCTGCAGGAATTCGAT-3'

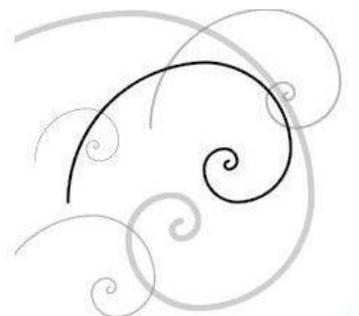
**Sonda MS2b-Cy3**

5'-AT\* GTCGACCTGCAGACAT\*GGGTGATCCTCAT\*GTTTTCTAGGCAATT\*A-3'

(con \* se indican las posiciones marcadas con Cy3).



## *Resultados*



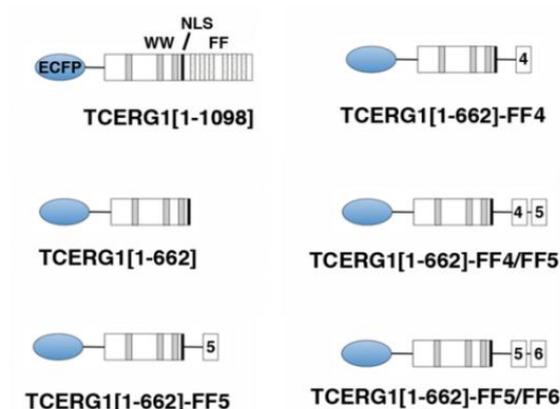


# 1. LOS DOMINIOS FF4 Y FF5 DE TCERG1 DIRIGEN PROTEÍNAS HACIA LA PERIFERIA DE LOS SPECKLES NUCLEARES

## 1.1. Identificación de la secuencia necesaria para la localización de TCERG1 en los speckles nucleares

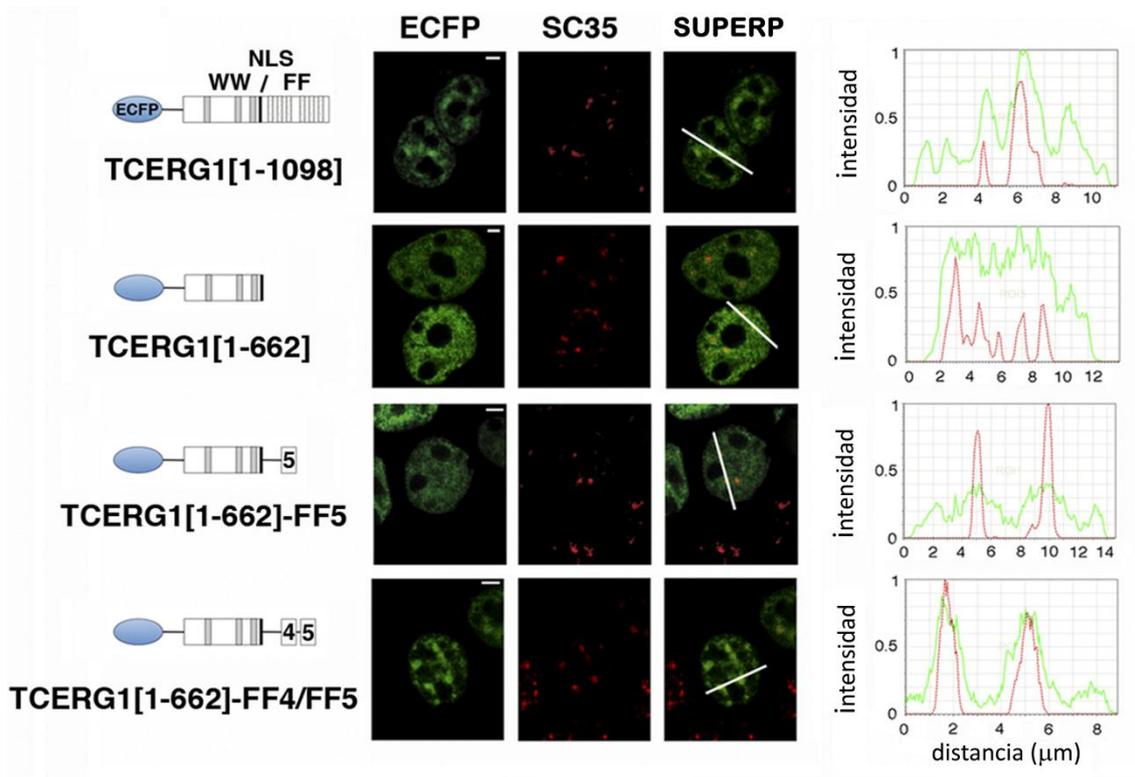
TCERG1 ha sido localizada en la periferia de los *speckles* nucleares, pero la secuencia específica responsable de su correcta localización no ha sido establecida. Estudios previos de nuestro grupo revelaron que un mutante de TCERG1 carente de los dominios FF5 y FF6 no es capaz de situarse en los *speckles* nucleares (Sanchez-Alvarez et al., 2006). De este resultado podemos inferir tres posibles interpretaciones; i) el dominio FF5 de TCERG1 es esencial para su correcta localización, ii) los dominios FF5 Y FF6 de TCERG1 son importantes, pero ninguno de ellos es esencial y iii) un número mínimo de dominios FF son necesarios para que TCERG1 se sitúe en los *speckles* nucleares.

Con el objetivo de identificar en detalle los elementos necesarios dentro de la porción carboxiterminal de TCERG1 para su correcta localización en los *speckles* nucleares, generamos una batería de construcciones que contenían la región aminoterminal de TCERG1 fusionada a diferentes dominios FF (Figura R-1). Todas ellas expresan la proteína ECFP (*enhanced cyan fluorescent protein*) así como la secuencia aminoacídica correspondiente a la proteína de la cápside del fago T7 en su parte aminoterminal. Las construcciones generadas de este modo conservan la señal de localización nuclear, que ha sido previamente descrita en la región central de la proteína (Sune et al., 1997).



**Figura R-1 (página anterior). Representación de los mutantes de TCERG1 usados para el estudio de los requerimientos de dominios FF en la correcta localización de la proteína.** En la porción amino terminal de cada uno de los mutante se ha fusionado la proteína ECFP (*enhanced cyan fluorescent protein*), representada con un óvalo de color azul. Se dibujan además los dominios más característicos de TCERG1: tres dominios WW en la región amino-terminal, seis dominios FF en la región carboxiterminal y la señal de localización nuclear (NLS) que se sitúa en la porción central de la proteína. Todos los mutantes usados en este estudio contienen la región aminoterminal de TCERG1 (aminoácidos 1-662).

Los estudios de localización subnuclear de estas construcciones se llevaron a cabo mediante transfecciones transitorias en células HEK293T. La localización de la versión completa de TCERG1 (aminoácidos 1 a 1098) corroboró la distribución ya observada en anteriores experimentos siendo idéntica a la que muestra la proteína endógena (Sánchez-Álvarez et al 2006), con una elevada acumulación en los *speckles* nucleares (Figura R-2, ECFP, TCERG1[1-1098]). La construcción que solo contiene la región aminoterminal de la proteína mostró una distribución totalmente difusa por el nucleoplasma, sin ninguna acumulación evidente en los *speckles* nucleares (Figura R-2, ECFP, TCERG1[1-662]), apoyando de este modo los datos previos del laboratorio y otros estudios anteriores (Smith et al., 2004; Sanchez-Alvarez et al., 2006). La sobre-expresión de la construcción que contiene la región aminoterminal fusionada únicamente al dominio FF5 de TCERG1 presentó un ligero enriquecimiento de la proteína en los *speckles* nucleares. Este resultado indica que este dominio es necesario para la localización de la proteína pero que, sin embargo, no es suficiente para una eficiente y correcta acumulación de TCERG1 en estos compartimentos nucleares (Figura R-2, ECFP, comparar la señal obtenida para la proteína TCERG1[1-1098] con la obtenida con la proteína TCERG1[1-662]-FF5). Para profundizar en los requerimientos proteicos necesarios para la correcta localización de la proteína, construimos una nueva proteína quimera que contuviese los dominios FF4 y FF5. Esta construcción mostró una fuerte acumulación en los *speckles* nucleares, idéntica a la obtenida con la proteína *wild-type*. (Figura R-2, ECFP, comparar la señal obtenida para la proteína TCERG1[1-1098] con la obtenida con la proteína TCERG1[1-662]-FF4/FF5)



**Figura R-2. Los dominios FF4/FF5 de TCERG1 son necesarios para la correcta localización de la proteína en los *speckles* nucleares.** Análisis mediante microscopía confocal de la distribución subnuclear de los mutantes ECFP-TCERG1. En el panel se muestra el doble marcaje realizado en células HEK293T transfectadas transitoriamente con las diferentes proteínas de fusión ECFP (en verde, columna izquierda: ECFP) y con el anticuerpo frente a SC35 (en rojo, columna central del panel: SC35). Se muestra además las imágenes resultantes de la superposición de ambas señales (panel derecho: SUPERP). El análisis cuantitativo de la intensidad de las señales de CFP y SC35 se muestra a la derecha de la imagen. Las líneas dibujadas en las imágenes superpuestas indican la zona de estudio de estos análisis. Barra=3 $\mu$ m.

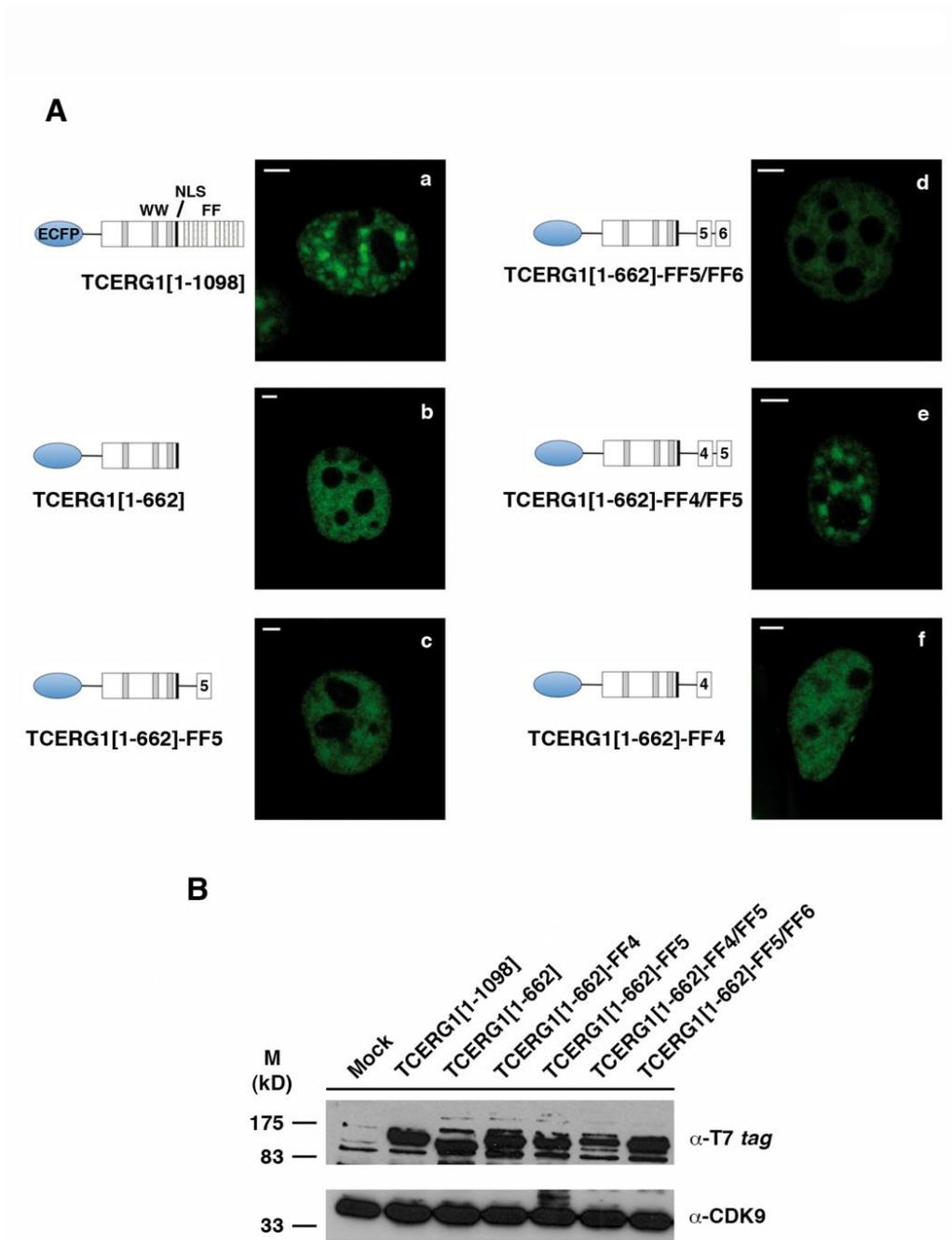
Para confirmar que la localización obtenida con cada una de las construcciones sobreexpresadas transitoriamente coincidía realmente con la de los *speckles* nucleares, los estudios de inmunofluorescencia se llevaron a cabo usando un marcaje específico frente al factor esencial de *splicing* SC35 que, de forma habitual, se usa para definir a estas estructuras. La superposición de las señales obtenidas con la construcción que contiene los dominios FF4 y FF5 y con el factor SC35 reflejó una perfecta colocalización de ambas proteínas en los *speckles* nucleares, de igual modo que ocurría con la proteína entera. Sin embargo, la construcción que solo contiene el dominio FF5 muestra una localización parcial en estos dominios, estando su señal más ampliamente distribuida

por el nucleoplasma (Figura R-2). La relación espacial entre las variantes de TCERG1 y SC35 fue estudiada mediante análisis cuantitativo en diferentes *speckles* nucleares (Figura R-2, derecha). El análisis reveló que los picos de intensidad de señal de la proteína *wild type* y la construcción TCERG1[1-662]-FF4/FF5 son coincidentes con la señal obtenida para SC35, mientras que para la construcción TCERG1[1-662]-FF5 el grado de coincidencia es mucho menor.

Para descartar que la localización observada con la construcción TCERG1[1-662]-FF4/FF5 no era debida a la unión de dos dominios FF cualesquiera, generamos una construcción adicional que contuviese los dominios FF5 y FF6 (TCERG1[1-662]-FF5/FF6). La expresión transitoria de esta proteína produjo un patrón de tinción similar al obtenido con la expresión de TCERG1[1-662]-FF5, sin una acumulación significativa de la proteína expresada en los *speckles* nucleares (Figura R-3, A, d). Por último, la construcción que contiene únicamente el dominio FF4 (TCERG1[1-662]-FF4) produjo un patrón de distribución homogéneo por todo el nucleoplasma (Figura R-3, A, f). El estudio por western blot de las diferentes construcciones sobre-expresadas demostró que la localización de cada uno de ellos no es debida a diferencias en los niveles de expresión (Figura R-3, B).

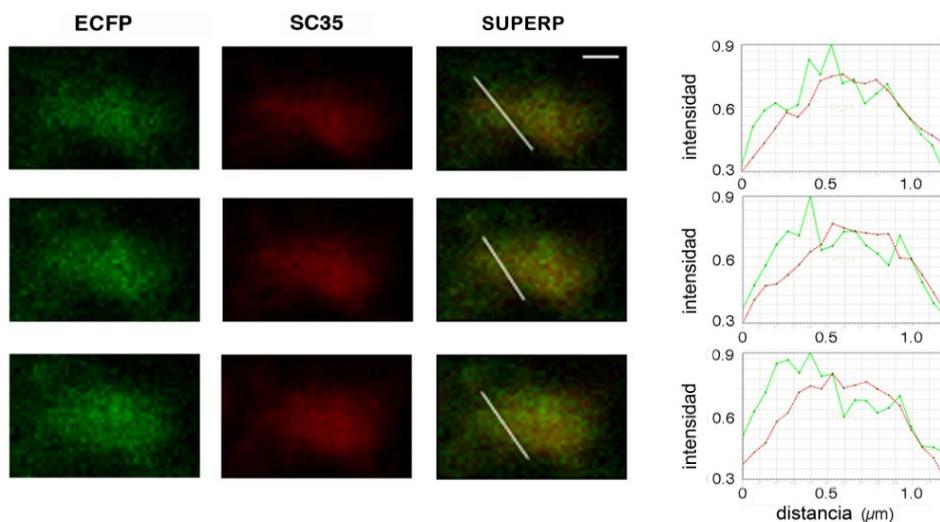
Por último, se generaron diversas construcciones de fusión con la ECFP conteniendo los dominios FF1, FF2 y FF3. Sin embargo, las proteínas expresadas formaron agregados nucleares que imposibilitaron el estudio de su localización espacial en la célula (datos no mostrados).

En conjunto, todos estos resultados sugieren que la secuencia correspondiente a los dominios FF4 y FF5 es la secuencia necesaria y suficiente para la correcta localización de TCERG1 en los *speckles* nucleares.



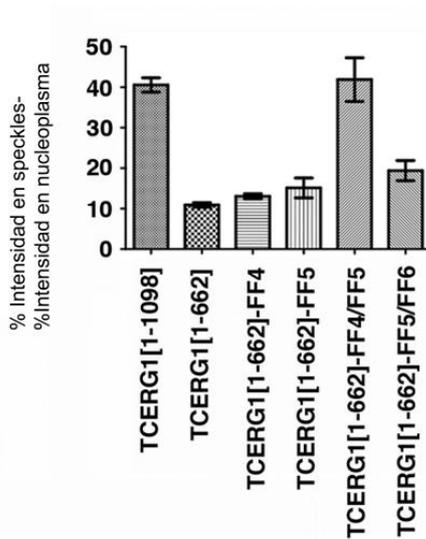
**Figura R-3. Análisis de la distribución espacial y la expresión de proteínas de fusión ECFP-TCERG1.** (A) Las diferentes proteínas de fusión ECFP-TCERG1, indicadas a la izquierda de cada imagen, fueron transfectadas transitoriamente en células HEK293T. Su patrón de distribución nuclear fue analizado mediante microscopía confocal. Barra=3µm. (B) Una muestra de los lisados celulares de células HEK293T transfectadas transitoriamente con las construcciones ECFP-TCERG1 fue analizada mediante western blot. Las diferentes construcciones fueron detectadas usando anticuerpos frente el epítipo T7. Como control de carga proteica se analizó la expresión de la quinasa Cdk9 (*Cyclin-dependent kinase 9*). El peso molecular de los marcadores (M, en kilodalton (kD)) queda indicado a la izquierda del panel.

Para analizar con más detalle el patrón de localización mediado por los dominios FF4/FF5, analizamos la distribución espacial tanto de la proteína *wild type* como la de la construcción TCERG1[1-662]-FF4/FF5 en secciones ópticas de 0.3  $\mu\text{m}$  a lo largo del eje Z de *speckles* individuales en relación a la distribución del factor SC35 (Figura R-4). Este exhaustivo análisis demostró que la versión *wild type* de la proteína y la construcción TCERG1 [1-662]-FF4/FF5 se distribuyen tanto de forma periférica como interna en los *speckles* nucleares, pero que quedan excluidas de su centro o núcleo. Por ello, el análisis cuantitativo de estas imágenes mostró que los picos de distribución de TCERG1 solapan parcialmente con los de SC35 pero no coinciden de manera exacta con éstos. Estos datos apoyan los resultados previos de los que disponía nuestro laboratorio que afirmaban que TCERG1 se encuentra enriquecido en la periferia de los *speckles* nucleares (Sanchez-Alvarez et al., 2006). Además demuestran que la localización mediada por los dominios FF4/FF5 también provoca un enriquecimiento en estas zonas concretas de los *speckles* nucleares.



**Figura R-4. TCERG1[1-662]-FF4/FF5 se localiza principalmente en la periferia y regiones externas del core de los *speckles* nucleares.** El estudio de secciones ópticas seriadas a lo largo del eje Z de *speckles* individuales (grosor de cada uno de los planos  $z=0,3 \mu\text{m}$ ) fue realizado a baja intensidad de ganancia de señal. La imagen muestra las señales por separado para ECFP (verde, columna izquierda: ECFP) y SC35 (rojo, columna central: SC35) así como la superposición de ambas (columna derecha: SUPERP). A la derecha de la imagen se muestran los análisis lineales de intensidad para ambos canales, evidenciando un solapamiento parcial pero no una coincidencia completa y una extensión del marcaje de TCERG1 hacia áreas periféricas del compartimento nuclear. Barra=  $3\mu\text{m}$ .

Los datos obtenidos fueron corroborados con el análisis cuantitativo y estadístico de la diferencia entre las intensidades de la señal obtenida mediante ECFP entre los *speckles* y el nucleoplasma (Figura R-5, ver “Materiales y métodos”, apartado de “Inmunofluorescencia, FISH, FRAP y procedimientos de microscopía” para una mayor explicación de la metodología usada en esta cuantificación) y confirmó el requerimiento de los dominios FF4 y FF5 para la correcta localización de TCERG1 en los *speckles* nucleares.



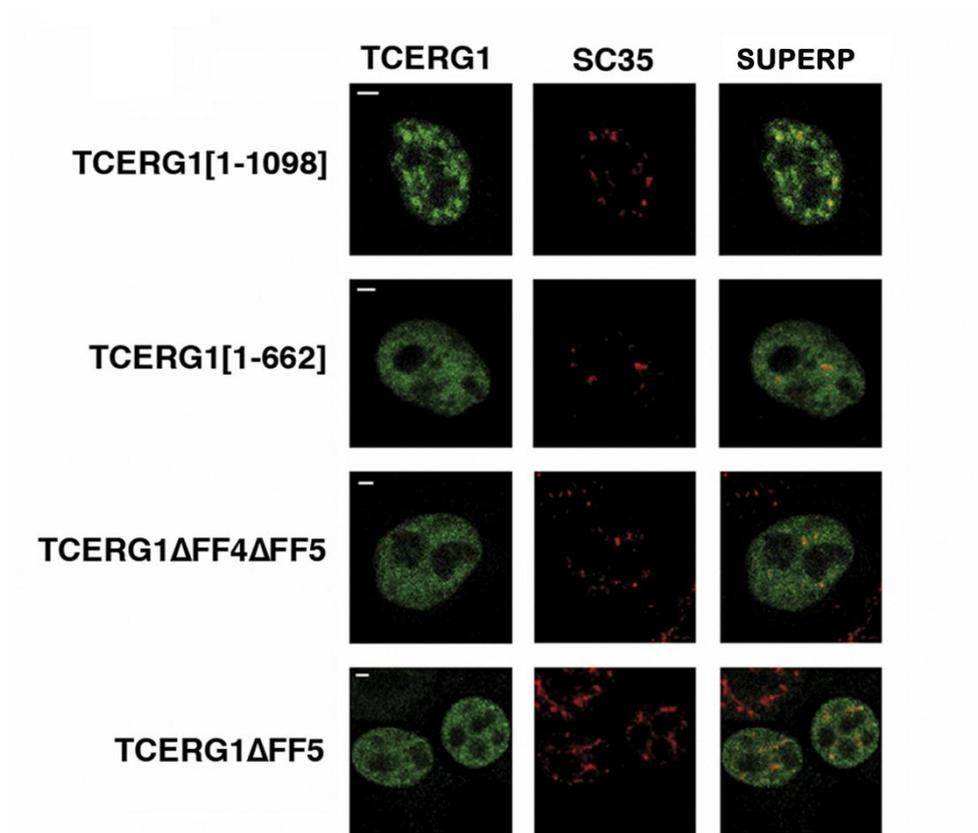
**Figura R-5. Medida de la fracción de TCERG1 presente en los *speckles* nucleares.** La cuantificación de las señales relativas de cada uno de los mutantes ECFP-TCERG1 se llevó a cabo en tres experimentos independientes tal y como se describe en el apartado “Materiales y métodos”. Los datos fueron analizados estadísticamente mediante el test *T-Student* que ofreció diferencias altamente significativas en todos los casos.

$p(\text{TCERG1}[1-1098] \text{ versus TCERG1}[1-662]) < 0.0001$ ;  $p(\text{TCERG1}[1-1098] \text{ versus TCERG1}[1-662]-\text{FF4}) = 0.0001$ ;  $p(\text{TCERG1}[1-1098] \text{ versus TCERG1}[1-662]-\text{FF5}) = 0.0011$ ;  $p(\text{TCERG1}[1-1098] \text{ versus TCERG1}[1-662]-\text{FF4}/\text{FF5}) = 0.8261$ ;  $p(\text{TCERG1}[1-662]-\text{FF4}/\text{FF5} \text{ versus TCERG1}[1-662]-\text{FF5}/\text{FF6}) = 0.0023$ ;  $p(\text{TCERG1}[1-662]-\text{FF4}/\text{FF5} \text{ versus TCERG1}[1-662]-\text{FF5}) = 0.0106$ ; and  $p(\text{TCERG1}[1-662]-\text{FF4}/\text{FF5} \text{ versus TCERG1}[1-662]-\text{FF5}/\text{FF6}) = 0.0192$ . Se prepresenta la media de los tres experimentos (media  $\pm$  desviación estándar (DS)).

El estudio de señalización de TCERG1 hacia los *speckles* nucleares lo basamos en el análisis de la localización espacial de construcciones que contenían diferentes combinaciones de la región aminoterminal de la proteína junto con uno o dos dominios FF. Sin embargo, era importante conocer el efecto de la delección de los dominios FF4 y FF5 en el contexto de la proteína *wild type*. Para ello expresamos TCERG1 *wild type* y el mutante TCERG1 $\Delta$ FF4 $\Delta$ FF5 (TCERG1 con delección de los dominios FF4 y FF5) de forma transitoria en células HEK293T. El estudio de localización se realizó de nuevo mediante microscopía confocal gracias a la presencia de la proteína ECFP en la zona aminoterminal de ambas construcciones. Como se muestra en la figura R-6, la delección de los dominios FF4 y FF5 de TCERG1 (TCERG1 $\Delta$ FF4 $\Delta$ FF5) hizo que la proteína mostrase una distribución homogénea por todo el nucleoplasma sin acumulación evidente en los *speckles*, que de nuevo fueron puestos de manifiesto mediante tinción

específica frente al factor esencial de *splicing* SC35. De esta manera se refuerza la conclusión obtenida acerca del papel de los dominios FF4/FF5 en la correcta localización de TCERG1 en la región de los *speckles* nucleares.

Estudios previos de nuestro laboratorio, ya citados anteriormente, habían demostrado que la única delección del dominio FF6 de TCERG1 no afecta a su distribución, lo cual concordaba con los datos de los que disponíamos hasta el momento. Además, sabíamos que una delección más amplia que elimina también el dominio FF5, sí interfiere con la distribución nuclear de la proteína (Sánchez-Álvarez M et al. 2006), impidiendo su localización en los *speckles* nucleares. Con el fin de comprobar si el dominio FF5 es esencial para la correcta localización de la proteína, transfectamos transitoriamente una construcción en la que éste dominio se encuentra delecionado (TCERG1 $\Delta$ FF5) y comprobamos su distribución espacial en el núcleo mediante inmunofluorescencia indirecta frente al epítipo T7 contenido en esta construcción (Figura R-6). La delección de este dominio resultó en una distribución difusa por todo el nucleoplasma, indicando que el dominio FF5 de TCERG1 es esencial para la asociación de la proteína con los *speckles* nucleares.



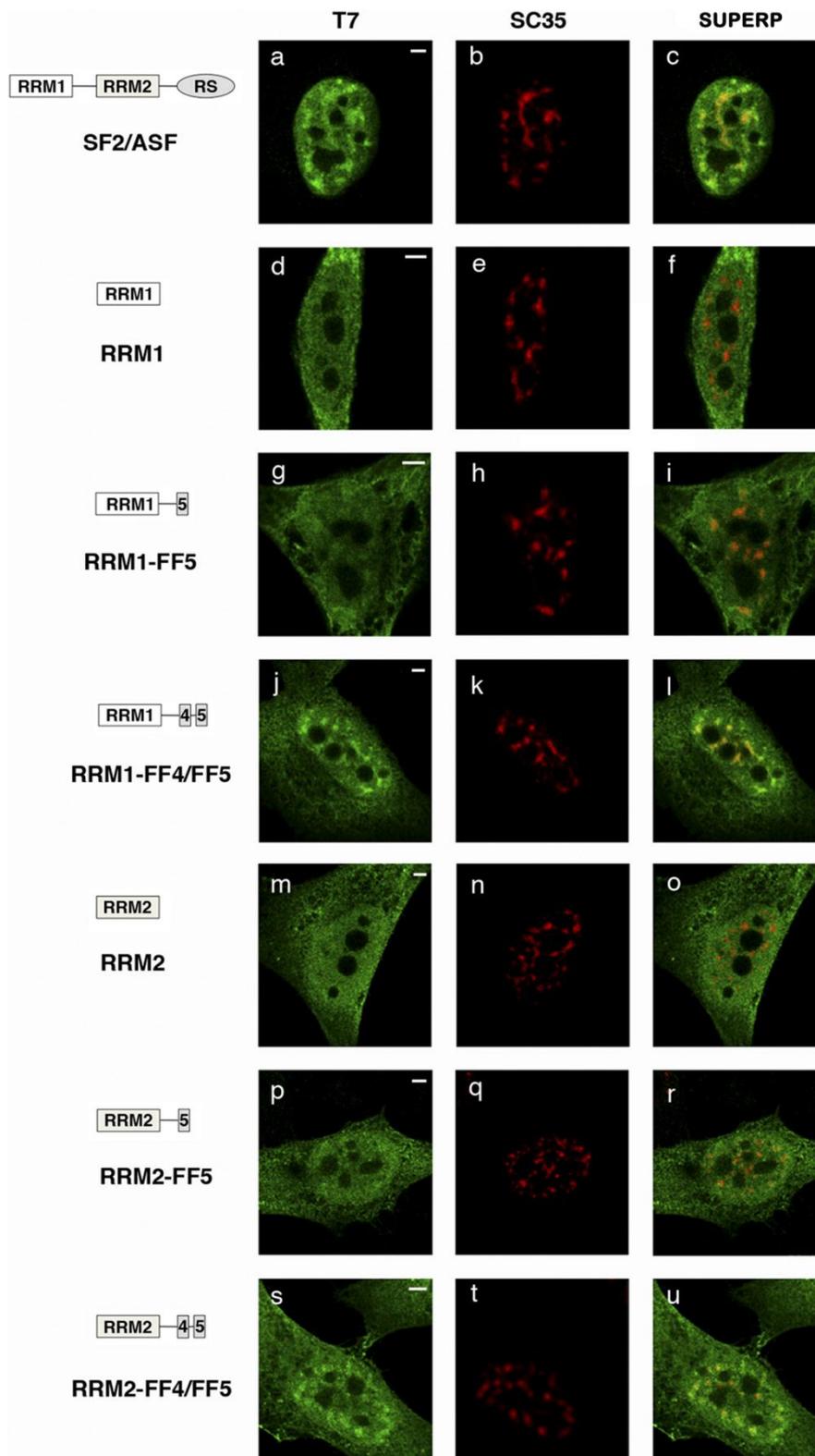
**Figura R-6. (página anterior) El dominio FF5 de TCERG1 es esencial para su correcta localización en los *speckles* nucleares.** Células HEK293T fueron transitoriamente transfectadas con cada uno de los plásmidos indicados y su distribución subnuclear fue analizada mediante microscopía confocal. La construcción TCERG1 [1-1098] mostró la localización característica de la proteína *wild type*, mientras que las construcciones TCERG1 $\Delta$ FF4 $\Delta$ FF5 y TCERG1 $\Delta$ FF5 mostraron una distribución difusa por el nucleoplasma, sin acumulación evidente en los *speckles* nucleares. Se muestran imágenes individuales de cada uno de los mutantes de TCERG1 (verde, columna izquierda), para SC35 (rojo, columna central: SC35), así como la superposición de ambas imágenes (columna derecha: SUPERP). Barra= 3 $\mu$ m.

### ***1.2. Los dominios FF4 y FF5 de TCERG1 representan una nueva señal de localización hacia la periferia de los speckles nucleares***

Una vez definida la señal responsable de la localización de TCERG1 en los *speckles* nucleares nos preguntamos si podría representar una nueva señal de localización general hacia estos dominios, es decir, si sería capaz de dirigir a otras proteínas hasta estos compartimentos nucleares. Para contestar a esa pregunta, decidimos usar el factor SRFS1 (conocido anteriormente como SF2/ASF), factor esencial de *splicing* que regula el procesamiento constitutivo y alternativo de diversos pre-ARNm. SRFS1 pertenece a la familia de las proteínas SR, todas ellas caracterizadas por la presencia de una o dos copias del motivo de unión a ARNRRM y un dominio carboxiterminal RS, rico en residuos de arginina y serina (Long, J.C. 2009). En concreto, SRSF1 presenta dos dominios RRM y un dominio RS, habiéndose descrito con exactitud el papel que cada uno de los motivos tiene en la localización celular, nuclear y en la regulación del *splicing* alternativo (Caceres et al., 1997). Para su correcta localización en los *speckles* nucleares, son necesarios al menos dos de los tres dominios que conforman SRSF1, lo que indica que en cada uno de los dominios individuales existen señales de localización débiles que funcionan de manera cooperativa. El trabajo de Cáceres y colaboradores demuestra como la expresión de proteínas SRSF1 que contienen cualquier combinación de dos de los tres dominios muestran una acumulación evidente en los *speckles* nucleares, mientras que mutantes de SRSF1 con un solo dominio RRM quedan dispersos por el nucleoplasma. Haciendo uso de estas construcciones delecionadas de SRSF1 (Cáceres J 1993) quisimos saber si los dominios FF4/FF5 de TCERG1 eran capaces de promover la localización en los *speckles* nucleares de mutantes de SRSF1 que han perdido su localización a los mismos. Para ello, generamos una nueva colección

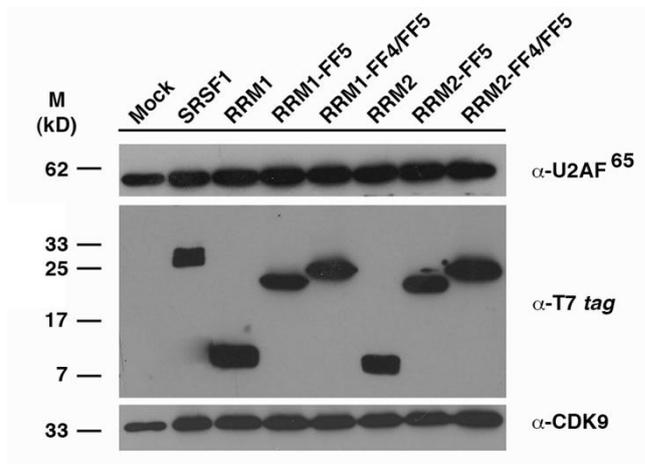
de construcciones en las que fusionamos los dominios FF5 y FF4/FF5 a la región carboxiterminal de los motivos RRM1 y RRM2 de SRSF1. Además, todas las proteínas de fusión contenían en su extremo aminoterminal la secuencia correspondiente a la proteína de la cápside del fago T7, que nos permitió la detección, tanto por western blot como por inmunofluorescencia, de la proteína exógena transfectada transitoriamente en células HeLa.

Los experimentos de inmunofluorescencia indirecta con anticuerpos frente a epítipo T7 y el estudio de la distribución mediante microscopía confocal mostraron que SRSF1 *wild type* presenta una localización exclusiva dentro del núcleo, con el patrón típico de localización en los *speckles* nucleares (Figura R-8, (a)) puesto de manifiesto mediante el marcaje frente a la proteína endógena SC35 (Figura R-9, (b y c)). Sin embargo, las construcciones que presentaban únicamente los dominios RRM1 y RRM2 exhibieron una localización tanto citoplasmática como nuclear, sin ninguna acumulación en los *speckles* nucleares y una distribución homogénea por todo el nucleoplasma (Figura R-9, (d-f y m-o), de acuerdo con datos ya publicados (Caceres et al., 1997). Las construcciones con los dominios RRM1 y RRM2 fusionados al dominio FF5 de TCERG1 provocaron pequeños cambios en el patrón obtenido para los dominios RRM, empezando a ser evidente cierta acumulación de la señal en los *speckles* nucleares sin llegar en ningún caso a alcanzar la distribución de la proteína *wild type* (Figura R-9, (g-i) y (p-r)). Sin embargo, la fusión de los dominios FF4/FF5 a RRM1 y RRM2 sí produjo una clara acumulación de las proteínas de fusión en los *speckles* nucleares (Figura R-9, (j-l y s-u). El estudio de western blot de los diferentes mutantes demostró una expresión equivalente para todos ellos (Figura R-9).



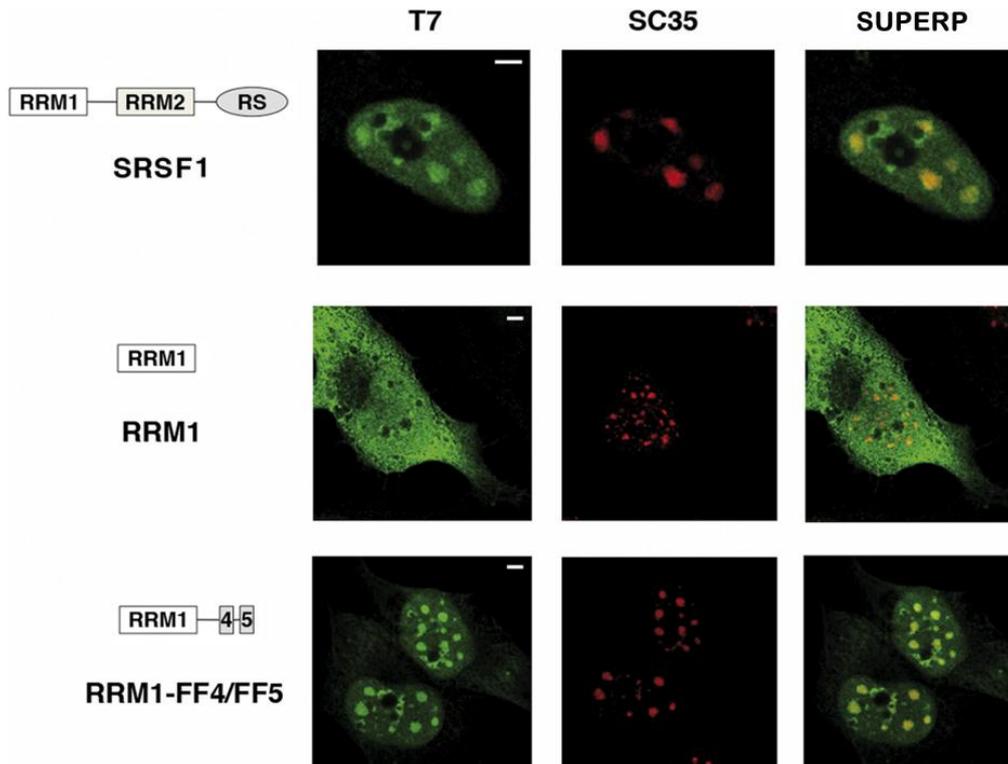
**Figura R-8.** Los dominios FF4/FF5 de TCERG1 localizan a mutantes delecionados de SRSF1 hacia los *speckles* nucleares. Células HeLa fueron transfectadas transitoriamente con las construcciones indicadas esquemáticamente en la parte izquierda de la imagen. El estudio de su patrón de distribución se realizó a través de experimentos de doble inmunofluorescencia indirecta frente al epítipo T7 (verde, columna izquierda: T7) y frente al factor SC35 (rojo, columna central: SC35).

(continuación pie Figura R-8) La superposición de ambas imágenes se muestra en la columna derecha de la imagen (SUPERP). La visualización de todas las construcciones se realizó mediante microscopía confocal. Barra= 3µm.



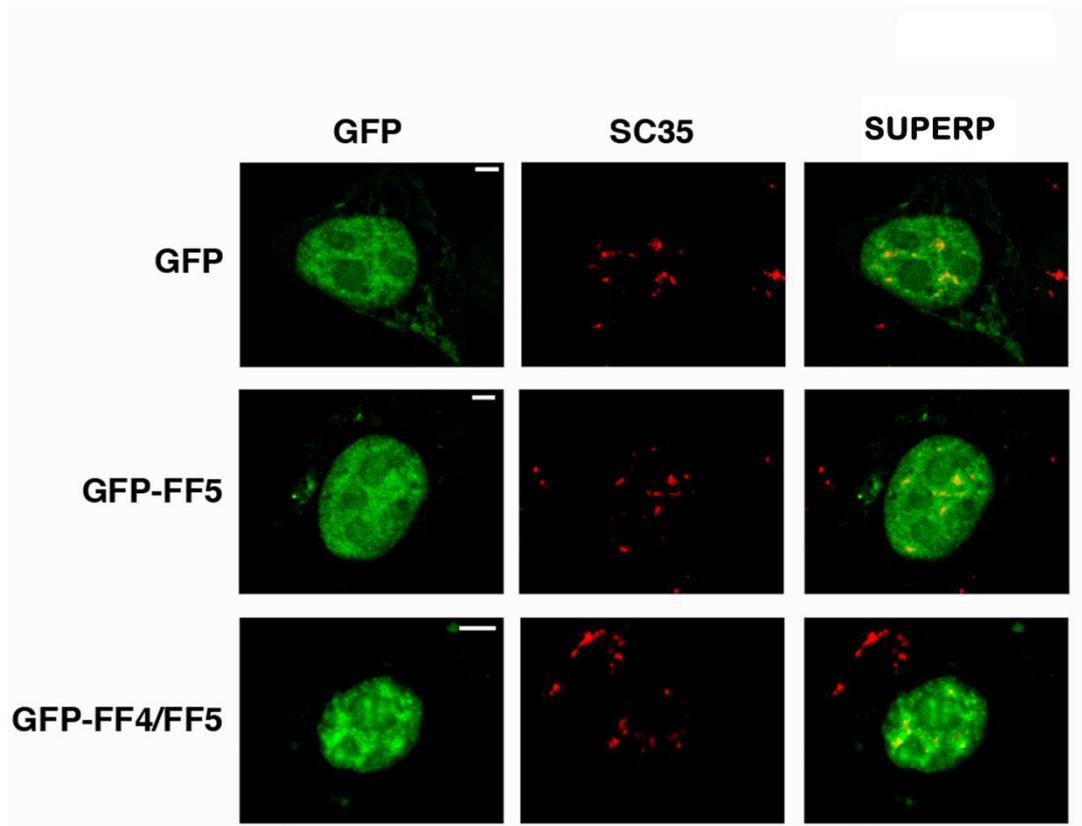
**Figura R-9. Análisis de la expresión de las proteínas de fusión SRSF1/FF (TCERG1).** Una muestra de los lisados celulares de células HeLa transfectadas transitoriamente con las construcciones indicadas fue analizada mediante western blot usando anticuerpos específicos frente al epítipo T7. Los controles de carga proteicos se llevaron a cabo estudiando la expresión de U2AF<sup>65</sup> y Cdk9. El peso molecular de los marcadores (M, en *kD*) queda indicado a la izquierda del panel.

Con el fin de conocer si el comportamiento de las proteínas de fusión generadas entre los dominios RRM de SRSF1 y los dominios FF de TCERG1 no había alterado el comportamiento de la proteína SRSF1 y éste era semejante al de otros factores de *splicing*, tratamos a las células HeLa transitoriamente transfectadas con el inhibidor de la ARN polIII  $\alpha$ -amanitina. Bajo estas condiciones de inhibición transcripcional, los factores de *splicing* implicados en el procesamiento de los transcritos recién generados, entre ellos TCERG1, son almacenados de manera más prominente en los *speckles* nucleares. Así, el número de estos orgánulos se reduce y además su morfología sufre un cambio muy característico, pasando a ser más redondeados y expandidos (O'Keefe, R.T. 1994; Sánchez-Álvarez, M. 2006). En estas condiciones de inhibición transcripcional, la distribución del mutante RRM1-FF4/FF5 mostró un cambio similar al obtenido para el patrón de SRSF1 *wild type* y SC35 (Figura R-10). De los datos obtenidos concluimos que la localización de las proteínas de fusión con la señal de localización hacia los *speckles* nucleares de TCERG1 no es dependiente de una transcripción activa en la célula y que el comportamiento de estas proteínas quimeras es el mismo que el de otros factores de *splicing* ante condiciones de inhibición transcripcional.



**Figura R-10.** La localización de la proteína de fusión RRM1-FF4/FF5 hacia los *speckles* nucleares no se ve alterada tras la inhibición de la transcripción. Células HeLa fueron transfectadas con cada uno de los plásmidos indicados esquemáticamente en la parte izquierda de la imagen y tratadas con 25  $\mu\text{g/ml}$  de  $\alpha$ -amanitina durante 6 horas a 37  $^{\circ}\text{C}$ . Posteriormente fueron procesadas para su estudio mediante inmunofluorescencia usando anticuerpos frente al epítipo T7 (verde, columna izquierda: T7) y frente al factor SC35 (rojo, columna central: SC35). Se muestran las imágenes individuales y la superposición de ambos canales (columna derecha: SUPERP). Barra= 3 $\mu\text{m}$ .

Finalmente quisimos comprobar el efecto de la inclusión de los dominios FF4/FF5 en otras proteínas heterólogas no relacionadas con el procesamiento del ARN. Para ello generamos proteínas de fusión que contenían la proteína GFP y los dominios FF5 o FF4/FF5. La construcción GFP-FF5 mostró un patrón de tinción similar al que exhibió la proteína GFP expresada de forma individual (Figura R-11, panel izquierdo), la cual se distribuye de forma difusa por el nucleoplasma aunque con pequeñas áreas de acumulación (ver por ejemplo Alvarez et al., 2003). La construcción GFP-FF4/FF5 se acumuló de forma clara en los *speckles* nucleares, de nuevo puestos de manifiesto mediante tinción frente al factor esencial de *splicing* SC35 (Figura R-11, panel central).



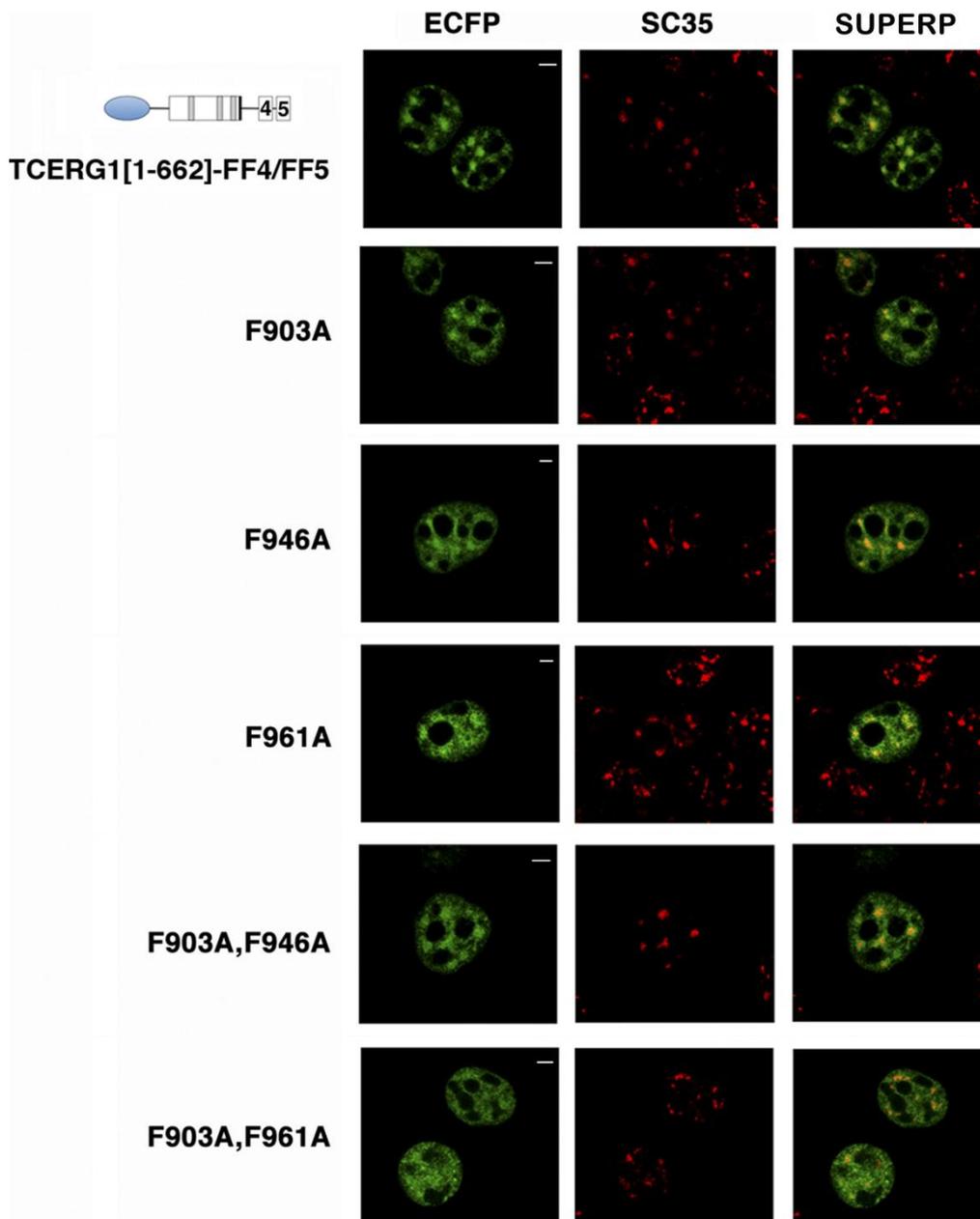
**Figura R-11. Los dominios FF4/FF5 de TCERG1 dirigen a la proteína GFP hacia los *speckles* nucleares.** Las construcciones GFP, GFP-FF5 y GFP-FF4/FF5 fueron transfectadas transitoriamente en células HeLa. Su distribución (verde, columna izquierda: GFP) se estudió mediante inmunofluorescencia en la que se empleó anticuerpos específicos para SC35 (rojo, columna central: SC35) para poner de manifiesto los *speckles* nucleares. En todos los casos la colocación de las proteínas GFP y SC35 se analizó mediante microscopía confocal. La figura muestra las imágenes individuales y la superposición de las mismas (columna derecha: SUPERP). Barra= 3 $\mu$ m.

En conjunto, el estudio de localización de los mutantes de SRSF1 y la proteína GFP demuestra que la región FF4/FF5 no solo es la región necesaria para la correcta localización de TCERG1, sino que también constituye una señal suficiente para dirigir a otras proteínas hacia los *speckles* nucleares. Por ello concluimos que los dominios FF4/FF5 cumplen los requisitos necesarios para ser una nueva señal general de localización a los *speckles* nucleares.

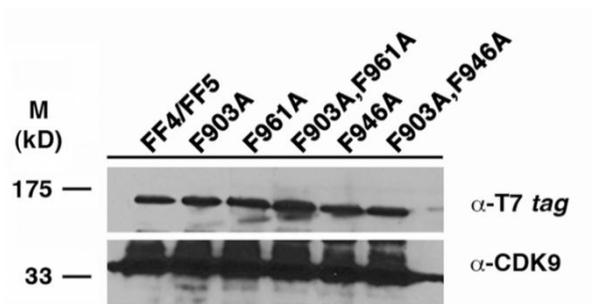
### ***1.3. Implicación de los residuos altamente conservados de fenilalanina de los dominios FF4/FF5 de TCERG1 en la localización a los speckles nucleares***

Debido al alto grado de conservación de los residuos de fenilalanina de los dominios FF, quisimos conocer la implicación de estos aminoácidos en la función de localización nuclear mediada por los dominios FF4/FF5. Para ello y partiendo de la construcción ECFP-TCERG1[1-662]-FF4/FF5, construimos una nueva colección de mutantes que contuviesen cambios de fenilalanina a alanina en las posiciones Phe-903, Phe-946 y Phe-961, así como los dobles mutantes F903A, F946A y F903A, F961A. Todos estos mutantes fueron transfectados transitoriamente en células HEK293T y su distribución espacial fue analizada como en veces anteriores mediante microscopía confocal. El cambio de la fenilalanina 1004 (F1004), situada en el dominio FF5 no pudo llevarse a cabo debido a la presencia de una secuencia cercana a ella rica en adeninas que imposibilitó la mutagénesis dirigida de este residuo en concreto.

La sobre-expresión transitoria de la construcción ECFP-TCERG1[1-662]-FF4/FF5 y las construcciones con mutaciones en una o en las dos fenilalaninas del dominio FF4 (F903A, F946A y el doble mutante F903A, F946A) mostraron un patrón de distribución similar, con una marcada acumulación en los *speckles* nucleares. La mutación que afectaba a la fenilalanina en posición 961 (F961A) dentro del dominio FF5, provocó una disminución de este enriquecimiento en los *speckles*. Por último, el doble mutante F903A, F961A mostró un patrón totalmente difuso por el nucleoplasma, sin ninguna acumulación evidente en los *speckles* nucleares (Figura R-12). Se obtuvieron los mismos resultados cuando las mutaciones puntuales descritas se realizaron en el contexto de la proteína TCERG1 *wild type*. La expresión de todos estos mutantes fue prácticamente idéntica tal y como se muestra por análisis de western blot (Figura R-13).



**Figura R-12. Residuos de fenilalanina estrictamente conservados dentro de los dominios FF4/FF5 son requeridos para la localización de la proteína en los *speckles* nucleares.** La construcción ECFP-TCERG1[1-662]-FF4/FF5 y las diferentes construcciones con una o dos mutaciones puntuales en los residuos de fenilalanina de los dominios FF4/FF5 fueron transfectadas transitoriamente en células HEK293T (verde, columna izquierda: ECFP). El marcaje para poner de manifiesto la presencia de los *speckles* nucleares se realizó con anticuerpos anti-SC35 (rojo, columna central: SC35). La superposición de ambas imágenes se muestra en la columna derecha del panel (SUPERP). Barra= 3 $\mu$ m.

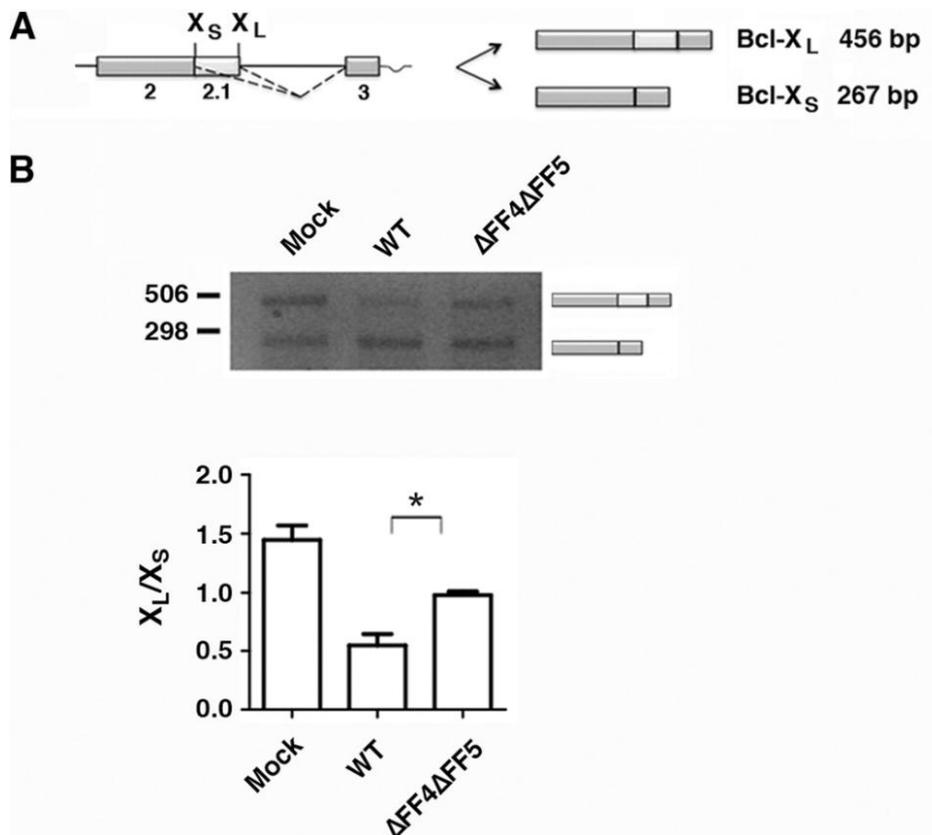


**Figura R-13. Análisis de la expresión de las construcciones de TCERG1 con mutaciones puntuales en los residuos de fenilalanina.** Una muestra de los lisados celulares de células HEK293T transfectadas transitoriamente con la construcción TCERG1[1-662]-FF4/FF5 y las diferentes construcciones con mutaciones en los residuos de fenilalanina fue analizada mediante western blot con anticuerpos específicos frente al epítipo T7 y Cdk9. El peso molecular de los marcadores (M, en kD) queda indicado a la izquierda del panel.

Estos resultados demuestran que TCERG1 puede mantener su localización típica en los *speckles* nucleares aun cuando los residuos F-903 y F-946, ambos en el dominio FF4, se encuentran sustituidos por alaninas. Esto podría deberse a que el plegamiento de este dominio se ve poco alterado por dichas mutaciones, lo cual es sugerido a la luz de nuestros resultados de resonancia magnética nuclear (NMR) que demuestran el mayor grado de estabilidad del dominio FF4. Además, se demuestra la gran implicación del dominio F-961, dentro del dominio FF5, en la correcta localización de TCERG1 en los *speckles* nucleares.

#### 1.4. Papel de los dominios FF4/FF5 en el efecto de TCERG1 sobre el *splicing* alternativo

Como se ha descrito en el apartado “Introducción”, TCERG1 está implicada en el proceso de *splicing*, habiéndose sugerido un papel regulador en el procesamiento de diferentes pre-ARNm. En este punto nos planteamos si los dominios FF4/FF5 de TCERG1 (es decir, una correcta localización subnuclear de la proteína) tendrían un papel en la regulación del *splicing* alternativo. Estudios recientes de nuestro laboratorio han demostrado que TCERG1 regula el procesamiento alternativo de los transcritos procedentes del gen apoptótico *Bcl-x* (Montes et al., 2012) (ver “Introducción”, apartado 3.2 y 3.3). El gen *Bcl-x* presenta dos sitios 5' de *splicing* (5'ss) alternativos en el exón 2. El uso del sitio 5'ss situado *upstream* da lugar a la forma corta o pro-apoptótica de la proteína ( $Bcl-x_S$ ), mientras que el uso del sitio 5'ss situado *downstream* genera una isoforma larga o anti-apoptótica ( $Bcl-x_L$ ) (Figura R-14, A) (Boise et al 1993).



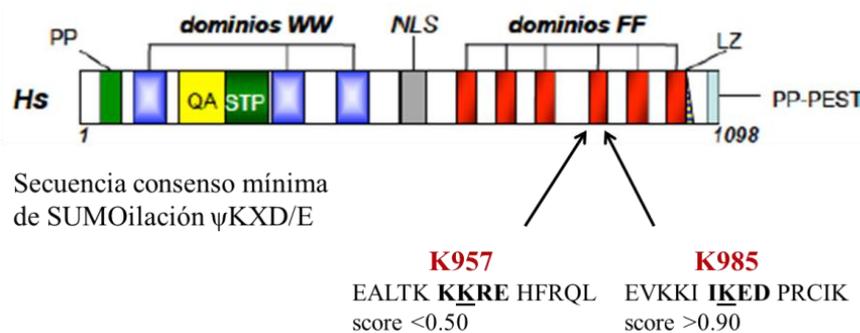
**Figura R-14 (página anterior). La delección de los dominios FF4/FF5 altera el efecto de TCERG1 sobre el *splicing* alternativo del *Bcl-x*.** (A) Esquema representativo de la estructura del minigen reportero del *Bcl-x*, con exones (cajas) e intrones (líneas). Se indican los 5' ss alternativos  $x_L$  y  $x_S$  y las dos isoformas que se generan ( $Bcl-x_L$  y  $Bcl-x_S$ ) mediante los dos eventos de *splicing* alternativo (línea punteada). (B) Efecto de la sobre-expresión de TCERG1 y TCERG1 $\Delta$ FF4 $\Delta$ FF5 sobre el *splicing* del minigen *Bcl-x*. Células HEK293T fueron transitoriamente transfectadas con las construcciones indicadas y procesadas por RT-PCR a las 44 horas aproximadamente después de la transfección. En el panel superior se muestra el gel de agarosa al 2% donde se resuelven los productos de *splicing* generados. La gráfica de la parte inferior de la imagen representa el análisis de densitometría como el ratio  $X_L/X_S$  de tres experimentos independientes (media  $\pm$ DS). \*p <0.05.

El estudio lo llevamos a cabo con un minigen reportero del gen *Bcl-x*, que reproduce el patrón de *splicing* endógeno del *Bcl-x* y ha sido muy útil para caracterizar la regulación de este gen (Garneau et al 2005; Montes et al 2012). Para ello cotransfectamos transitoriamente en células HEK293T un vector de expresión vacío o los vectores de expresión de TCERG1 *wild type* o de TCERG1 con la delección de los dominios FF4/FF5 (TCERG1 $\Delta$ FF4 $\Delta$ FF5) junto con el minigen reportero del *Bcl-x*. Tras extraer el ARN total de las células transfectadas con las diferentes combinaciones de plásmidos analizamos mediante RT-PCR el patrón de *splicing* generado en cada condición. Como se muestra en la figura R-14, B, la sobre-expresión de TCERG1 *wild type* favoreció el uso del 5' ss situado *upstream*, provocando por tanto un aumento de la isoforma corta de *Bcl-x* ( $Bcl-x_S$ ). Por su parte, la transfección del mutante de TCERG1 que no contiene los dominios FF4/FF5 afectó en menor medida este cambio de isoforma respecto a la situación control. Este resultado sugiere la implicación de estos dominios en la elección del sitio de *splicing* 5' en el gen *Bcl-x* y apoya nuestros datos estructurales que proponen a los dominios FF4/FF5 como una unidad funcional. Asimismo, estos datos sugieren que el efecto funcional de TCERG1 podría estar condicionado por una correcta localización de la proteína en la periferia de los *speckles* nucleares.

## 1.5. Estudio del mecanismo molecular implicado en la localización de TCERG1 en los speckles nucleares

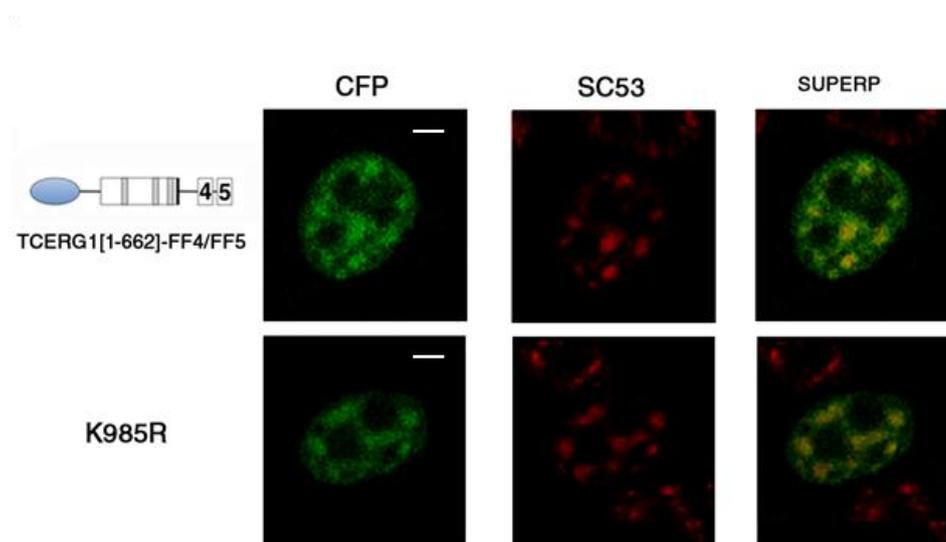
### 1.5.1 El residuo K985, potencial aceptor de SUMO, no influye en la localización de TCERG1 a los speckles nucleares

Estudios previos de nuestro laboratorio han demostrado que las lisinas 503 (lys503) y 608 (lys608) de TCERG1 son modificadas mediante SUMOilación. Estas modificaciones no afectan a la función de *splicing* alternativo llevada a cabo por TCERG1 ni a su localización en los *speckles* nucleares. Sin embargo, mutaciones en las lisinas aceptoras de SUMO potencian la actividad transcripcional de TCERG1, indicando que ésta se encuentra negativamente regulada por la modificación post-traducciona mediada por SUMO (Sánchez-Alvarez et al 2010). El algoritmo SUMOplot (Abgent), que permite la predicción de potenciales sitios de SUMOilación en una secuencia aminoacídica determinada, revela la presencia de dos lisinas dentro del dominio FF5 de TCERG1 enmarcadas en secuencias consenso mínimas aceptoras de SUMOilación ( $\psi$ KXD/E) (Figura-R15). La lisina en posición 957 (K957) muestra una puntuación de 0.48 sobre 1, mientras que la lisina en posición 985 (K985) presenta una puntuación de 0.94, siendo este valor igual que el que ofrecen las lisinas 503 y 608.



**Figura R-15. El dominio FF5 de TCERG1 contiene dos potenciales dianas de SUMOilación.** La imagen muestra la estructura de TCERG1, señalando las lisinas en posición 957 y 985 (K957 y K985) enmarcadas dentro de secuencias consenso mínimas de SUMOilación (resaltadas en negrita para cada una de las lisinas).  $\psi$  es un residuo hidrofóbico de gran tamaño, X puede ser cualquier aminoácido y K la lisina a modificar. Se representan además los dominios proteicos más característicos de TCERG1 (para una mayor explicación, ver la Figura I-6 del apartado de “Introducción”).

Debido a la alta probabilidad de la lisina 985 de ser SUMOilada y a la influencia que esta modificación post-traducciona l tiene en la localización subnuclear de determinadas proteínas (ver “Introducción”, apartado 4.3.2), quisimos conocer si la SUMOilación de K985 podría estar mediando su correcta distribución en los *speckles* nucleares. Para ello decidimos crear una construcción ECFP-TCERG1[1-662]-FF4/FF5 que contuviese el cambio de lisina a arginina en la posición 985 (K985R) y estudiar su localización nuclear mediante microscopía confocal (Figura R-16).



**Figura R-16. La localización nuclear de TCERG1 es independiente de la modificación de la lisina en posición 985.** Células HEK293T fueron transfectadas con las construcciones ECFP-TCERG1[1-662]-FF4/FF5 y la construcción portadora de la mutación K985R (indicadas en la parte izquierda de la imagen) y procesadas para su estudio mediante inmunofluorescencia. El panel muestra la señal de CFP (verde, columna izquierda: CFP) para cada una de las construcciones y la señal de SC35 (rojo, columna central: SC35). Se muestra además la superposición de ambas imágenes (columna derecha: SUPERP). Barra= 3µm.

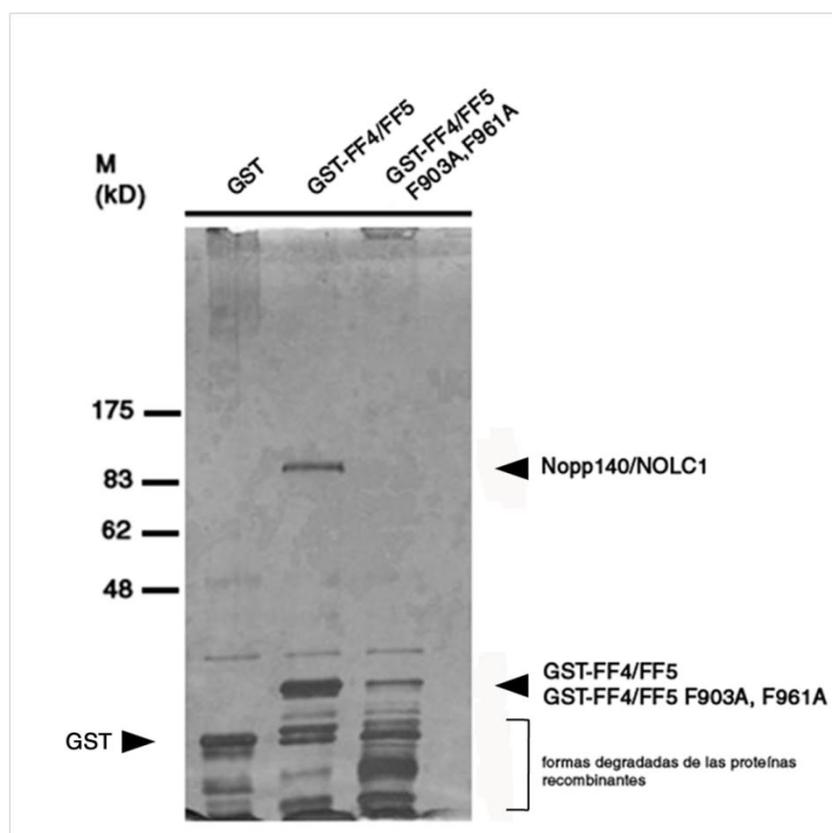
Al igual que en experimentos anteriores, la sobre-expresión transitoria de la construcción TCERG1 wild *type* en células HEK293T mostró una distribución similar a la de la proteína endógena, es decir, fuertemente acumulada en los *speckles* nucleares, puestos de manifiesto mediante tinción frente al factor esencial de *splicing* SC35. El cambio de la lisina en posición 985 a arginina (K985R) no produjo ningún efecto en la localización subnuclear de la construcción portadora de la mutación puntual, que mostró también una acumulación significativa en los *speckles* nucleares. Se obtuvieron

idénticos resultados al cambiar la lisina 985 (lys985) en la construcción ECFP-TCERG1[1-1098]. Estos resultados demuestran que la sustitución de la lisina en posición 985 por arginina no provoca ningún cambio en la localización de TCERG1 y que una potencial SUMOilación del dominio FF5 de TCERG1 a través de esa lisina no estaría mediando su correcta localización en los *speckles* nucleares.

### ***1.5.2 La proteína nucleolar Nopp140 (NOLC1) interacciona in vitro con los dominios FF4/FF5 de TCERG1***

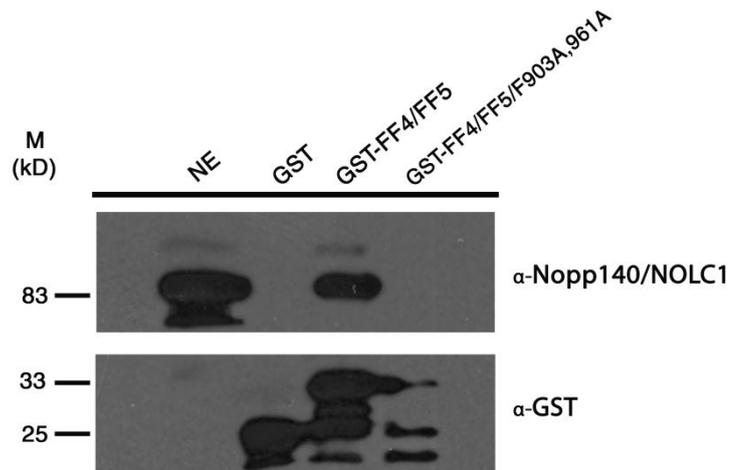
Es posible que la localización mediada por los dominios FF4/FF5 de TCERG1 a los *speckles* nucleares sea debida a interacciones con otras proteínas. Con el fin de identificar proteínas de unión a los dominios FF4/FF5, generamos construcciones que contuviesen los dominios FF4/FF5 de TCERG1 o FF4/FF5 con la doble mutación F903, 961A fusionadas a la proteína GST (GST-FF4/FF5 y GST-FF4/FF5/F903A,F961A). Tras la producción de las proteínas recombinantes en la cepa BL21 pLys1 de *E.Coli*, éstas fueron inmovilizadas en una resina de agarosa acoplada a glutatión y usadas para capturar proteínas contenidas en extractos nucleares de células HeLa (ver apartado 5 de “Materiales y métodos”, Ensayos de *pull down*). Las proteínas unidas fueron separadas mediante geles SDS-PAGE y reveladas mediante tinción de plata (Figura R-17). El análisis de estos geles reveló la presencia de una proteína de aproximadamente 83 kD asociada de forma específica a los dominios FF4/FF5 y no a la proteína GST. No podemos descartar que la ausencia de unión de esta proteína a la versión mutada de los dominios (GST-FF4/FF5/ F903, F961A) pueda deberse a la incorrecta expresión de esta proteína de fusión que generó numerosos productos de degradación, tal y como se muestra en las figuras R-17 y R-18. Nuevos experimentos *in vivo* con la construcción GST-FF4/FF5 F903A, F961A tratan de confirmar la ausencia de esta interacción.

La banda correspondiente a la proteína identificada (Figura R-17) y una banda de tamaño y localización equivalente en el carril de la proteína GST y GST-FF4/FF5/F903A, F961A fueron escindidas y mandadas para su estudio mediante espectrometría de masas (MALDI TOF-TOF) a la Unidad de Proteómica del Centro Nacional de Biotecnología (CSIC). El resultado de las masas obtenidas, en dos experimentos independientes, identificó a la fosfoproteína de localización preferente en el nucléolo Nopp140 (NOLC1) en las muestras correspondientes al carril de GST-FF4/FF5.



**Figura R-17. Identificación de las interacciones proteicas establecidas a través de los dominios FF4/FF5 de TCERG1.** Las proteínas recombinantes GST, GST-FF4/FF5 y GST-FF4/FF5/F903A, F961A fueron usadas para unir proteínas contenidas en extractos nucleares de células HeLa. En la figura se muestra la tinción de plata del gel SDS-PAGE al 10% en el que se separaron las proteínas. El ensayo incluyó a la proteína GST sola como control de uniones inespecíficas. La posición de la banda correspondiente a la proteína de 83 kD identificada mediante el experimento de *pull down* es señalada mediante una flecha a la derecha de la imagen. Se indica también la localización relativa de las proteínas recombinantes GST, GST-FF4/FF5 y GST-FF4/FF5/F903A, F961A así como las posibles formas degradadas de ambas proteínas. El peso molecular de los marcadores (M, en kD) queda indicado a la izquierda del panel.

La presencia de esta proteína en las muestras de los precipitados obtenidos en los experimentos de *pull-down* fue confirmada mediante western blot con anticuerpos específicos frente a NOLC1 (Figura R-18). Estos resultados sugieren que la localización de TCERG1 a los *speckles* nucleares podría estar mediada a través de su interacción con la proteína nucleolar Nopp140. Experimentos en curso tratan de confirmar la interacción de TCERG1 con NOLC1 *in vivo* así como la localización específica dentro del núcleo donde tiene lugar dicha interacción.

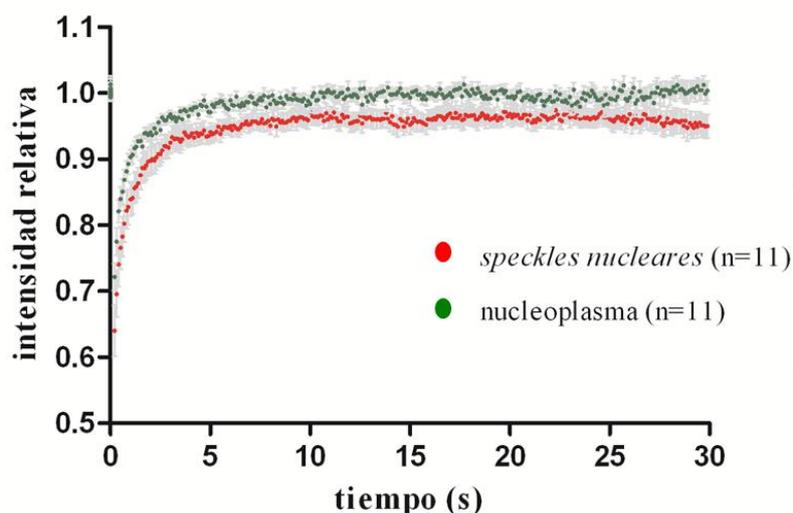
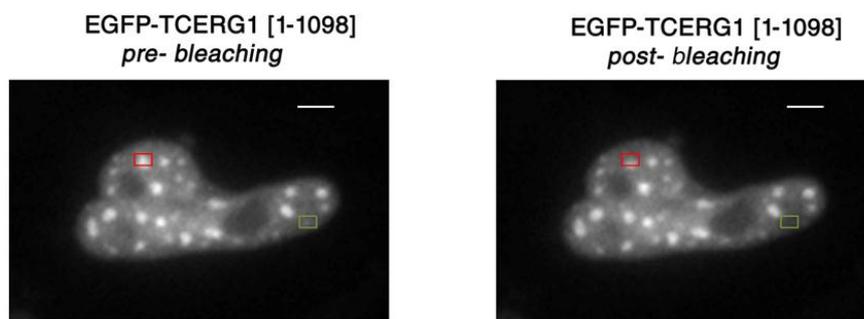


**Figura R-18. Los dominios FF4/FF5 de TCERG1 interactúan *in vitro* con la proteína nucleolar Nopp140/NOLC1.** Una muestra de los eluidos obtenidos en los experimentos de *pull down* fue analizada mediante western blot. La presencia de Nopp140/NOLC1 fue corroborada mediante anticuerpos específicos frente a esta proteína ( $\alpha$ -Nopp140/NOLC1), demostrando su interacción específica con los dominios FF4/FF5 de TCERG1. Las diferentes proteínas recombinantes GST son puestas de manifiesto con anticuerpos específicos frente a ella ( $\alpha$ -GST). El carril izquierdo del gel pone de manifiesto la presencia de NOLC1 en una muestra de extracto nuclear (NE) de células HeLa. El peso molecular de los marcadores (M, en kD) queda indicado a la izquierda del panel.

## 2. DINÁMICA DE TCERG1 EN CÉLULA VIVA

### 2.1 Cinética de TCERG1 en los *speckles* nucleares y nucleoplasma en célula viva

Estudios previos han demostrado que los factores de *splicing* presentan una elevada movilidad y un rápido intercambio entre los *speckles* nucleares y el nucleoplasma circundante (Misteli et al 1997; Phair and Misteli, 2000; Rino J et al. 2007). Para conocer la cinética de TCERG1 en célula viva, llevamos a cabo experimentos de FRAP (ver el apartado de “Inmunofluorescencia, FISH, FRAP y procedimientos de microscopía” de “Materiales y métodos”, para una mayor explicación del procedimiento) en células U2OS transfectadas transitoriamente con una construcción conteniendo la proteína EGFP fusionada al extremo aminoterminal de TCERG1. La proteína generada mostró una distribución similar a la de la proteína de fusión ECFP-TCERG1[1-1098] y a la de la proteína TCERG1 endógena, con una distribución difusa por el nucleoplasma y una clara acumulación en los *speckles* nucleares (Figura R-19, A). La fluorescencia debida a la expresión de la proteína EGFP-TCERG1 de una área de una superficie de  $0.96 \mu\text{m}^2$  en la región de los *speckles* nucleares y del nucleoplasma fue irreversiblemente eliminada usando un láser de alta intensidad. Estas áreas fueron tomadas lo más alejadas posibles la una de la otra para evitar la interferencia en la recuperación de la señal en ambas zonas. La toma secuencial de imágenes durante 30 segundos permitió observar la recuperación de la fluorescencia en ambos compartimentos, confirmándose la movilidad de la población nucleoplasmática de TCERG1 y su intercambio entre los *speckles* y el nucleoplasma (Figura R-19, B). El tiempo medio de recuperación ( $t_{1/2}$ ) obtenido fue inferior a 1 segundo en ambos casos, lo que demuestra el rápido movimiento de TCERG1 en el núcleo (Figura R-19, C). El análisis de la fracción inmóvil (F.I.) mostró que existe una pequeña población de la proteína que se encuentra inmovilizada en los *speckles* nucleares (F.I. = 14%), un comportamiento observado previamente para otros factores de *splicing* como SRSF1, U2AF<sup>65</sup> o SRm160 (Phair and Misteli, 2000; Wagner et al 2004; Rino J et al 2007). En el nucleoplasma, sin embargo, no se detectó F.I. alcanzándose una recuperación completa de la fluorescencia en aproximadamente 10 segundos (Figura R-19, B).



	$t_{1/2}$ (s)	F.M. (%)	F.I. (%)
<i>speckles</i> nucleares	0.5	86	14
nucleoplasma	0.5	100	0

**Figura R-19. Análisis mediante FRAP del comportamiento dinámico de TCERG1 en diferentes compartimentos nucleares.** (A) Imagen representativa de una célula U2OS transitoriamente transfectada con la construcción EGFP-TCERG1[1-1098]. Se muestran las imágenes antes (*pre-bleach*) y justo después de la eliminación de la fluorescencia (*post-bleach*) en los *speckles* nucleares (cuadrado rojo) y en el nucleoplasma (cuadrado verde). Barra= 3 $\mu$ m. (B) Curvas de recuperación de la fluorescencia de EGFP-TCERG1 en los *speckles* nucleares (en rojo) y en el nucleoplasma (en verde). Cada punto de la curva corresponde al valor medio de intensidad relativa del número de células indicadas entre paréntesis (n) que mostraron niveles similares de expresión de la proteína EGFP-TCERG1[1-1098]. La intensidad de la señal fue corregida con el valor de la intensidad *background* y el total de fluorescencia perdida en la

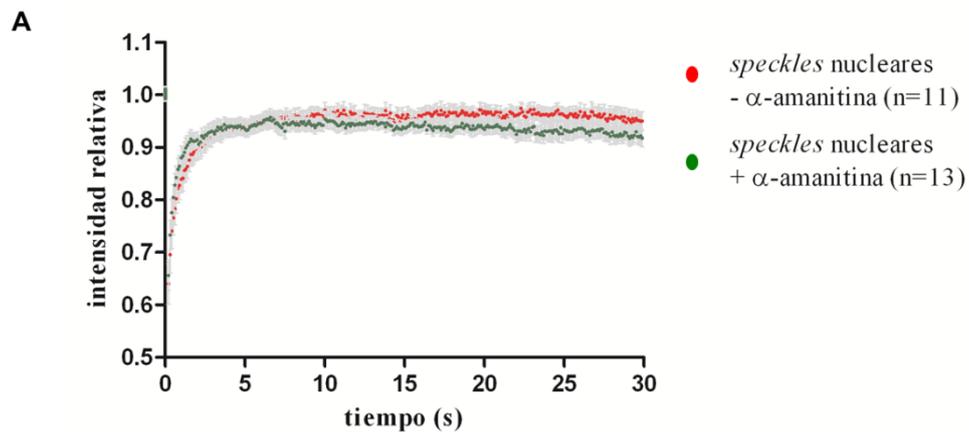
(continuación pie Figura R-19) célula a lo largo del experimento. Las barras representan la DS para cada medida. (C) Valores experimentales obtenidos a partir de las curvas de recuperación de la fluorescencia normalizadas a 0 para el primer valor de intensidad *post-bleach*.  $t_{1/2}$ , tiempo medio de recuperación; F.M., fracción móvil; F.I., fracción inmóvil (Ver “Materiales y métodos”, “Inmunofluorescencia, FISH, FRAP y procedimientos de microscopía”, para mayor detalle del tratamiento de los datos).

Los datos obtenidos muestran que el comportamiento cinético de TCERG1 es prácticamente idéntico en los *speckles* nucleares y en el nucleoplasma, quizás por su localización específica en la periferia de estos dominios.

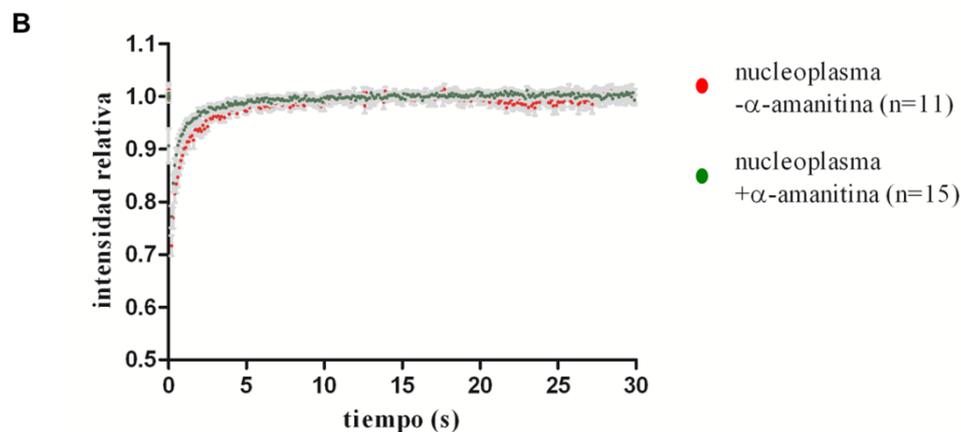
## ***2.2 El intercambio de TCERG1 entre los speckles nucleares y el nucleoplasma es independiente de la existencia de transcripción activa en la célula***

Como ya se ha comentado con anterioridad, los factores de *splicing* se acumulan de manera más prominente en los *speckles* nucleares cuando los procesos de transcripción y procesamiento de los pre-ARNm se encuentran inhibidos. El número de estos orgánulos se reduce y además su morfología sufre un cambio muy característico, pasando a ser más redondeados y expandidos (O’Keefe, R.T. 1994; Sánchez-Álvarez, M. 2006). A pesar de esta aparente inmovilización de las proteínas en los *speckles* nucleares, se ha demostrado que los factores de *splicing* presentan una mayor movilidad en el núcleo ante condiciones de inhibición de la transcripción, probablemente por la falta de interacción con otros componentes del *spliceosoma* y la ausencia de sitios de unión en los transcritos recién sintetizados (Phair y Misteli 2000; Rino J et al. 2007; Huranová M et al. 2010). Para conocer la cinética de TCERG1 en ausencia de actividad transcripcional, células U2OS fueron transitoriamente transfectadas con la construcción EGFP-TCERG1[1-1098] y tratadas con el inhibidor de la ARN polIII  $\alpha$ -amanitina. Tras el tratamiento, la movilidad de TCERG1 fue estudiada mediante FRAP en las mismas condiciones descritas con anterioridad y su dinámica comparada con la observada en células sin tratar. Como se muestra en la , figura R-20, A, B, el  $t_{1/2}$  de recuperación de TCERG1 en las células tratadas con  $\alpha$ -amanitina es menor con respecto a las células sin tratar, tanto en los *speckles* nucleares como en el nucleoplasma, sugiriendo que

permanece menos tiempo unida a otras proteínas de las maquinarias de transcripción y *splicing* en ambos compartimentos. El efecto de este menor  $t_{1/2}$  se refleja en un desplazamiento de las curvas de recuperación de la fluorescencia hacia la izquierda (Figura R-20, A, B, curvas en verde). Aunque el tratamiento con  $\alpha$ -amanitina resultó en una completa recuperación de la fluorescencia en el nucleoplasma, la F.I. de TCERG1 presente en los *speckles* nucleares no desapareció tras la inhibición de la transcripción, como se ha descrito para otros factores de *splicing* (Rino J et al 2010), sino que aumentó (F.I. = 23%) (Figura R-20, A). Todos estos resultados implican que, independientemente del estado transcripcional de la célula, TCERG1 presenta una elevada movilidad en el núcleo. Además sugieren la existencia de sitios de unión para TCERG1 en la periferia de los *speckles* nucleares no dependientes del proceso de transcripción y *splicing*, lo que sugiere la existencia de una población establemente unida que podría ejercer un papel de organización dentro de estos orgánulos.



<i>speckles nucleares</i>	$t_{1/2}$ (s)	F.M. (%)	F.I. (%)
- $\alpha$ -amanitina	0.5	86	14
+ $\alpha$ -amanitina	0.25	77	23

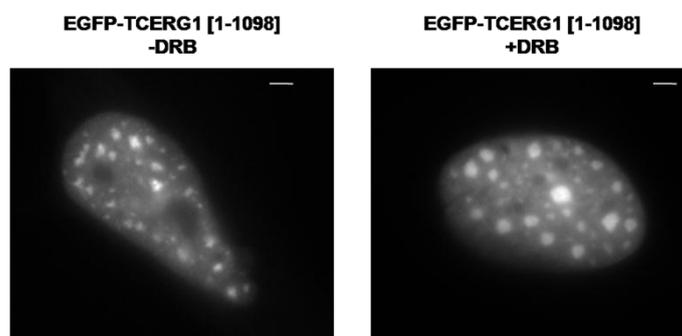


nucleoplasma	$t_{1/2}$ (s)	F.M. (%)	F.I. (%)
- $\alpha$ -amanitina	0.5	100	0
+ $\alpha$ -amanitina	0.3	100	0

**Figura R-20. Efecto de la inhibición transcripcional mediante  $\alpha$ -amanitina en la cinética de TCERG1.** Experimentos de FRAP fueron llevados a cabo de forma simultánea en los *speckles* nucleares (A) y nucleoplasma (B) de células U2OS transitoriamente transfectadas con EGFP-TCERG1[1-1098] sin tratar y tratadas durante 2 horas con  $\alpha$ -amanitina a una concentración de 0.1  $\mu\text{g}/\mu\text{l}$ . Los paneles superiores de cada imagen muestran las curvas de recuperación de la fluorescencia del número de células expresado entre paréntesis (n) que mostraron niveles similares de expresión de la construcción EGFP-TCERG1. Cada dato de las curvas representa la media de intensidad relativa para cada punto  $\pm$  la DS para cada

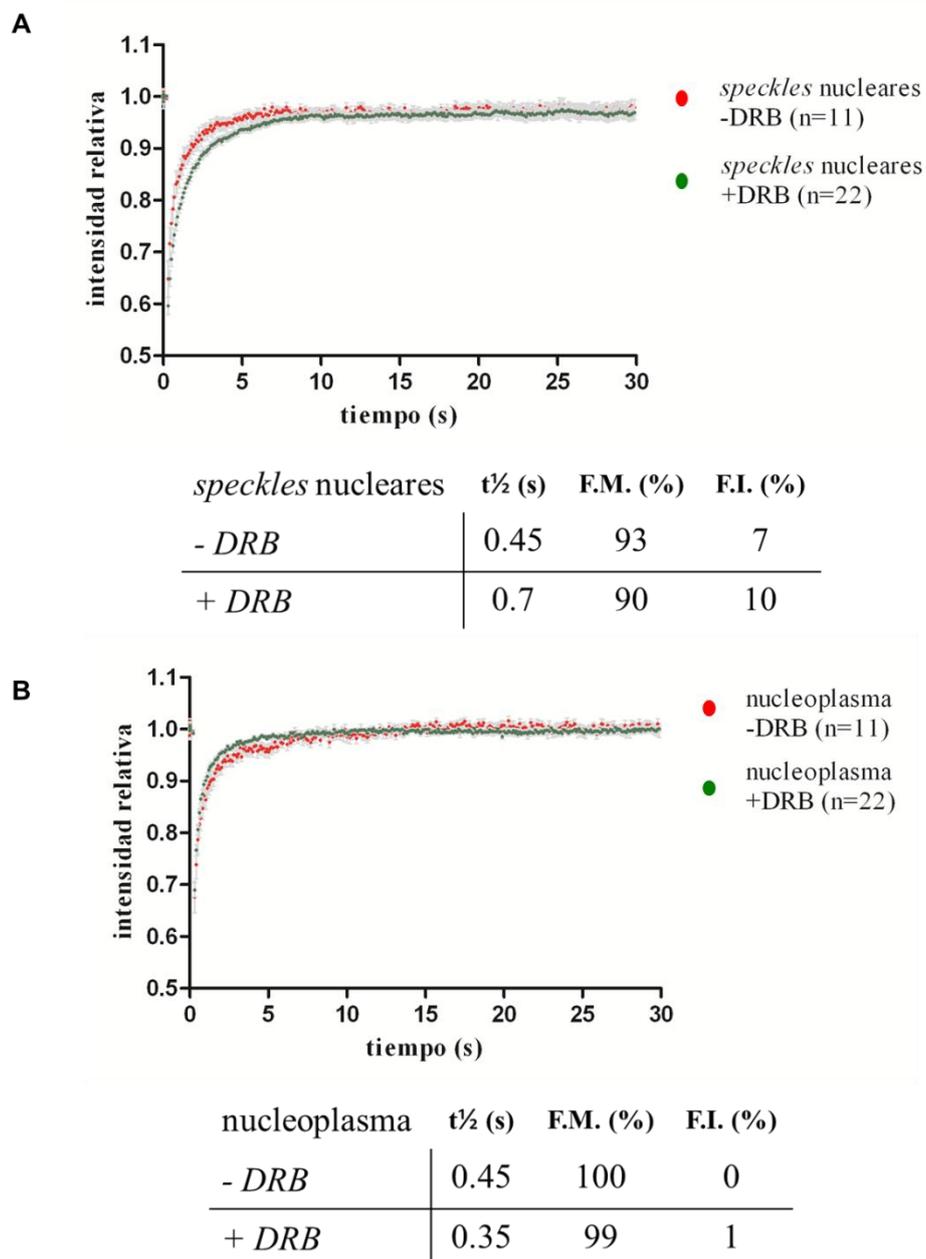
(continuación pie de Figura R-20) medida. Las tablas muestran los valores experimentales obtenidos a partir de las curvas de recuperación de la fluorescencia normalizadas a 0 para el primer valor de intensidad *post-bleach*.  $t_{1/2}$ ; tiempo medio de recuperación, F.M.; fracción móvil (en %); F.I.; fracción inmóvil (en %) (Ver “Materiales y métodos, “Inmunofluorescencia, FISH, FRAP y procedimientos de microscopía”, para mayor detalle del tratamiento de los datos).

Quisimos conocer también el efecto de la inhibición transcripcional ejercida por DRB (5,6-dicloro-1- $\beta$ -ribofuranosilbenzimidazol) sobre el comportamiento dinámico de TCERG1. Este compuesto inhibe a la kinasa Cdk9 y, por tanto, la fosforilación de la serina 2 del extremo CTD de la RNA polIII y la elongación transcripcional. El uso de DRB, al igual que la  $\alpha$ -amanitina, provoca una disminución de la población nucleoplasmática de TCERG1 y su mayor acumulación en los *speckles* nucleares (Figura R-21). Estos resultados apoyan nuestros datos anteriores y muestran que la acumulación de TCERG1 en estos compartimentos es independiente de la fosforilación del dominio CTD de la ARN polII a través de la quinasa Cdk9.



**Figura R-21. El inhibidor transcripcional DRB induce una mayor acumulación de TCERG1 en los *speckles* nucleares de células U2OS.** Imágenes de inmunofluorescencia de células U2OS que expresan transitoriamente EGFP-TCERG1[1-1098] fueron tomadas antes (panel izquierdo) y después (panel derecho) del tratamiento con DRB a 100  $\mu$ M durante 2 horas. La inhibición de la transcripción mediante DRB provoca una redistribución de TCERG1 desde el nucleoplasma hasta los *speckles* nucleares, que sufren un cambio drástico en su morfología. Barra= 0.3  $\mu$ m.

Células U2OS transitoriamente transfectadas con la construcción EGFP-TCERG1[1-1098], fueron tratadas con DRB y sometidas a experimentos de FRAP en las mismas condiciones que en experimentos anteriores. El  $t_{1/2}$  de recuperación obtenido en los *speckles* nucleares resultó ser superior respecto a células en las que la transcripción no se encuentra inhibida y a células tratadas con  $\alpha$ -amanitina (Figura R-22, A). Este retardo en el movimiento, que se refleja en un desplazamiento de la curva de recuperación de la fluorescencia hacia la derecha (Figura R-22, A), refuerza la idea de una conexión entre el reclutamiento de TCERG1 en la periferia de los *speckles* nucleares y su papel en la fase de elongación de la transcripción. En el nucleoplasma, el tratamiento con DRB provocó un  $t_{1/2}$  de recuperación inferior al obtenido en células sin tratar, al igual que ocurría en la inhibición mediante  $\alpha$ -amanitina en este compartimento (observar el desplazamiento de la curva hacia la izquierda, Figura R-22, B). Los datos de la F.I. (Figura R-22, A, B) no mostraron cambios respecto a células sin tratar, lo que una vez más, sugiere la presencia de sitios de unión para TCERG1 en los *speckles* nucleares independientes de la presencia de transcripción activa en la célula.



**Figura R-22. Efecto de la inhibición transcripcional mediante DRB en la cinética de TCERG1.**

Experimentos de FRAP fueron realizados en las mismas condiciones descritas anteriormente sobre regiones de los *speckles* nucleares y del nucleoplasma de células U2OS transitoriamente transfectadas con la construcción EGFP-TCERG1[1-1098] sin tratar o tratadas con DRB a 100  $\mu$ M durante 2 horas. Se muestran las curvas de recuperación de la fluorescencia de ambos compartimentos (A, *speckles* nucleares y B, nucleoplasma) así como los valores experimentales obtenidos a partir de la normalización de las curvas a 0 para el primer valor post-*bleach*. Cada curva representa el valor medio de intensidad relativa  $\pm$  la DS para cada punto en el número de células indicado en la gráfica (n).  $t_{1/2}$ , tiempo medio de recuperación; F.M., fracción móvil; F.I., fracción inmóvil (Ver “Materiales y métodos, “Inmunofluorescencia, FISH, FRAP y procedimientos de microscopía”, para mayor detalle del tratamiento de los datos).

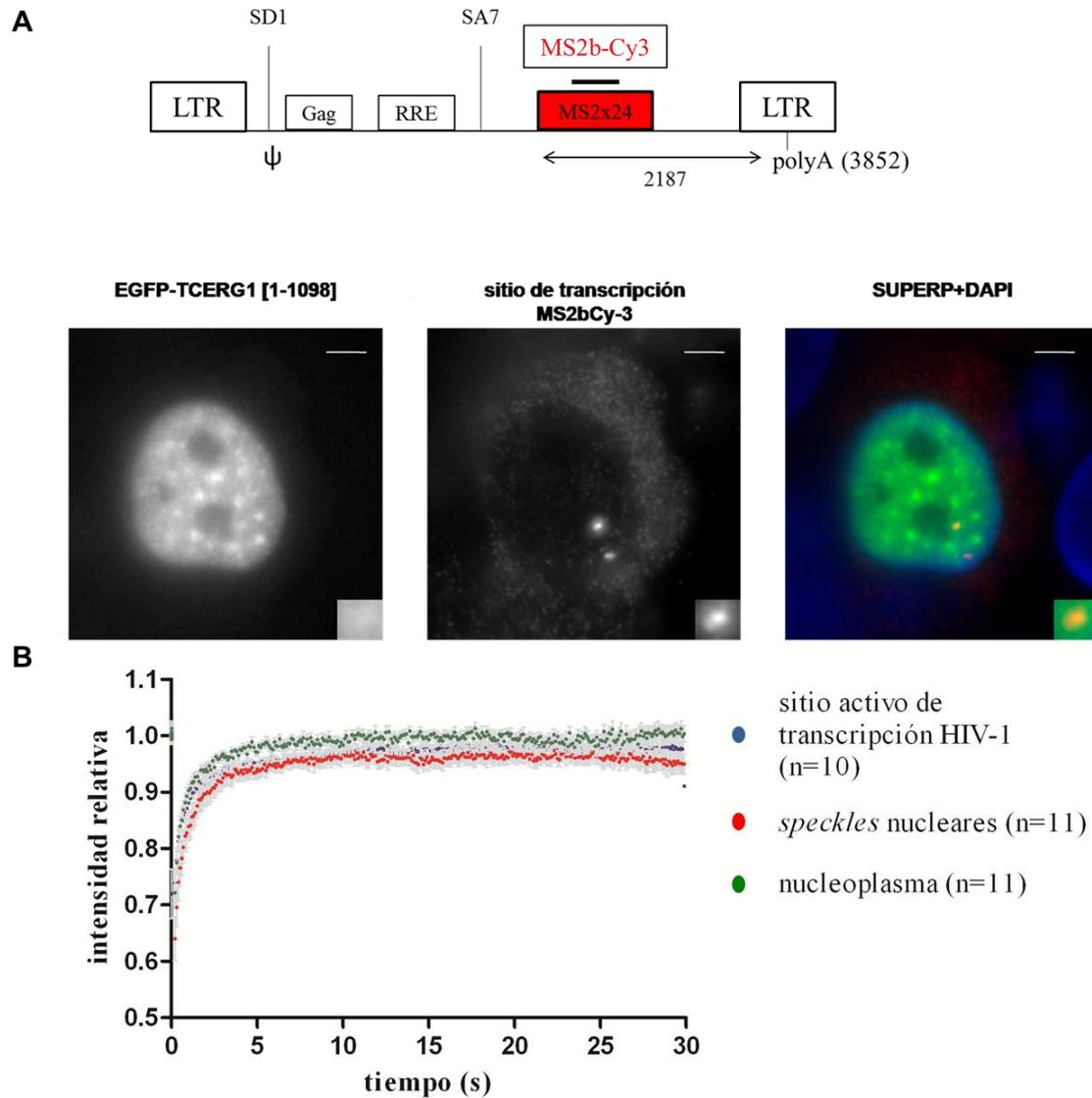
### ***2.3 TCERG1 se localiza en el sitio activo de transcripción del VIH-1 en célula viva***

TCERG1 está implicado en la regulación transcripcional de VIH-1 (ver, “Introducción” apartado 3.2). Por ello nos decidimos a estudiar su posible localización en el sitio activo de transcripción (SAT) viral así como su dinámica de reclutamiento hacia este lugar. Las células U2OS utilizadas en los experimentos de FRAP descritos hasta ahora expresan establemente un vector reportero de la transcripción del VIH-1 con todos los elementos necesarios para la producción del ARNm viral. Además, este reportero contiene 24 sitios de unión para la proteína de la cápside del fago MS2, lo que permite su detección en célula viva (reportero: pExo-MS2x24; línea celular: pU2OS\_Exo1) (Figura R-23, A). Cuando estas células son transfectadas con un plásmido de expresión de la proteína viral Tat se estimula la transcripción del reportero integrado produciendo transcritos que se distribuyen de forma homogénea por el citoplasma y concentrados en un punto específico del núcleo que se corresponde con el lugar donde está teniendo lugar de forma activa la transcripción (Boireau S et al. 2007). El empleo de sondas o proteínas fluorescentes frente a la región MS2 permiten visualizar estos transcritos en célula fijada y célula viva respectivamente, con una alta sensibilidad.

Mediante experimentos de FISH en células pU2OS\_Exo1 transitoriamente transfectadas con la construcción EGFP-TCERG1[1-1098] se observó la co-localización de TCERG1 en el SAT del VIH-1 (Figura R-23, A), situado en la mayoría de los casos en regiones próximas a los *speckles* nucleares. Estudios previos han demostrado que el ensamblaje del *spliceosoma* en el sistema pExo-MS2x24 queda bloqueado tras el reclutamiento del factor U1snRNP (S. Boireau, comunicación personal). La localización de TCERG1 en este sistema sugiere su asociación constitutiva con la maquinaria transcripcional y un reclutamiento mediado por factores asociados al SAT antes de la formación del complejo A del *spliceosoma* (U2snRNP).

Para estudiar el comportamiento dinámico de TCERG1, el SAT del VIH-1 se identificó *in vivo* mediante la proteína de fusión cherry-MS2 donde también se pudo observar la acumulación de TCERG1. Los experimentos de FRAP, eliminando la fluorescencia correspondiente a EGFP-TCERG1 en un área de  $0.96 \mu\text{m}^2$  del SAT, mostraron que el  $t_{1/2}$  en esta localización era similar al obtenido en los *speckles* nucleares y nucleoplasma (Figura 23, C, D). El estudio de la F.I. en el SAT mostró un valor igualmente similar al de los *speckles* nucleares (F.I.= 12%) (Figura R-23, C). De nuevo, esta similitud en el

comportamiento dinámico de TCERG1 en todos los compartimentos estudiados podría explicarse por su localización específica en la periferia de los *speckles* nucleares.

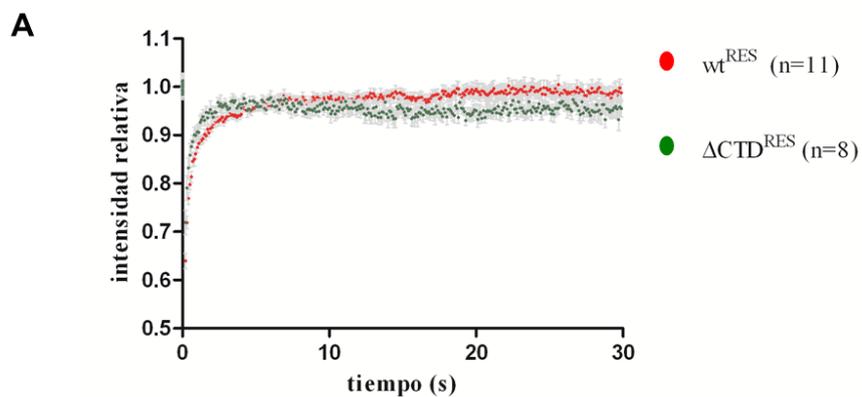


**Figura R-23 (página anterior). TCERG1 localiza en el sitio activo de transcripción del VIH-1. Comportamiento dinámico (A)** Representación esquemática de la construcción pExo-MS2x24 establemente transfectada en células U2OS. El sistema reportero contiene todos los elementos necesarios para síntesis de ARN viral. 5'LTR, *long terminal repeat* 5'; SD1, sitio donador de *splicing*;  $\psi$ , señal de empaquetamiento; RRE, elemento de unión a Rev; SA7, sitio aceptor de *splicing*, flanqueado por secuencias reguladoras (secuencia exónica potenciador del *splicing* y ESS3); 3'LTR, *long terminal repeat* 3'. Se muestra también la secuencia MS2x24 (24 sitios de unión para la proteína viral MS2), región en la que hibrida la sonda MS2bCy3. Células pU2OS\_Exo1, que contienen este reportero, fueron transfectadas transitoriamente con la construcción EGFP-TCERG1[1-1098] y sometidas a experimentos de FISH 24 horas después empleando la sonda MS2b-Cy3. A la izquierda de la imagen se muestra a la localización de TCERG1 en el núcleo de estas células, acumulada en los *speckles* nucleares (EGFP-TCERG1[1-1098]). La imagen central muestra el SAT del VIH-1 donde se acumulan los ARNm recién sintetizados (MS2bCy3). En la parte derecha se muestra la superposición de ambas imágenes más la tinción con DAPI del núcleo (SUPERP+DAPI). En las ampliaciones de las imágenes (esquina inferior derecha) se observa que TCERG1 localiza en el SAT del VIH-1. Barra= 3 $\mu$ m. **(B)** Curvas de recuperación de la fluorescencia en el SAT del VIH-1 (en azul), en los *speckles* nucleares (en rojo) y en el nucleoplasma (en verde). Al igual que en los experimentos anteriores, cada punto de la curva corresponde al valor medio de intensidad relativa del número de células indicadas entre paréntesis, las cuales mostraron niveles similares de expresión de la proteína EGFP-TCERG1[1-1098]. Las barras representan la DS para cada medida. **(C)** Valores experimentales obtenidos a partir de las curvas de recuperación de la fluorescencia normalizadas a 0 para el primer valor de intensidad *post-bleach*.  $t_{1/2}$ ; tiempo medio de recuperación; F.M., fracción móvil; F.I., fracción inmóvil (Ver “Materiales y métodos”, “Inmunofluorescencia, FISH, FRAP y procedimientos de microscopía”, para mayor detalle del tratamiento de los datos).

#### ***2.4 El reclutamiento de TCERG1 al SAT del VIH-1 es independiente del dominio CTD de la ARN polIII.***

Como se ha descrito anteriormente una de las funciones mejor definidas para los *speckles* nucleares es la de servir como “suministro” de factores de *splicing* hasta zonas cercanas donde la transcripción está teniendo lugar de forma activa. Para este reclutamiento se ha descrito una total dependencia del dominio CTD de la ARN polIII (Misteli y Spector, 1999). Dada la interacción de TCERG1 con el dominio CTD (Suñé et al 1997; Cart y et al 2000; Goldstrohom et al 2001) quisimos conocer su implicación en el correcto reclutamiento de TCERG1 hasta el SAT del VIH-1. Para ello células pU2OS\_Exo1 fueron cotransfectadas con la construcción EGFP-TCERG1[1-1098] y plásmidos que expresan polimerasas resistentes a  $\alpha$ -amanitina que contenían un CTD intacto (wt<sup>RES</sup>) o completamente deletado ( $\Delta$ CTD<sup>RES</sup>). Tras el tratamiento con  $\alpha$ -

amanitina el SAT del VIH-1, puesto de manifiesto mediante la proteína MS2-cherry, siguió siendo visible debido a la transcripción mediada por ambas polimerasas mutantes. Mediante inmunofluorescencia corroboramos la presencia de TCERG1 tanto en el SAT promovido por la polimerasa  $wt^{RES}$  como por la polimerasa  $\Delta CTD^{RES}$ , indicando que el dominio CTD de la ARN polII no es esencial para el reclutamiento de TCERG1 hasta el SAT del VIH-1. Sin embargo, los estudios de FRAP, realizados en las mismas condiciones que en experimentos anteriores, mostraron que el  $t_{1/2}$  de recuperación de TCERG1 hasta el SAT del VIH-1 en el contexto de la transcripción mediada por la polimerasa  $\Delta CTD^{RES}$  es menor (es decir, mayor movilidad de la proteína) comparado con su comportamiento bajo la influencia de la polimerasa  $wt^{RES}$  (Figura R-24, A, B). La sobre-expresión de la polimerasa  $\Delta CTD^{RES}$  provocó un aumento de la F.I. de TCERG1 en comparación con lo que sucede con la polimerasa  $wt^{RES}$  (F.I. = 17% *versus* 3%) (Figura R-24, C). Todos estos resultados sugieren que aunque el reclutamiento de TCERG1 hacia el SAT del VIH-1 es independiente de la presencia de un CTD íntegro, este dominio modifica sus propiedades cinéticas en el SAT probablemente regulando la población de TCERG1 que es transferida al complejo competente en elongación, donde establece interacciones con las maquinarias de transcripción y *splicing*.



**B**

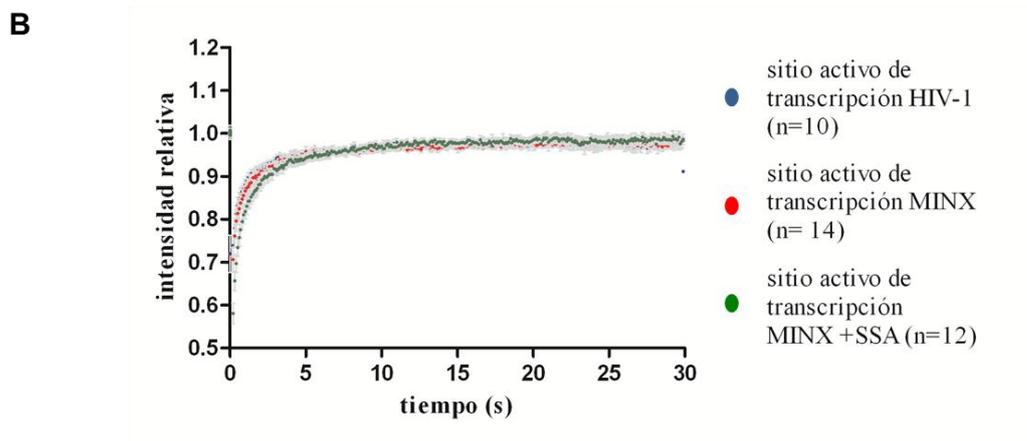
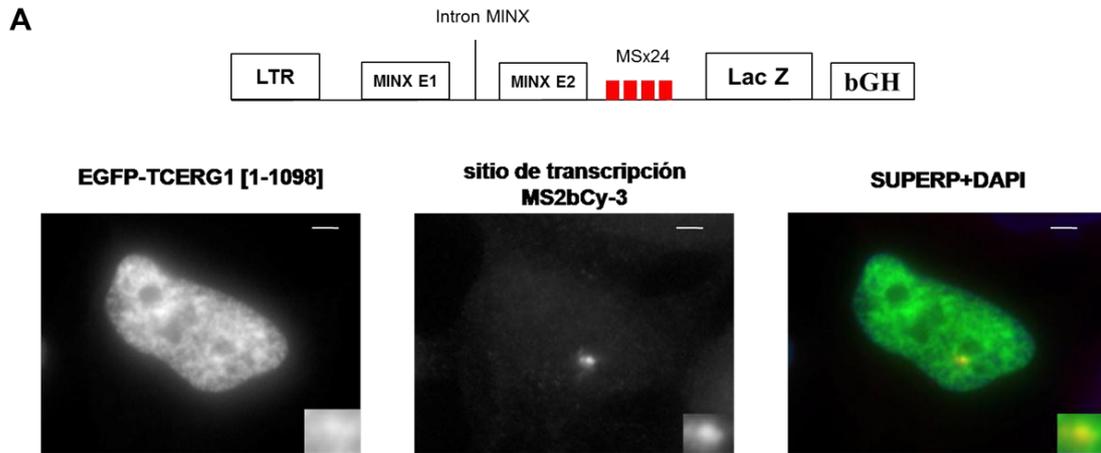
	$t_{1/2}$ (s)	F.M. (%)	F.I. (%)
$wt^{RES}$	0.4	97	3
$\Delta CTD^{RES}$	0.2	83	17

**Figura R-24 (página anterior). El comportamiento cinético de TCERG1 es regulado por el dominio CTD de la ARN polIII.** Células pU2OS\_Exo1 fueron transfectadas transitoriamente con la construcción EGFP-TCERG1[1-1098] y vectores de expresión de versiones mutadas de la ARN polIII resistentes a  $\alpha$ -amanitina ( $wt^{RES}$  y  $\Delta CTD^{RES}$ ). Tras 2 horas de tratamiento con  $\alpha$ -amanitina 0.1  $\mu\text{g}/\mu\text{l}$  las células fueron sometidas a estudio de la recuperación de fluorescencia (FRAP) en el SAT del VIH-1-. **(A)** La figura muestra las curvas de recuperación de la fluorescencia (en rojo para la sobre-expresión del mutante  $wt^{RES}$  y en verde para el mutante  $\Delta CTD^{RES}$ ). Cada punto de la curva corresponde al valor medio de intensidad del número de células indicado entre paréntesis  $\pm$  la DS. Las células incluidas en el estudio mostraron niveles similares de expresión de la proteína EGFP-TCERG1[1-1098] (no mostrado). **(B)** Valores experimentales obtenidos a partir de las curvas de recuperación de la fluorescencia normalizadas a 0 para el primer valor de intensidad *post-bleach*.  $t_{1/2}$ , tiempo medio de recuperación; F.M., fracción móvil; F.I., fracción inmóvil (Ver “Materiales y Resultados”, “Inmunofluorescencia, FISH, FRAP y procedimientos de microscopía”, para mayor detalle del tratamiento de los datos).

### ***2.5 El splicing co-transcripcional modifica la dinámica de reclutamiento de TCERG1 hacia el SAT.***

Dada la implicación de TCERG1 en la regulación del proceso de *splicing* (ver, “Introducción” apartado 3.3) quisimos estudiar el comportamiento dinámico de TCERG1 en un contexto en el que el *splicing* esté teniendo lugar de manera co-transcripcional. Para ello usamos células U2OS establemente transfectadas con el vector reportero MINX\_MS2in (Figura R-25, A) (Schmidt U et al. 2011). Este vector contiene la secuencia intrónica MINX, derivada del genoma del adenovirus, flanqueada por dos exones. El vector contiene 4 sitios de unión para la proteína MS2 que permiten la detección del SAT mediante experimentos de FISH y FRAP en célula fijada o en célula viva respectivamente. La secuencia del gen *LacZ* y el sitio de corte y poliadenilación del gen *bGH* se sitúan *downstream* de la secuencia MINX. La transcripción de este reportero se encuentra regulada por la secuencia promotora contenida en el extremo 5’LTR del VIH y es inducida, al igual que el reportero pExo-MS2x24 mediante la transfección transitoria de un vector de expresión de la proteína viral Tat (Schmidt U et al. 2011). Estudios previos han demostrado que en este sistema todos los componentes del *spliceosoma* son reclutados hasta el SAT donde el *splicing* ocurre de manera co-transcripcional (Spiluttini, B et al. 2010; Schmidt U et al. 2011).

El análisis mediante hibridación *in situ* en células MINX transitoriamente transfectadas con la construcción EGFP-TCERG1[1-1098] mostró la presencia de TCERG1 en el SAT y *splicing* del sistema MINX (R-25, A).



**C**

	$t_{1/2}$ (s)	F.M. (%)	F.I. (%)
sitio de transcripción HIV-1	0.45	88	12
sitio de transcripción <i>MINX</i>	0.55	93	7
sitio de transcripción MINX+SSA	0.55	97	3

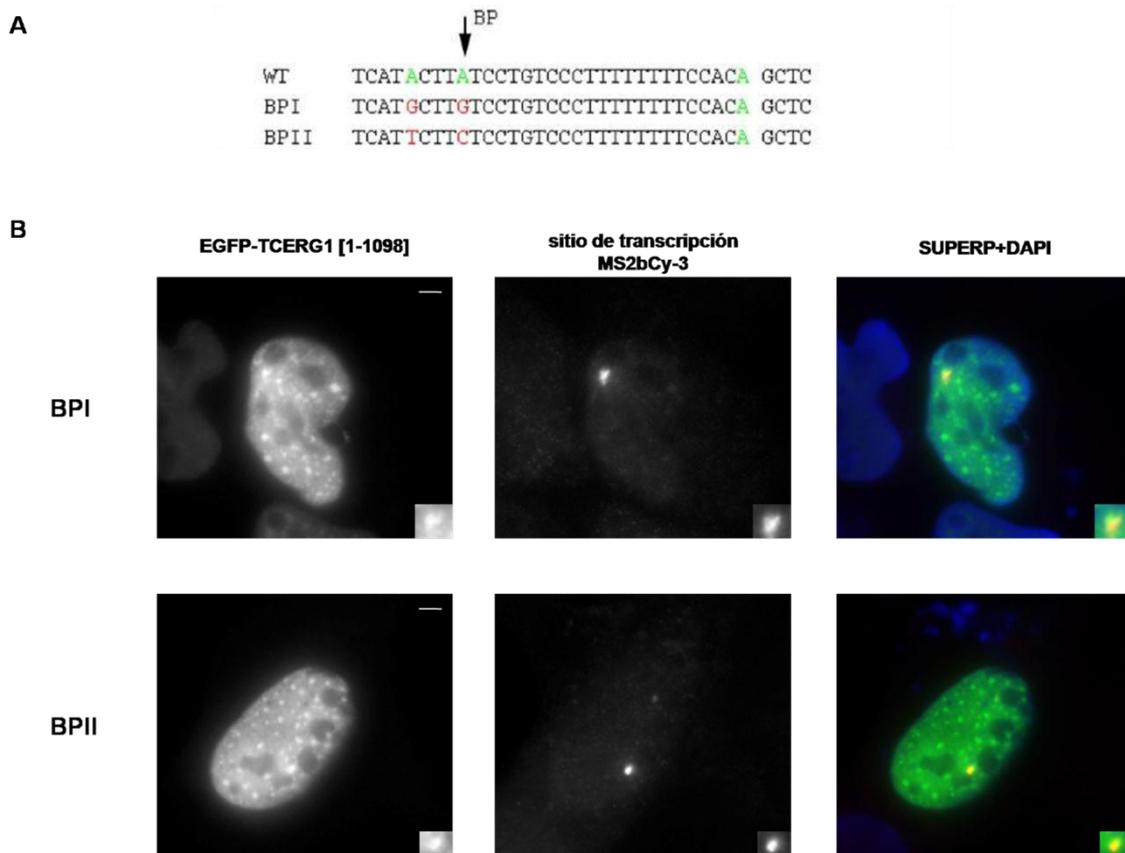
**Figura R-25. TCERG1 localiza en el sitio de transcripción y *splicing* del MINX. Comportamiento dinámico.** (A) Esquema representativo del sistema MINX\_MS2in. Se indican los elementos presentes en el vector: 5'LTR, *long terminal repeat* 5'; MINX E1 y MINX E2, secuencias exónicas derivadas del genoma de adenovirus; LacZ, gen *Lac Z*; bGH, secuencias del sitio de corte y poliadenilación del gen de la hormona de crecimiento bovina. Se indican también los 4 sitios de unión para la proteína MS2. Las células U2OS conteniendo este reportero fueron transfectadas transitoriamente con la construcción EGFP-TCERG1[1-1098] y sometidas a experimentos de FISH 24 horas después de la transfección empleando la

(continuación pie de Figura R- 25) sonda MS2b-Cy3. A la izquierda de la imagen se muestra la localización de TCERG1 en el núcleo de estas células, observándose su acumulación en los *speckles* nucleares (EGFP-TCERG1[1-1098]). La imagen central muestra el SAT donde se acumulan los ARNm recién sintetizados (Ms2bCy3). En la parte derecha se muestra la superposición de ambas imágenes más la tinción con DAPI del núcleo (SUPERP+DAPI). Las ampliaciones de las imágenes (esquina inferior derecha) muestran que TCERG1 localiza en el SAT donde el *splicing* ocurre de manera co-transcripcional. Barra= 3 $\mu$ m. **(B)** Curvas de recuperación de la fluorescencia en el SAT del VIH-1 (en azul), en el SAT y *splicing* del sistema MINX (en rojo) y en este mismo lugar en células tratadas con el inhibidor *spliceostatina* (SSA) (en verde). Cada punto de la curva corresponde al valor medio de intensidad relativa  $\pm$  la DS del número de células indicadas entre paréntesis que mostraron niveles similares de expresión de la proteína EGFP-TCERG [1-1098]. **(C)** Valores experimentales obtenidos a partir de las curvas de recuperación de la fluorescencia normalizadas a 0 para el primer valor de intensidad *post-bleach*.  $t_{1/2}$ , tiempo medio de recuperación; F.M., fracción móvil; F.I., fracción inmóvil (Ver “Materiales y métodos”, “Inmunofluorescencia, FISH, FRAP y procedimientos de microscopía”, para mayor detalle del tratamiento de los datos).

A continuación nos propusimos conocer las propiedades cinéticas del reclutamiento de TCERG1 hacia el SAT del sistema MINX. Mediante experimentos de FRAP realizados en las mismas condiciones que los experimentos anteriores, TCERG1 mostró un ligero enlentecimiento (mayor  $t_{1/2}$ ) de la proteína respecto al sistema pExo-MS2x24 (Figura R-25, B, C), probablemente debido su interacción con la maquinaria de *splicing* que está siendo reclutada hacia el SAT. Este enlentecimiento, que conlleva el desplazamiento hacia la derecha de la curva de recuperación de la fluorescencia (Figura R-25, B), fue más evidente cuando las células fueron tratadas durante 3 horas con *spliceostatina* (SSA), una pequeña molécula que permite el reclutamiento de todos los componentes del *spliceosoma* pero que sin embargo inhibe la reacción de *splicing* (Schmidt U et al. 2011) (Figura R-25, B, C). La F.I. obtenida en estos experimentos mostró valores inferiores respecto al sistema pExo-MS2x24 (F.I.= 5% y 3% en los sistemas MINX y MINX + SSA, respectivamente) (Figura R-25, C). Todos estos datos apoyan un modelo de interacción de TCERG1 con la maquinaria de *splicing in vivo*, probablemente con componentes del *spliceosoma* que actúan *downstream* del factor U2snRNP en un contexto co-transcripcional.

Para conocer con exactitud los componentes de la maquinaria de *splicing* implicados en el reclutamiento de TCERG1 hacia el SAT del MINX usamos dos nuevas líneas de células U2OS. Estas líneas contienen establemente integrados sistemas reporteros

basados en el sistema MINX\_MS2in con diferentes mutaciones en el sitio *branch point* (BP1 y BP2) (Figura R-26, A). En ambos sistemas el ensamblaje del *spliceosoma* queda bloqueado tras la asociación del factor U1snRNP de forma similar a lo que ocurre en el sistema pExo-MS2x24 (Schmidt U., comunicación personal). Las células BP1 y BP2 fueron transitoriamente transfectadas con la construcción EGFP-TCERG1[1-1098] y sometidas a experimentos de FISH con la sonda MS2bCy3. En ambos casos pudimos observar la localización de TCERG1 en el SAT (Figura R-26, B).

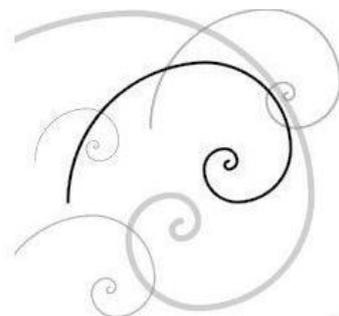


**Figura R-26. Mutaciones en el sitio *branch point* (BP) que impiden el reclutamiento de factores de *splicing* no afectan al reclutamiento de TCERG1 al SAT.** (A) Secuencia del sitio 3' de *splicing* (3'ss) del intrón MINX contenido en el sistema reportero MINX\_MS2in establemente integrado en células U2OS. Se muestra la secuencia *wild type* (WT) y las dos secuencias con mutaciones en el sitio BP (en rojo se resaltan los nucleótidos sustituidos en ambos mutantes). (B) Células U2OS, que contienen establemente transfectado los vectores reporteros MINX\_MS2in BPI (imágenes superiores) y MINX\_MS2in BPII (imágenes inferiores) fueron transfectadas transitoriamente con la construcción EGFP-TCERG1[1-1098] y sometidas a experimentos de FISH 24 h después de la transfección empleando la sonda MS2b-Cy3. A la izquierda de las imágenes se muestra a la localización de TCERG1 en el núcleo de estas células, observándose su acumulación en los *speckles* nucleares (EGFP-TCERG1[1-1098]). La imagen central muestra el SAT donde se acumulan los ARNm recién sintetizados (Ms2bCy3). En la parte

(continuación pie de Figura R- 26) derecha se muestra la superposición de ambas imágenes más la tinción con DAPI del núcleo (SUPERP+DAPI). En las ampliaciones de las imágenes (esquina inferior derecha) se observa que TCERG1 localiza en el SAT incluso cuando el ensamblaje del spliceosoma queda detenido. Barra= 3 $\mu$ m.



## *Conclusiones*





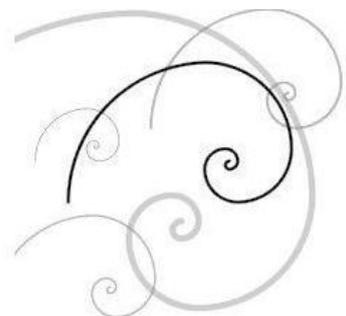
1. La secuencia correspondiente a los dominios FF4 y FF5 (aminoácidos 878 a 1021) es la secuencia necesaria y suficiente para la correcta localización de TCERG1 en la periferia de los *speckles* nucleares. Dentro de esta secuencia, el dominio FF5 (aminoácidos 956-1004) es esencial para la correcta localización de TCERG1 en los *speckles* nucleares aunque por sí mismo no es capaz de mediar dicha localización.
2. Los dominios FF4/FF5 representan una nueva señal general de localización hacia la periferia de los *speckles* nucleares ya que constituye una señal suficiente para dirigir a otras proteínas hacia esas regiones nucleares.
3. Los residuos de fenilalanina F-903 y F-946 del dominio FF4 no son esenciales de forma individual o conjunta para la correcta distribución de TCERG1 en los *speckles* nucleares. El residuo F-961 del dominio FF5, es importante para dicha distribución. La combinación de los residuos F-903 y F-961 es crítica para la correcta localización de TCERG1 en los *speckles* nucleares.
4. La delección de los dominios FF4 y FF5 altera el efecto modulador de TCERG1 sobre el *splicing* alternativo del gen *Bcl-x* disminuyendo, respecto a la proteína *wild type*, el uso del 5' *ss* situado *upstream* en el exón 2 y que da lugar a la isoforma corta de *Bcl-x* (*Bcl-x<sub>S</sub>*).
5. La lisina K-985, dentro del dominio FF5 y posible sitio aceptor de SUMO, no influye en la localización de TCERG1 en los *speckles* nucleares. La SUMOilación de dicha lisina no tiene un papel relevante como mecanismo molecular implicado en la distribución de TCERG1.
6. Los dominios FF4 y FF5 interactúan *in vitro* con la proteína nucleolar Nopp140/NOLC1.
7. TCERG1 presenta una elevada movilidad en el interior del núcleo eucariota. Su tiempo medio de recuperación ( $t_{1/2}$ ) es menor a 1 segundo en los *speckles* nucleares y el nucleoplasma, mostrando en ambos compartimentos un comportamiento prácticamente idéntico. Existe una población de TCERG1 que se encuentra establemente unida a los *speckles* nucleares.
8. El intercambio dinámico de TCERG1 entre los *speckles* nucleares y el nucleoplasma circundante es independiente del estado transcripcional de la célula.

9. TCERG1 localiza en el sitio activo de transcripción del HIV-1, siendo su reclutamiento hasta este lugar independiente del dominio CTD de la ARN polIII. El comportamiento dinámico de este reclutamiento es prácticamente idéntico al observado para los *speckles* nucleares y el nucleoplasma.

10. TCERG1 localiza en el sitio de transcripción activa del MINX. El *splicing* co-transcripcional provoca un enlentecimiento en la dinámica de reclutamiento de TCERG1 hacia el sitio de transcripción activa del MINX.

11. El bloqueo del ensamblaje del *spliceosoma* tras el reclutamiento del factor U1snRNP no impide el reclutamiento de TCERG1 hacia el sitio de transcripción activa.

## *Discusión*





Los dominios FF4 y FF5, que comprenden la secuencia desde el aminoácido 878 hasta el aminoácido 1021, son los dominios responsables de la localización de TCERG1 en la periferia de los *speckles* nucleares. Ambos dominios son además necesarios y suficientes para la localización de proteínas heterólogas dentro de estas regiones.

Análisis previos de nuestro laboratorio, utilizando deleciones seriadas de la región carboxiterminal de TCERG1, ya demostraron un papel esencial del dominio FF5 para una correcta localización de la proteína en la periferia de los *speckles* nucleares (Sánchez-Álvarez M et al. 2006). En un principio estos resultados plantearon varias posibilidades: 1) que FF5 fuese el elemento estructural dentro de TCERG1 responsable de dicha localización, 2) que fuese necesaria una combinación específica de dominios entre los que se encontraría el dominio FF5 o , 3) basado en la redundancia observada previamente para las uniones que se establecen a través de los dominios FF (Smith et al 2004), dentro de la secuencia de TCERG1 debiese existir un número mínimo de estos dominios con independencia de su combinación.

Los experimentos llevados a cabo durante esta tesis han demostrado que el dominio FF5 de TCERG1 es esencial para la acumulación de la proteína en los *speckles* nucleares, ya que un mutante que no contiene este dominio no localiza en estas regiones. Sin embargo, los estudios basados en mutantes que contienen la región amino terminal de TCERG1 fusionada a diferentes dominios FF han demostrado que el dominio FF5 no es suficiente por sí mismo para inducir la acumulación de la proteína en los *speckles* nucleares siendo necesaria la presencia del dominio FF4, y no otro dominio FF cualquiera, para la correcta localización de TCERG1. El requerimiento absoluto de estos dominios demuestran que a lo largo de la secuencia de TCERG1 no existen señales de localización redundantes como se ha demostrado para otros factores como SRSF1 (Caceres JF et al. 1997) en donde diferentes dominios pueden ser delecionados sin que exista una pérdida de la localización completa de la proteína en los *speckles* nucleares. Además sugieren la posibilidad de que la localización de TCERG1 se realiza mediante un mecanismo directo, con necesidad absoluta de una señal de localización, ya que a diferencia de lo que se ha descrito para otros factores de *splicing*, como el factor Tra de *Drosophila* (Hedley et al 1995), la deleción de esta señal impide por completo su concentración en los *speckles* nucleares. Como veremos más adelante, este mecanismo directo probablemente este mediado por interacciones proteicas a través de los dominio FF4/FF5.

Comparando la señal responsable de la localización de TCERG1 en los *speckles* nucleares con señales ya identificadas que median esta misma localización (Li and Bingham 1991; Hedley, Amrein et al. 1995; Cáceres, Misteli et al. 1997; Ko, Kelly et al. 2001; Alvarez, Estivill et al. 2003; (Jagiello, Van Eynde et al. 2000; Eilbracht and Schmidt-Zachmann 2001; Wagner, Chiosea et al. 2003), los dominios FF4/FF5 de TCERG1 representan una nueva señal de localización no descrita hasta el momento. A diferencia de los dominios RS, uno de las señales mejor caracterizadas, la localización mediada por los dominios FF4/FF5 a los *speckles* nucleares es independiente de la NLS. En aquellos casos en los que el dominio RS actúa como señal de localización hacia los *speckles* nucleares, una correcta localización hacia el núcleo depende de la señalización ejercida por estos dominios (Cáceres JF et al. 1997) lo que guarda relación con el solapamiento existente entre una NLS bipartita y la secuencia del dominio RS. Sin embargo, para TCERG1 ambos procesos son independientes, y la no localización de la proteína en los *speckles* nucleares no supone la pérdida de su localización nuclear.

### **La SUMOilación y TCERG1**

Una de las preguntas más interesantes que debíamos resolver a partir de estos resultados era la de qué mecanismos moleculares median la señalización ejercida por los dominios FF4/FF5. Estudios previos de nuestro laboratorio han demostrado que TCERG1 es un sustrato de SUMOilación (Sánchez-Álvarez M et al 2010). Aunque la presencia de SUMO en las lisinas 503 (lys503) y 608 (lys608) no afectan a su correcta localización en los *speckles* nucleares, no quisimos descartar un efecto de esta modificación post-traducciona en la señalización llevada a cabo por los dominios FF4/FF5, ya que las lisinas estudiadas hasta el momento se encuentran en la región amino-terminal de TCERG1, carente de función en la distribución subnuclear de la proteína. A pesar de la alta probabilidad de la lisina K985, localizada en el dominio FF5, de ser un sustrato de SUMOilación, los estudios de mutagénesis realizados parecen descartar el papel de SUMO como la base molecular sobre la que se sustenta la localización mediada por los dominios FF4/FF5. Hay que destacar en este punto que la función de la SUMOilación en la organización estructural de un orgánulo nuclear parece tener un efecto casi exclusivo en el mantenimiento de los cuerpos PML, donde la mayoría de sus

componentes contienen motivos de interacción con SUMO (SIM) (Valérie Lallemand-Breitenbach and Hugues de Thé, 2010).

### **TCERG1 podría comportarse como una proteína SR**

La fosforilación es la modificación post-traduccional con más relevancia en el mantenimiento, organización y regulación de la actividad de los *speckles* nucleares (ver apartado 4.3.1 de “Introducción”) (Misteli y Spector, 1997; Misteli T et al 1997). TCERG1 puede ser también modificada mediante fosforilación (Shimada M et al. 1999; Yu LR et al 2007) lo que plantea la posibilidad de un papel de esta modificación en su localización subnuclear. El análisis de posibles sitios de fosforilación, usando el algoritmo de predicción NetPhos 2.0, reveló que a lo largo de la secuencia de TCERG1 existen numerosos residuos de serina y treonina susceptibles de ser fosforilados. Un gran número de estos residuos se concentran en los dominios FF4 y FF5, dando lugar a una secuencia que recuerda en gran medida a los dominios RS de las proteínas SR. De forma interesante, un análisis exhaustivo de la secuencia de TCERG1 demuestra además la presencia del tramo aminoacídico R-E-R-E-Q-H-K-R (Figura D-1) que presenta una gran similitud con el motivo de anclaje (*docking motif*) consenso de la quinasa SRPK1 (R-X-R/K-X-X-X-R) (Ngo et al 2005). Esta secuencia se sitúa en posición amino-terminal respecto al dominio FF4, de manera similar a como se sitúa la secuencia R-V-K-V-D-G-P-R de SRSF1, definida como su *docking motif* para SRPK1, respecto al dominio RS susceptible de fosforilación (Ngo et al 2005). El alineamiento entre especies demuestra que la secuencia R-E-R-E-Q-H-K-R de TCERG1 se encuentra altamente conservada a lo largo de la evolución (Figura D-1).

a) *Docking motif* **SRPK1**  
R-X-R/K-X-X-X-R

b)

<i>Hs</i>	RSEQTKEID <b>REREQHK</b> REEAIQNFKA <sub>1005</sub>
<i>Mm</i>	RSEQTKEID <b>REREQHK</b> REEAIQNFKA <sub>1007</sub>
<i>Dr</i>	RSEQTKEID <b>REREQHK</b> REEAVQHFKA <sub>807</sub>

**Figura D-1 (página anterior). TCERG1 contiene una potencial secuencia de anclaje para la quinasa SRPK1. (a)** Secuencia consenso para el motivo de unión de la quinasa SRPK1 (Ngo et al 2005). **(b)** Alineamiento mediante *ClustalW2* de la secuencia de TCERG1, y sus homólogos, que contiene el potencial motivo de unión para SRPK1 *Hs: Homo sapiens, Mm: Mus musculus, Dr: Danio rerio*. Al final de cada secuencia se indica la posición que ocupa el último aminoácido.

Una atractiva posibilidad es entonces, que al igual que ocurre para otros factores de *splicing*, concretamente para las proteínas SR, TCERG1 deba su localización, desde y hacia los *speckles*, a múltiples eventos de fosforilación mediados por SRPK1. Sin embargo, la delección de la secuencia correspondiente a la posible señal de anclaje para esta quinasa no tiene ningún efecto en la localización de TCERG1 en los *speckles* nucleares (datos no mostrados). Futuros experimentos intentarán determinar el papel exacto que la fosforilación ejerce sobre el comportamiento de TCERG1 así como las posibles quinasas encargadas de su fosforilación.

### **Un posible papel para la unión ácidos nucleicos**

Aunque los dominios FF han sido clasificados típicamente como motivos de unión a fosfopéptidos, principalmente al dominio CTD fosforilado de la ARN polII (Carty, Goldstrohm et al. 2000; Morris and Greenleaf 2000; Allen, Friedler et al. 2002), existen numerosas evidencias que demuestran que estos dominios funcionan como plataformas sobre las que se ensamblan numerosos factores de las maquinarias de transcripción y *splicing* (Smith et al 2004; Gash A et al 2006; Chung S et al 1999; Ester C, Uetz, P, 2008) e incluso pueden unir moléculas de ARN. El factor esencial de *splicing* de levaduras yPrp40 fue inicialmente identificado como una proteína asociada al componente U1 snRNA del *spliceosoma* (Kao HY, Siliciano, PG, 1996), habiéndose determinado que el cambio de la serina en posición 240, situada en el dominio FF2, a fenilalanina (S240F) impide esta asociación. El dominio FF2 de yPrp40 presenta una naturaleza altamente básica, con un pKa de 9.9 lo que sugiere la posibilidad de que otros dominios FF cargados también positivamente puedan funcionar como dominios de unión a ARN. De acuerdo con esta idea, los dominios FF1-3, FF5 y FF6 de TCERG1 (pKa 9.2, 9.1, 6.9, 9.1 y 9.3 respectivamente, (Gash et al 2005)) podrían interaccionar con ARN. En este sentido, Pawson y colaboradores han llegado a observar una

interacción de naturaleza iónica entre los dominios FF1, FF2 y FF3 de TCERG1 y ARN oligomérico bacteriano (Murphy JM et al 2009). Resultados recientes del laboratorio han demostrado que TCERG1 se asocia específicamente con los transcritos endógenos del gen *Bcl-X* (Montes et al, 2012), y de forma directa, nuestros resultados también parecen demostrar la interacción de los dominios FF1 y FF2 de TCERG1 con ARN (datos no mostrados, Ruiz L y Macías MJ). Apoyando esta posible unión a ARN, hrp130, homólogo de TCERG1 en *C. tetans* se asocia a los transcritos recién sintetizados a partir del anillo BR3 (*Balbani ring 3*) (Sun X, Zhao et al 2004). Aunque en hrp130 no se ha descrito un motivo clásico de unión a ARN, tipo RRM, su región aminoterminal presenta una caja RGG, típicamente presentes en proteínas de unión a ARN (Burd CG and Dreyfuss, 1994). TCERG1 y hrp130 se encuentran altamente conservadas con un grado de similitud de entre el 80-90% en la región aminoterminal y un 60-90% en la región carboxiterminal. Este alto grado de conservación evolutiva entre ambas proteínas y los datos de los que disponemos hasta ahora abren la interesante posibilidad de que TCERG1 pueda ser una proteína de unión a ARN. Futuros experimentos deberán ir encaminados a conocer los ligandos de ARN a los que se une TCERG1 y el papel que estas posibles interacciones tendrían en el contexto de los procesos de transcripción, *splicing* y el acoplamiento entre ambos.

Esta más que probable interacción de TCERG1 con ARN plantea la duda de si esta unión puede estar implicada en su localización en los *speckles* nucleares. Datos previos del laboratorio han demostrado que la presencia de TCERG1 en la periferia de los *speckles* nucleares es independiente de la síntesis de nuevos transcritos (Sánchez-Álvarez et al 2006). Lo que no habíamos dilucidado hasta el momento es si TCERG1 puede relacionarse con la población estable de ARN presentes en estos dominios (ver apartado 2.2 de “Introducción”) debiendo a ellos su distribución subnuclear. El tratamiento con ARNasa (datos no mostrados) demostró que la asociación de TCERG1 no se encuentra mediada por ARN, al igual que ocurre con SC35.

## **TCERG1 podría localizar en el nucléolo antes de asociarse con los speckles nucleares**

Aunque no hemos detectado ninguna modificación post-traducciona que sirva de base al establecimiento de interacciones proteicas con los dominios FF4 y FF5 de TCERG1 quisimos identificar proteínas de unión a estos dominios que pudiesen explicar su papel como señal de localización en los *speckles* nucleares. Nuestros resultados de los experimentos de *pull-down* han demostrado la asociación *in vitro* de Nopp140 (NOLC1) a los dominios FF4/FF5 de TCERG1. De manera preliminar, esta unión parece específica para la versión *wild type* de los dominios FF4 y FF5, ya que mutaciones en los residuos F-903 y F-961 impiden dicha asociación. Aunque nuevos experimentos *in vivo* son necesarios tanto para confirmar la asociación específica de Nopp140 con los dominios FF4 y FF5, como para confirmar la ausencia de tal asociación con su versión mutada, estos datos abren interesantes posibilidades.

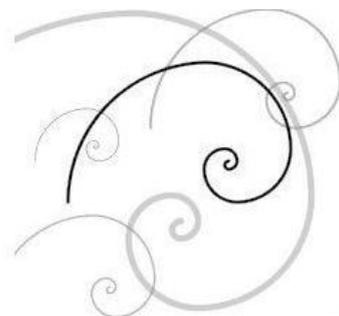
Nopp140 (NOLC1) fue inicialmente caracterizada como una proteína nucleolar de 130 kD situada principalmente en el componente fibrilar denso (CFD) del nucléolo (Meier and Blobel, 1992; Pai, CY, HK Chen et al. 1995), aunque también descrito su localización en el componente fibrilar (CF) (Thiry, M et al. 2009). A diferencia de otras proteínas nucleolares, Nopp140 no contiene motivos de unión a ARN. Su estructura se caracteriza por la presencia de un dominio central rico en residuos de serina que se alternan con repeticiones de lisina, arginina y prolina. La mayoría de las serinas de este dominio central son fosforiladas por la quinasa caseína I, que hace de Nopp140 una de las proteínas más fosforiladas de la célula (Li et al 1997; Meier and Blobel, 1992; Meier, 1996). Dentro del nucléolo, Nopp140 interacciona con factores como NAP57 (Meier and Blobel, 1994), diferentes snoRNPs (Yang et al. 2000), la ARN polII (Chen et al. 1999), o el factor de transcripción TFII-I (Miau et al 1997), todos ellos componentes de las maquinarias de transcripción y procesamiento de los transcritos de los pre-ARNr. Estas interacciones demuestran la implicación de Nopp140 en procesos de biogénesis y transporte de las snoRNPs. En el contexto de localización y dinámica nuclear, quizás una de las interacciones más interesantes que establece Nopp140 es con la coilina, proteína mayoritaria de los cuerpos de Cajal (Isaac C et al. 1998), orgánulo donde Nopp140 también se encuentra presente. Su presencia en ambos compartimentos y la regulación que sobre la estructura de los cuerpos de Cajal ejerce Nopp140, han llevado

a proponer para ella una posible función de chaperona en el movimiento proteico desde el nucléolo hasta los cuerpos de Cajal (Isaac C et al. 1998).

Ante estos resultados, es muy atractiva la idea de que TCERG1 interactúe con Nopp140 en el nucléolo antes de su asociación con los *speckles* nucleares. Al igual que ocurre para los componentes de los cuerpos de Cajal, Nopp140 podría ejercer funciones de chaperona para TCERG1 hasta los *speckles* nucleares, dependiendo de esta unión la correcta localización de la proteína en estos dominios. Apoyando esta posible interacción, estudios de proteómica han demostrado la presencia de TCERG1 en el nucléolo (Andersen JS et al 2005) así como la presencia de Nopp140 en los *speckles* nucleares (Naitoh N, et al. 2004). El paso transitorio de componentes de los *speckles* nucleares por el nucléolo ha sido demostrado previamente. Las proteínas SR no localizan en los *speckles* nucleares inmediatamente después de la mitosis, sino que se concentran transitoriamente alrededor de las regiones organizadoras de nucléolo (NORs) en donde podrían sufrir procesos iniciales de maduración/modificación y ensamblaje (Sacco-Bubulya P et al. 2004). En estas regiones las proteínas SR se encuentran en estado hipofosforilado, lo que junto con la presencia de la quinasa Ctk/STY apoya la idea de una modificación y/o interacción previa de estos factores antes de asociarse a los sitios de transcripción activa y/o los *speckles* nucleares. De igual modo, el paso de TCERG1 por el nucléolo, tras finalizar la mitosis, podría ser importante para posibles modificaciones post-traduccionales y su correcto reclutamiento hasta los sitios de transcripción y *speckles* nucleares. La idea del nucléolo, como un orgánulo distribuidor de factores de la maquinaria de *splicing* tendría una gran importancia tras la formación del núcleo, en la entrada de la célula en interfase, cuando los *speckles* nucleares aún no se han formado.



## *Bibliografía*





Allen, M., A. Friedler, et al. (2002). "The structure of an FF domain from human HYPA/FBP11." *J Mol Biol* **323**(3): 411-416.

Alvarez, M., X. Estivill, et al. (2003). "DYRK1A accumulates in splicing speckles through a novel targeting signal and induces speckle disassembly." *J Cell Sci* **116**(Pt 15): 3099-3107.

Andrade, L. E., E. M. Tan, et al. (1993). "Immunocytochemical analysis of the coiled body in the cell cycle and during cell proliferation." *Proc Natl Acad Sci U S A* **90**(5): 1947-1951.

Andresen, J. M., J. Gayan, et al. (2007). "The relationship between CAG repeat length and age of onset differs for Huntington's disease patients with juvenile onset or adult onset." *Ann Hum Genet* **71**(Pt 3): 295-301.

Arango, M., S. Holbert, et al. (2006). "CA150 expression delays striatal cell death in overexpression and knock-in conditions for mutant huntingtin neurotoxicity." *J Neurosci* **26**(17): 4649-4659.

Aubol, B. E., S. Chakrabarti, et al. (2003). "Processive phosphorylation of alternative splicing factor/splicing factor 2." *Proc Natl Acad Sci U S A* **100**(22): 12601-12606.

Baldin, V., M. Militello, et al. (2008). "A novel role for PA28gamma-proteasome in nuclear speckle organization and SR protein trafficking." *Mol Biol Cell* **19**(4): 1706-1716.

Bedford, M. T. and P. Leder (1999). "The FF domain: a novel motif that often accompanies WW domains." *Trends Biochem Sci* **24**(7): 264-265.

Bernard, D., K. V. Prasanth, et al. "A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression." *EMBO J* **29**(18): 3082-3093.

Bickmore, W. A. and H. G. Sutherland (2002). "Addressing protein localization within the nucleus." *EMBO J* **21**(6): 1248-1254.

Birney, E., S. Kumar, et al. (1993). "Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors." *Nucleic Acids Res* **21**(25): 5803-5816.

Bohne, J., S. E. Cole, et al. (2000). "Expression analysis and mapping of the mouse and human transcriptional regulator CA150." *Mamm Genome* **11**(10): 930-933.

Bond, C. S. and A. H. Fox (2009). "Paraspeckles: nuclear bodies built on long noncoding RNA." *J Cell Biol* **186**(5): 637-644.

Bonet, R., X. Ramirez-Espain, et al. (2008). "Solution structure of the yeast URN1 splicing factor FF domain: comparative analysis of charge distributions in FF domain structures-FFs and SURPs, two domains with a similar fold." *Proteins* **73**(4): 1001-1009.

Bonet, R., L. Ruiz, et al. (2009). "NMR structural studies on human p190-A RhoGAPFF1 revealed that domain phosphorylation by the PDGF-receptor alpha requires its previous unfolding." *J Mol Biol* **389**(2): 230-237.

Boronenkov, I. V., J. C. Loijens, et al. (1998). "Phosphoinositide signaling pathways in nuclei are associated with nuclear speckles containing pre-mRNA processing factors." *Mol Biol Cell* **9**(12): 3547-3560.

Brown, J. M., J. Green, et al. (2008). "Association between active genes occurs at nuclear speckles and is modulated by chromatin environment." *J Cell Biol* **182**(6): 1083-1097.

Caceres, J. F., T. Misteli, et al. (1997). "Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity." *J Cell Biol* **138**(2): 225-238.

Cao, W., S. F. Jamison, et al. (1997). "Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro." *RNA* **3**(12): 1456-1467.

Carmo-Fonseca, M., J. Ferreira, et al. (1993). "Assembly of snRNP-containing coiled bodies is regulated in interphase and mitosis--evidence that the coiled body is a kinetic nuclear structure." *J Cell Biol* **120**(4): 841-852.

Carter, K. C., D. Bowman, et al. (1993). "A three-dimensional view of precursor messenger RNA metabolism within the mammalian nucleus." *Science* **259**(5099): 1330-1335.

Carter, K. C., K. L. Taneja, et al. (1991). "Discrete nuclear domains of poly(A) RNA and their relationship to the functional organization of the nucleus." *J Cell Biol* **115**(5): 1191-1202.

Carty, S. M., A. C. Goldstrohm, et al. (2000). "Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II." *Proc Natl Acad Sci U S A* **97**(16): 9015-9020.

Chattopadhyay, B., S. Ghosh, et al. (2003). "Modulation of age at onset in Huntington's disease and spinocerebellar ataxia type 2 patients originated from eastern India." Neurosci Lett **345**(2): 93-96.

Chen, D. and S. Huang (2001). "Nucleolar components involved in ribosome biogenesis cycle between the nucleolus and nucleoplasm in interphase cells." J Cell Biol **153**(1): 169-176.

Chen, L. L. and G. G. Carmichael (2009). "Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA." Mol Cell **35**(4): 467-478.

Clemson, C. M., J. N. Hutchinson, et al. (2009). "An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles." Mol Cell **33**(6): 717-726.

Cmarko, D., P. J. Verschure, et al. (1999). "Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection." Mol Biol Cell **10**(1): 211-223.

Colwill, K., L. L. Feng, et al. (1996). "SRPK1 and Clk/Sty protein kinases show distinct substrate specificities for serine/arginine-rich splicing factors." J Biol Chem **271**(40): 24569-24575.

Colwill, K., T. Pawson, et al. (1996). "The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution." EMBO J **15**(2): 265-275.

Deckert, J., K. Hartmuth, et al. (2006). "Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions." Mol Cell Biol **26**(14): 5528-5543.

Dias, A. P., K. Dufu, et al. (2010). "A role for TREX components in the release of spliced mRNA from nuclear speckle domains." Nat Commun **1**: 97.

Ding, J. H., X. Y. Zhong, et al. (2006). "Regulated cellular partitioning of SR protein-specific kinases in mammalian cells." Mol Biol Cell **17**(2): 876-885.

Dow, E. C., H. Liu, et al. (2010). "T-loop phosphorylated Cdk9 localizes to nuclear speckle domains which may serve as sites of active P-TEFb function and exchange between the Brd4 and 7SK/HEXIM1 regulatory complexes." J Cell Physiol **224**(1): 84-93.

Dundr, M. and T. Misteli (2001). "Functional architecture in the cell nucleus." Biochem J **356**(Pt 2): 297-310.

Dundr, M. and T. Misteli (2010). "Biogenesis of nuclear bodies." Cold Spring Harb Perspect Biol **2**(12): a000711.

Dundr, M., T. Misteli, et al. (2000). "The dynamics of postmitotic reassembly of the nucleolus." J Cell Biol **150**(3): 433-446.

Dye, B. T. and J. G. Patton (2001). "An RNA recognition motif (RRM) is required for the localization of PTB-associated splicing factor (PSF) to subnuclear speckles." Exp Cell Res **263**(1): 131-144.

Eilbracht, J. and M. S. Schmidt-Zachmann (2001). "Identification of a sequence element directing a protein to nuclear speckles." Proc Natl Acad Sci U S A **98**(7): 3849-3854.

Fakan, S. (1994). "Perichromatin fibrils are in situ forms of nascent transcripts." Trends Cell Biol **4**(3): 86-90.

Fakan, S. and R. van Driel (2007). "The perichromatin region: a functional compartment in the nucleus that determines large-scale chromatin folding." Semin Cell Dev Biol **18**(5): 676-681.

Ferreira, J. A., M. Carmo-Fonseca, et al. (1994). "Differential interaction of splicing snRNPs with coiled bodies and interchromatin granules during mitosis and assembly of daughter cell nuclei." J Cell Biol **126**(1): 11-23.

Fidan, Z., A. Younis, et al. (2011). "Chemical synthesis of the third WW domain of TCERG 1 by native chemical ligation." J Pept Sci **17**(9): 644-649.

Fox, A. H., C. S. Bond, et al. (2005). "P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner." Mol Biol Cell **16**(11): 5304-5315.

Gama-Carvalho, M., R. D. Krauss, et al. (1997). "Targeting of U2AF65 to sites of active splicing in the nucleus." J Cell Biol **137**(5): 975-987.

Gasch, A., S. Wiesner, et al. (2006). "The structure of Prp40 FF1 domain and its interaction with the crn-TPR1 motif of Clf1 gives a new insight into the binding mode of FF domains." J Biol Chem **281**(1): 356-364.

Goldstrohm, A. C., T. R. Albrecht, et al. (2001). "The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1." Mol Cell Biol **21**(22): 7617-7628.

Goldstrohm, A. C., A. L. Greenleaf, et al. (2001). "Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing." Gene **277**(1-2): 31-47.

Gross, S. D., J. C. Loijens, et al. (1999). "The casein kinase Ialpha isoform is both physically positioned and functionally competent to regulate multiple events of mRNA metabolism." J Cell Sci **112** ( Pt **16**): 2647-2656.

Grummt, I. (2006). "Actin and myosin as transcription factors." Curr Opin Genet Dev **16**(2): 191-196.

Gui, J. F., W. S. Lane, et al. (1994). "A serine kinase regulates intracellular localization of splicing factors in the cell cycle." Nature **369**(6482): 678-682.

Hagopian, J. C., C. T. Ma, et al. (2008). "Adaptable molecular interactions guide phosphorylation of the SR protein ASF/SF2 by SRPK1." J Mol Biol **382**(4): 894-909.

Hall, L. L., K. P. Smith, et al. (2006). "Molecular anatomy of a speckle." Anat Rec A Discov Mol Cell Evol Biol **288**(7): 664-675.

Handwerker, K. E. and J. G. Gall (2006). "Subnuclear organelles: new insights into form and function." Trends Cell Biol **16**(1): 19-26.

Hebert, M. D. and A. G. Matera (2000). "Self-association of coilin reveals a common theme in nuclear body localization." Mol Biol Cell **11**(12): 4159-4171.

Hedley, M. L., H. Amrein, et al. (1995). "An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine-rich splicing factor." Proc Natl Acad Sci U S A **92**(25): 11524-11528.

Herrmann, C. H. and M. A. Mancini (2001). "The Cdk9 and cyclin T subunits of TAK/P-TEFb localize to splicing factor-rich nuclear speckle regions." J Cell Sci **114**(Pt 8): 1491-1503.

Holbert, S., I. Denghien, et al. (2001). "The Gln-Ala repeat transcriptional activator CA150 interacts with huntingtin: neuropathologic and genetic evidence for a role in Huntington's disease pathogenesis." Proc Natl Acad Sci U S A **98**(4): 1811-1816.

Holt, I., S. Mittal, et al. (2007). "Defective mRNA in myotonic dystrophy accumulates at the periphery of nuclear splicing speckles." Genes Cells **12**(9): 1035-1048.

Hu, Q., Y. S. Kwon, et al. (2008). "Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules." Proc Natl Acad Sci U S A **105**(49): 19199-19204.

Hu, Y., I. Kireev, et al. (2009). "Large-scale chromatin structure of inducible genes: transcription on a condensed, linear template." J Cell Biol **185**(1): 87-100.

Huang, S., T. J. Deerinck, et al. (1994). "In vivo analysis of the stability and transport of nuclear poly(A)+ RNA." J Cell Biol **126**(4): 877-899.

Huang, S. and D. L. Spector (1996). "Dynamic organization of pre-mRNA splicing factors." J Cell Biochem **62**(2): 191-197.

Huang, S. and D. L. Spector (1996). "Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription." J Cell Biol **133**(4): 719-732.

Hutchinson, J. N., A. W. Ensminger, et al. (2007). "A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains." BMC Genomics **8**: 39.

Iborra, F. J., A. Pombo, et al. (1996). "Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei." J Cell Sci **109** ( Pt **6**): 1427-1436.

Ingham, R. J., K. Colwill, et al. (2005). "WW domains provide a platform for the assembly of multiprotein networks." Mol Cell Biol **25**(16): 7092-7106.

Jablonski, J. A., A. L. Amelio, et al. (2010). "The transcriptional transactivator Tat selectively regulates viral splicing." Nucleic Acids Res **38**(4): 1249-1260.

Jagatheesan, G., S. Thanumalayan, et al. (1999). "Colocalization of intranuclear lamin foci with RNA splicing factors." J Cell Sci **112** ( Pt **24**): 4651-4661.

Jagiello, I., A. Van Eynde, et al. (2000). "Nuclear and subnuclear targeting sequences of the protein phosphatase-1 regulator NIPP1." J Cell Sci **113** Pt **21**: 3761-3768.

Jiang, W., R. Sordella, et al. (2005). "An FF domain-dependent protein interaction mediates a signaling pathway for growth factor-induced gene expression." Mol Cell **17**(1): 23-35.

Johnson, C., D. Primorac, et al. (2000). "Tracking COL1A1 RNA in osteogenesis imperfecta. splice-defective transcripts initiate transport from the gene but are retained within the SC35 domain." *J Cell Biol* **150**(3): 417-432.

Jurica, M. S., L. J. Licklider, et al. (2002). "Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis." *RNA* **8**(4): 426-439.

Kaiser, T. E., R. V. Intine, et al. (2008). "De novo formation of a subnuclear body." *Science* **322**(5908): 1713-1717.

Ko, T. K., E. Kelly, et al. (2001). "CrkRS: a novel conserved Cdc2-related protein kinase that colocalizes with SC35 speckles." *J Cell Sci* **114**(Pt 14): 2591-2603.

Krause, S., S. Fakan, et al. (1994). "Immunodetection of poly(A) binding protein II in the cell nucleus." *Exp Cell Res* **214**(1): 75-82.

Kruhlak, M. J., M. A. Lever, et al. (2000). "Reduced mobility of the alternate splicing factor (ASF) through the nucleoplasm and steady state speckle compartments." *J Cell Biol* **150**(1): 41-51.

Kushner, D. J. (1969). "Self-assembly of biological structures." *Bacteriol Rev* **33**(2): 302-345.

Lamond, A. I. and D. L. Spector (2003). "Nuclear speckles: a model for nuclear organelles." *Nat Rev Mol Cell Biol* **4**(8): 605-612.

Lerner, M. R., J. A. Boyle, et al. (1981). "Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus." *Science* **211**(4480): 400-402.

Li, H. and P. M. Bingham (1991). "Arginine/serine-rich domains of the su(wa) and tra RNA processing regulators target proteins to a subnuclear compartment implicated in splicing." *Cell* **67**(2): 335-342.

Lin, K. T., R. M. Lu, et al. (2004). "The WW domain-containing proteins interact with the early spliceosome and participate in pre-mRNA splicing in vivo." *Mol Cell Biol* **24**(20): 9176-9185.

Mabon, S. A. and T. Misteli (2005). "Differential recruitment of pre-mRNA splicing factors to alternatively spliced transcripts in vivo." *PLoS Biol* **3**(11): e374.

Macias, M. J., S. Wiesner, et al. (2002). "WW and SH3 domains, two different scaffolds to recognize proline-rich ligands." *FEBS Lett* **513**(1): 30-37.

Makarov, E. M., O. V. Makarova, et al. (2002). "Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome." *Science* **298**(5601): 2205-2208.

Mao, Y. S., H. Sunwoo, et al. (2011). "Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs." *Nat Cell Biol* **13**(1): 95-101.

Mao, Y. S., B. Zhang, et al. (2011). "Biogenesis and function of nuclear bodies." *Trends Genet* **27**(8): 295-306.

Marshall, N. F., J. Peng, et al. (1996). "Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase." *J Biol Chem* **271**(43): 27176-27183.

Marshall, N. F. and D. H. Price (1995). "Purification of P-TEFb, a transcription factor required for the transition into productive elongation." *J Biol Chem* **270**(21): 12335-12338.

Matera, A. G., M. Izaguirre-Sierra, et al. (2009). "Nuclear bodies: random aggregates of sticky proteins or crucibles of macromolecular assembly?" *Dev Cell* **17**(5): 639-647.

Matera, A. G. and D. C. Ward (1993). "Nucleoplasmic organization of small nuclear ribonucleoproteins in cultured human cells." *J Cell Biol* **121**(4): 715-727.

McFie, P. J., G. L. Wang, et al. (2006). "Identification of a co-repressor that inhibits the transcriptional and growth-arrest activities of CCAAT/enhancer-binding protein alpha." *J Biol Chem* **281**(26): 18069-18080.

Mermoud, J. E., P. T. Cohen, et al. (1994). "Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism." *EMBO J* **13**(23): 5679-5688.

Mintz, P. J., S. D. Patterson, et al. (1999). "Purification and biochemical characterization of interchromatin granule clusters." *EMBO J* **18**(15): 4308-4320.

Misteli, T. (1999). "RNA splicing: What has phosphorylation got to do with it?" *Curr Biol* **9**(6): R198-200.

Misteli, T. (2001). "The concept of self-organization in cellular architecture." *J Cell Biol* **155**(2): 181-185.

Misteli, T., J. F. Caceres, et al. (1998). "Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo." *J Cell Biol* **143**(2): 297-307.

Misteli, T., J. F. Caceres, et al. (1997). "The dynamics of a pre-mRNA splicing factor in living cells." *Nature* **387**(6632): 523-527.

Misteli, T. and D. L. Spector (1996). "Serine/threonine phosphatase 1 modulates the subnuclear distribution of pre-mRNA splicing factors." *Mol Biol Cell* **7**(10): 1559-1572.

Misteli, T. and D. L. Spector (1999). "RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo." *Mol Cell* **3**(6): 697-705.

Moen, P. T., Jr., C. V. Johnson, et al. (2004). "Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis." *Mol Biol Cell* **15**(1): 197-206.

Molenaar, C., A. Abdulle, et al. (2004). "Poly(A)+ RNAs roam the cell nucleus and pass through speckle domains in transcriptionally active and inactive cells." *J Cell Biol* **165**(2): 191-202.

Montanuy, I., R. Torremocha, et al. (2008). "Promoter influences transcription elongation: TATA-box element mediates the assembly of processive transcription complexes responsive to cyclin-dependent kinase 9." *J Biol Chem* **283**(12): 7368-7378.

Montes, M., A. Cloutier, et al. (2012). "TCERG1 regulates alternative splicing of the Bcl-x gene by modulating the rate of RNA polymerase II transcription." *Mol Cell Biol* **32**(4): 751-762.

Morris, D. P. and A. L. Greenleaf (2000). "The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II." *J Biol Chem* **275**(51): 39935-39943.

Murphy, J. M., D. F. Hansen, et al. (2009). "Structural studies of FF domains of the transcription factor CA150 provide insights into the organization of FF domain tandem arrays." *J Mol Biol* **393**(2): 409-424.

Nakayasu, H. and K. Ueda (1984). "Small nuclear RNA-protein complex anchors on the actin filaments in bovine lymphocyte nuclear matrix." *Cell Struct Funct* **9**(4): 317-325.

Nayler, O., F. Schnorrer, et al. (1998). "The cellular localization of the murine serine/arginine-rich protein kinase CLK2 is regulated by serine 141 autophosphorylation." *J Biol Chem* **273**(51): 34341-34348.

Neubauer, G., A. King, et al. (1998). "Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex." *Nat Genet* **20**(1): 46-50.

Ngo, J. C., S. Chakrabarti, et al. (2005). "Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2." *Mol Cell* **20**(1): 77-89.

Ngo, J. C., K. Giang, et al. (2008). "A sliding docking interaction is essential for sequential and processive phosphorylation of an SR protein by SRPK1." *Mol Cell* **29**(5): 563-576.

Nicolis, G. (1977). "Dissipative structures and biological order." *Adv Biol Med Phys* **16**: 99-113.

Niedojadlo, J., C. Perret-Vivancos, et al. "Transcribed DNA is preferentially located in the perichromatin region of mammalian cell nuclei." *Exp Cell Res* **317**(4): 433-444.

Ninomiya, K., N. Kataoka, et al. "Stress-responsive maturation of Clk1/4 pre-mRNAs promotes phosphorylation of SR splicing factor." *J Cell Biol* **195**(1): 27-40.

O'Keefe, R. T., A. Mayeda, et al. (1994). "Disruption of pre-mRNA splicing in vivo results in reorganization of splicing factors." *J Cell Biol* **124**(3): 249-260.

Patel, N. A., C. E. Chalfant, et al. (2001). "Insulin regulates alternative splicing of protein kinase C beta II through a phosphatidylinositol 3-kinase-dependent pathway involving the nuclear serine/arginine-rich splicing factor, SRp40, in skeletal muscle cells." *J Biol Chem* **276**(25): 22648-22654.

Patel, N. A., S. Kaneko, et al. (2005). "Molecular and genetic studies imply Akt-mediated signaling promotes protein kinase CbetaII alternative splicing via phosphorylation of serine/arginine-rich splicing factor SRp40." *J Biol Chem* **280**(14): 14302-14309.

Pearson, J. L., T. J. Robinson, et al. (2008). "Identification of the cellular targets of the transcription factor TCERG1 reveals a prevalent role in mRNA processing." *J Biol Chem* **283**(12): 7949-7961.

Pederson, T. (2008). "As functional nuclear actin comes into view, is it globular, filamentous, or both?" *J Cell Biol* **180**(6): 1061-1064.

Pestic-Dragovich, L., L. Stojiljkovic, et al. (2000). "A myosin I isoform in the nucleus." *Science* **290**(5490): 337-341.

Peterlin, B. M. and D. H. Price (2006). "Controlling the elongation phase of transcription with P-TEFb." *Mol Cell* **23**(3): 297-305.

Phair, R. D. and T. Misteli (2000). "High mobility of proteins in the mammalian cell nucleus." *Nature* **404**(6778): 604-609.

Politz, J. C., R. A. Tuft, et al. (1999). "Movement of nuclear poly(A) RNA throughout the interchromatin space in living cells." *Curr Biol* **9**(6): 285-291.

Politz, J. C., R. A. Tuft, et al. (2006). "Rapid, diffusional shuttling of poly(A) RNA between nuclear speckles and the nucleoplasm." *Mol Biol Cell* **17**(3): 1239-1249.

Pombo, A. and P. R. Cook (1996). "The localization of sites containing nascent RNA and splicing factors." *Exp Cell Res* **229**(2): 201-203.

Prasanth, K. V., M. Camiolo, et al. (2010). "Nuclear organization and dynamics of 7SK RNA in regulating gene expression." *Mol Biol Cell* **21**(23): 4184-4196.

Prasanth, K. V., S. G. Prasanth, et al. (2005). "Regulating gene expression through RNA nuclear retention." *Cell* **123**(2): 249-263.

Rafalska, I., Z. Zhang, et al. (2004). "The intranuclear localization and function of YT521-B is regulated by tyrosine phosphorylation." *Hum Mol Genet* **13**(15): 1535-1549.

Rappsilber, J., U. Ryder, et al. (2002). "Large-scale proteomic analysis of the human spliceosome." *Genome Res* **12**(8): 1231-1245.

Ren, L., J. R. McLean, et al. (2011). "Systematic two-hybrid and comparative proteomic analyses reveal novel yeast pre-mRNA splicing factors connected to Prp19." *PLoS One* **6**(2): e16719.

Ricciardi, S., C. Kilstrup-Nielsen, et al. (2009). "CDKL5 influences RNA splicing activity by its association to the nuclear speckle molecular machinery." *Hum Mol Genet* **18**(23): 4590-4602.

Sacco-Bubulya, P. and D. L. Spector (2002). "Disassembly of interchromatin granule clusters alters the coordination of transcription and pre-mRNA splicing." *J Cell Biol* **156**(3): 425-436.

Saitoh, N., C. S. Spahr, et al. (2004). "Proteomic analysis of interchromatin granule clusters." *Mol Biol Cell* **15**(8): 3876-3890.

Salichs, E., A. Ledda, et al. (2009). "Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment." *PLoS Genet* **5**(3): e1000397.

Sanchez-Alvarez, M., A. C. Goldstrohm, et al. (2006). "Human transcription elongation factor CA150 localizes to splicing factor-rich nuclear speckles and assembles transcription and splicing components into complexes through its amino and carboxyl regions." *Mol Cell Biol* **26**(13): 4998-5014.

Sanchez-Alvarez, M., M. Montes, et al. (2010). "Differential effects of sumoylation on transcription and alternative splicing by transcription elongation regulator 1 (TCERG1)." *J Biol Chem* **285**(20): 15220-15233.

Sasaki, Y. T., T. Ideue, et al. (2009). "MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles." *Proc Natl Acad Sci U S A* **106**(8): 2525-2530.

Schmidt, U., K. Richter, et al. (2006). "In vivo BiFC analysis of Y14 and NXF1 mRNA export complexes: preferential localization within and around SC35 domains." *J Cell Biol* **172**(3): 373-381.

Schul, W., R. van Driel, et al. (1998). "A subset of poly(A) polymerase is concentrated at sites of RNA synthesis and is associated with domains enriched in splicing factors and poly(A) RNA." *Exp Cell Res* **238**(1): 1-12.

Sharma, A., H. Takata, et al. (2010). "Son is essential for nuclear speckle organization and cell cycle progression." *Mol Biol Cell* **21**(4): 650-663.

Shi, Y. and J. L. Manley (2007). "A complex signaling pathway regulates SRp38 phosphorylation and pre-mRNA splicing in response to heat shock." *Mol Cell* **28**(1): 79-90.

Shopland, L. S., C. V. Johnson, et al. (2003). "Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods." *J Cell Biol* **162**(6): 981-990.

Shopland, L. S., C. V. Johnson, et al. (2002). "Evidence that all SC-35 domains contain mRNAs and that transcripts can be structurally constrained within these domains." *J Struct Biol* **140**(1-3): 131-139.

Shpargel, K. B., J. K. Ospina, et al. (2003). "Control of Cajal body number is mediated by the coilin C-terminus." *J Cell Sci* **116**(Pt 2): 303-312.

Smith, M. J., S. Kulkarni, et al. (2004). "FF domains of CA150 bind transcription and splicing factors through multiple weak interactions." *Mol Cell Biol* **24**(21): 9274-9285.

Snaar, S., K. Wiesmeijer, et al. (2000). "Mutational analysis of fibrillarin and its mobility in living human cells." *J Cell Biol* **151**(3): 653-662.

Spector, D. L., X. D. Fu, et al. (1991). "Associations between distinct pre-mRNA splicing components and the cell nucleus." *EMBO J* **10**(11): 3467-3481.

Spector, D. L. and A. I. Lamond (2011). "Nuclear speckles." *Cold Spring Harb Perspect Biol* **3**(2).

Spector, D. L., W. H. Schrier, et al. (1983). "Immunoelectron microscopic localization of snRNPs." *Biol Cell* **49**(1): 1-10.

Sune, C. and M. A. Garcia-Blanco (1999). "Transcriptional cofactor CA150 regulates RNA polymerase II elongation in a TATA-box-dependent manner." *Mol Cell Biol* **19**(7): 4719-4728.

Sune, C., T. Hayashi, et al. (1997). "CA150, a nuclear protein associated with the RNA polymerase II holoenzyme, is involved in Tat-activated human immunodeficiency virus type 1 transcription." *Mol Cell Biol* **17**(10): 6029-6039.

Sunwoo, H., M. E. Dinger, et al. (2009). "MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles." *Genome Res* **19**(3): 347-359.

Tazi, J., U. Kornstadt, et al. (1993). "Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing." *Nature* **363**(6426): 283-286.

Trinkle-Mulcahy, L., P. Ajuh, et al. (1999). "Nuclear organisation of NIPP1, a regulatory subunit of protein phosphatase 1 that associates with pre-mRNA splicing factors." *J Cell Sci* **112** (Pt 2): 157-168.

Tripathi, V., J. D. Ellis, et al. "The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation." *Mol Cell* **39**(6): 925-938.

Vecerova, J., K. Koberna, et al. (2004). "Formation of nuclear splicing factor compartments is independent of lamins A/C." *Mol Biol Cell* **15**(11): 4904-4910.

Visa, N., F. Puvion-Dutilleul, et al. (1993). "Intranuclear distribution of poly(A) RNA determined by electron microscope in situ hybridization." *Exp Cell Res* **208**(1): 19-34.

Wagner, S., S. Chiosea, et al. (2003). "The spatial targeting and nuclear matrix binding domains of SRm160." *Proc Natl Acad Sci U S A* **100**(6): 3269-3274.

Wang, H. Y., W. Lin, et al. (1998). "SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells." *J Cell Biol* **140**(4): 737-750.

Wei, X., S. Somanathan, et al. (1999). "Three-dimensional visualization of transcription sites and their association with splicing factor-rich nuclear speckles." *J Cell Biol* **146**(3): 543-558.

Xiao, S. H. and J. L. Manley (1997). "Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing." *Genes Dev* **11**(3): 334-344.

Xie, S. Q., S. Martin, et al. (2006). "Splicing speckles are not reservoirs of RNA polymerase II, but contain an inactive form, phosphorylated on serine2 residues of the C-terminal domain." *Mol Biol Cell* **17**(4): 1723-1733.

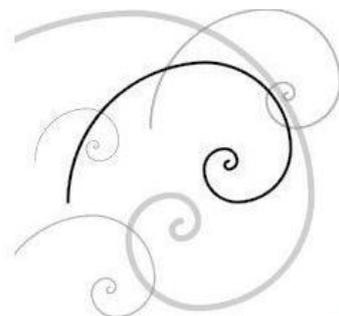
Yuryev, A., M. Patturajan, et al. (1996). "The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins." *Proc Natl Acad Sci U S A* **93**(14): 6975-6980.

Zhang, B., G. Arun, et al. (2012). "The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult." *Cell Rep* **2**(1): 111-123.

Zhong, X. Y., J. H. Ding, et al. (2009). "Regulation of SR protein phosphorylation and alternative splicing by modulating kinetic interactions of SRPK1 with molecular chaperones." Genes Dev **23**(4): 482-495.

Zhou, Z., L. J. Licklider, et al. (2002). "Comprehensive proteomic analysis of the human spliceosome." Nature **419**(6903): 182-185.

# *Publicaciones*





# The FF4 and FF5 Domains of Transcription Elongation Regulator 1 (TCERG1) Target Proteins to the Periphery of Speckles<sup>\*S</sup>

Received for publication, September 16, 2011, and in revised form, March 23, 2012. Published, JBC Papers in Press, March 27, 2012, DOI 10.1074/jbc.M111.304782

Noemí Sánchez-Hernández<sup>‡</sup>, Lidia Ruiz<sup>§</sup>, Miguel Sánchez-Álvarez<sup>‡1</sup>, Marta Montes<sup>‡</sup>, María J. Macías<sup>§¶</sup>, Cristina Hernández-Munain<sup>||</sup>, and Carlos Suñé<sup>‡2</sup>

From the <sup>‡</sup>Department of Molecular Biology and <sup>||</sup>Cell Biology and Immunology, Instituto de Parasitología y Biomedicina “López Neyra” Consejo Superior de Investigaciones Científicas (IPBLN-CSIC), 18100 Armilla, Spain, the <sup>§</sup>Structural and Computational Biology Programme, Institute for Research in Biomedicine (IRB), 08028 Barcelona, Spain, and the <sup>¶</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010, Barcelona, Spain

**Background:** Coordinated transcription and splicing occurs at the periphery of speckles.

**Results:** The FF4 and FF5 domains of transcription elongation regulator 1 (TCERG1) form a structural unit that directs proteins to the periphery of speckles.

**Conclusion:** The FF4 and FF5 domains constitute a novel speckle periphery-targeting signal.

**Significance:** This speckle periphery-targeting signal might participate in the coordination of transcription and splicing.

Transcription elongation regulator 1 (TCERG1) is a human factor implicated in interactions with the spliceosome as a coupler of transcription and splicing. The protein is highly concentrated at the interface between speckles (the compartments enriched in splicing factors) and nearby transcription sites. Here, we identified the FF4 and FF5 domains of TCERG1 as the amino acid sequences required to direct this protein to the periphery of nuclear speckles, where coordinated transcription/RNA processing events occur. Consistent with our localization data, we observed that the FF4 and FF5 pair is required to fold in solution, thus suggesting that the pair forms a functional unit. When added to heterologous proteins, the FF4-FF5 pair is capable of targeting the resulting fusion protein to speckles. This represents, to our knowledge, the first description of a targeting signal for the localization of proteins to sites peripheral to speckled domains. Moreover, this “speckle periphery-targeting signal” contributes to the regulation of alternative splicing decisions of a reporter pre-mRNA *in vivo*.

The mammalian cell nucleus is a highly dynamic organelle that contains numerous morphologically defined structures, some of which have been implicated in essential processes such as RNA biogenesis. Nuclear speckles are one of these nuclear bodies, and studies on their composition, structure, and behavior have provided useful information for understanding the functional compartmentalization of the cell nucleus (1). Speckles are enriched in pre-mRNA splicing factors and are located in the interchromatin region of the nucleoplasm of mammalian cells (2–4). They appear as 20–50 irregular regions per mammalian nucleus that are generally defined by immunofluorescence staining of RNA-processing factors, such as the serine/arginine-rich (SR)<sup>3</sup> splicing factor SC35 (5). Speckles are dynamic structures, and they become round and increase in size upon transcriptional or splicing inhibition (6). It is believed that speckles are storage/assembly sites for splicing components and that transcription and pre-mRNA splicing do not occur within these structures. However, a significant proportion of RNA polymerase II-mediated transcription in the cell nucleus is associated with the periphery of speckles (7–9). Nascent transcript formation near the speckle compartment results in the recruitment of splicing factors from these nuclear bodies to the processing site, and this exchange rate might be regulated by continuous phosphorylation and dephosphorylation events. Thus, phosphorylation of SR proteins is necessary for their recruitment from nuclear speckles to sites of transcription *in vivo* (10). In summary, speckles appear to modulate the relative concentration of processing factors at active transcription sites, thus acting as an architectural integrator of the dynamic molecular associations that are involved in the coordination of transcription and RNA processing.

<sup>\*</sup> This work was supported by Spanish Ministry of Science and Innovation Grants BFU2008-01599 and BFU2011-24577, by Foundation for Research and Prevention of AIDS in Spain Grant FIPSE-36768/08, and by Andalusian Government Excellence Project CVI-4626/2009 (to C. S.). This work was also supported by Spanish Ministry of Science and Innovation Grants BFU2009-08796 (to C. H. M.) and CSL2009-00080 (to M. J. M.), and by the Institute for Research in Biomedicine (to M. J. M.). This work was also supported by the European Region Development Fund, ERDF (FEDER), by a fellowship from the Spanish Ministry of Education (Formación Profesorado Universitario/FPU program) (to M. M.), and by a fellowship from the Consejo Superior de Investigaciones Científicas (Junta para la Ampliación de Estudios, CSIC/JAE program) (to N. S. H.).

<sup>S</sup> This article contains supplemental Figs. S1–S6.

<sup>1</sup> Present address: Dynamical Cell Systems Team, Section of Cellular and Molecular Biology, Institute of Cancer Research, SW3 6JB London, UK.

<sup>2</sup> To whom correspondence should be addressed: Instituto de Parasitología y Biomedicina “López Neyra” (CSIC), PT Ciencias de la Salud, Avenida del Conocimiento s/n, Armilla 18100, Spain. Tel.: 34-958181645; Fax: 34-958181632; E-mail: csune@ipb.csic.es.

<sup>3</sup> The abbreviations used are: SR, serine/arginine-rich; RRM, RNA recognition motif; TCERG1, transcription elongation regulator 1; ECFP, enhanced cyan fluorescent protein; DSC, differential scanning calorimetry; HSQC, heteronuclear single-quantum correlation; SUMO, small ubiquitin-like modifier.

## Novel Speckle Periphery-targeting Signal

Although significant progress has been made on the role of speckles in gene expression, little is known about the sequence motifs responsible for the accumulation of splicing factors at the speckle region. In the case of the SR family of proteins, the RNA recognition motif (RRM) and the RS domain direct these splicing factors to the nuclear speckles (11–14). Other regions of specific splicing factors can also act as targeting signals to nuclear speckles, such as the threonine-proline repeats found in SF3b155 (15) and the arginine-, proline-, and serine-rich domains of SRm160 (16). In the case of protein kinases CrkRS and DYRK1A, the RS domain and a histidine-rich region, respectively, are required for localization to speckles (17–19). To date, no localization signal has been clearly defined to target proteins to the interface between speckles and surrounding transcription sites.

TCERG1 participates in transcriptional elongation and alternative splicing of pre-mRNAs, and a role for this protein in coordinating both processes has been proposed (20, 21). TCERG1 is composed of 1098 residues (22) and contains three WW domains at its N terminus followed by six FF domains at its C terminus. TCERG1 was first described as a transcriptional elongation regulator and was initially found in HIV-1 Tat-responsive HeLa nuclear extract fractions (22, 23). However, accumulating evidence indicates a potential role of TCERG1 in splicing and, hence, in the coupling between transcription and splicing. TCERG1 affects the alternative pre-mRNA splicing of  $\beta$ -globin,  $\beta$ -tropomyosin, CD44, and fibronectin splicing reporters (24–27) and of putative cellular targets identified by microarray analysis following TCERG1 knockdown (26). Consistent with a potential role in the coupling of transcription and splicing, TCERG1 localizes at the interface of splicing factor-rich nuclear speckles and what are presumably nearby transcription sites (21), and it associates with RNA polymerase II and with elongation and splicing components (21, 24, 28, 29).

In this study, we identified the FF4 and FF5 domains of TCERG1 as the region required to direct this protein to the periphery of nuclear speckles. We performed NMR-based analyses and observed that although the FF4 domain is folded and stable, the FF5 domain is not. However, when both domains are expressed as a pair, the folded properties of FF5 are improved. These observations suggest that both domains form a functional unit and provide insights into the nature of FF protein domains. Moreover, our data demonstrate that both of these FF domains specifically direct the localization of fused unrelated proteins to these nuclear regions. Therefore, we defined the FF4 and FF5 domains as novel targeting signals for the localization of proteins at the interface between speckles and what are presumably nearby transcription sites. Placing our data in a functional context, this “speckle periphery-targeting sequence” contributes to the regulation of alternative splicing decisions of a reporter pre-mRNA *in vivo*.

### EXPERIMENTAL PROCEDURES

**Plasmids**—The intermediate vector pEF-ECFP and eukaryotic expression plasmid pEFBOS/ECFP/T7-TCERG1(1–1098) have been described previously (27). pEFBOS/ECFP/T7-TCERG1(1–662) was generated using EcoRI fragments obtained from the pEFBOST7-TCERG1 parental vector (23) by

standard cloning procedures. pEFBOS/ECFP/T7-TCERG1(1–662)-FF mutants were constructed by inserting appropriate FF repeat domain sequences into the BstBI-digested pEFBOS/ECFP/T7-TCERG1(1–662) vector. The enhanced green fluorescent protein expression vectors pEGFP-FF5 and pEGFP-FF4/FF5 were obtained by inserting appropriate PCR-amplified products into the XhoI and XhoI/BamHI sites of pEGFP-C1 (Invitrogen), respectively. pEFBOS/ECFP/T7-TCERG1 $\Delta$ FF4/FF5 was generated by cloning two PCR products encompassing FF1 to FF3 and FF6 sequences, respectively, into the pEFBOS/ECFP/T7-TCERG1 (1–662) plasmid by standard cloning procedures. pEFBOS/GFP/T7-TCERG1 $\Delta$ FF5 was kindly provided by Mariano Garcia-Blanco (Duke University Medical Centre, Durham, NC). The phenylalanine-to-alanine mutants (F903A, F961A, F903A,F961A, F946A, and F903A,F946A) were created by mutagenesis into pEFBOS/ECFP/T7-TCERG1(1–662)-FF4/FF5 vector.

The wild-type and mutant SRSF1 expression vectors (SRSF1, RRM1, and RRM2) were described previously (14) and kindly provided by Javier F. Cáceres (MRC, Edinburgh). RRM1 and RRM2 fragments with XbaI and BamHI/XhoI sites were obtained by PCR and inserted into the same epitope-tagged expression vector as wild-type SRSF1 (pCGT7-SF2), yielding pCGT7-RRM1 and pCGT7RRM2 intermediate vectors. The RRM mutants containing FF repeat domains were created using appropriate XhoI/BglII fragments obtained by PCR from pEFBOS-ECFP-T7-TCERG1 (1–1098) and inserted into XhoI/BamHI-digested pCGT7-RRM vectors.

All constructs were verified by DNA sequencing. Plasmids were transformed into DH5 $\alpha$  cells for selection.

**Proteins**—Constructs of the FF4 (878–956 fragment) and for the pair of FF4–FF5 domains (878–1022 fragment) of TCERG1 were amplified by PCR and subcloned into a pETM-11 (a gift from the European Molecular Biology Laboratory-Heidelberg Protein Expression Facility) using NcoI and HindIII sites. All constructs were confirmed by sequencing.

Unlabeled,  $^{15}\text{N}$ -labeled,  $^{13}\text{C}$ ,  $^{15}\text{N}$  and  $^2\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$ -labeled proteins were expressed in *Escherichia coli* BL21 (DE3) in Luria Broth medium or minimal medium (M9) using either  $\text{H}_2\text{O}$  or  $\text{D}_2\text{O}$  (99.89%, CortecNet) enriched with  $^{15}\text{NH}_4\text{Cl}$  and/or D- $^{13}\text{C}$  glucose as the sole sources of carbon and nitrogen, respectively (30). *E. coli* extracts were lysed using an EmulsiFlex-C5 (Avestin) cell disrupter equipped with an in-house developed Peltier temperature controller system. Soluble fusion proteins were purified by nickel affinity chromatography (HiTrap chelating High Performance column, GE Healthcare), and samples were eluted using buffer (20 mM Tris, 10 mM Imidazole, 150 mM NaCl) with EDTA. After nickel affinity purification, the proteins were cut with the Tobacco Etch Virus (TEV) protease and further purified by gel filtration on a HiLoad<sup>TM</sup> Superdex<sup>TM</sup> 75 prepgrade (GE Healthcare). All samples were prepared in 20 mM sodium phosphate buffer, 130 mM NaCl, 0.5 mM  $\text{NaN}_3$  in 90%  $\text{H}_2\text{O}$ , 10%  $\text{D}_2\text{O}$  or 100%  $\text{D}_2\text{O}$  (pH 5.8). To avoid aggregation of the FF4–FF5 sample, the buffer was supplemented with 5% glycerol.

**Antibodies**—Antibodies against the T7 tag were purchased from Bethyl and used at dilutions of 1:20,000 and 1:1000 for immunoblotting and immunofluorescence, respectively. Anti-

body against CDK9 (catalog no. sc-484, Santa Cruz Biotechnology) was used at a dilution of 1:500. Antibodies against U2AF<sup>65</sup> were kindly provided by J. Valcárcel (CRG, Barcelona, Spain) and were used at a dilution of 1:500. Antibody against splicing factor SC35 (catalog no. S4045, Sigma) was used at a dilution of 1:4000. For Western blot analysis, primary antibodies were detected using HRP-conjugated secondary antibodies to rabbit and mouse (PerkinElmer Life Sciences). These were generally used at dilutions of 1:5000. For immunofluorescence, we used Alexa Fluor 488-conjugated goat anti-mouse and Alexa Fluor 647-conjugated goat anti-rabbit from Molecular Probes and were generally used at dilutions of 1:500.

**Cell Culture and Transfection Assays**—HeLa and HEK293T cells were grown and maintained as described previously (27). Transfection assays were carried out using protocols described previously (21, 27). Cells were transfected by using calcium phosphate and/or Lipofectamine 2000 reagent (Invitrogen) according to the protocols of the manufacturer. Empty vector was used to keep the total amount of nucleic acid constant.

**Immunofluorescence, Image Processing, and Quantification**—HeLa and HEK293T cells were grown on coverslips and transfected with the constructs indicated in the legends to the figures. Approximately 24 h after transfection (50–60% confluence), cells were fixed with 3.5% paraformaldehyde in PBS buffer (pH 7.4) for 45 min on ice. Cells were washed three times with PBS, permeabilized in PBS containing 0.5% Triton X-100 for 5 min at room temperature, and blocked in PBS containing 2.5% BSA overnight at 4 °C. Cells were incubated with primary antibodies at appropriate dilutions in PBS containing 0.1% BSA for 1 h at room temperature (humidity chamber) and subsequently washed extensively with 0.1% BSA in PBS and incubated with appropriate secondary antibodies under the conditions described previously. After staining, cells were rinsed five times with 0.1% BSA in PBS and three more times with PBS. Coverslips were then mounted onto glass slides using ProLong Gold antifade reagent (Molecular Probes). Images were acquired with an inverted Leica SP2 confocal microscope, using an HCX PL APO CS 40.0 × 1.25 OIL UV objective. In cases where double immunofluorescence was performed, images were all taken simultaneously. GFP, CFP, and Alexa Fluor 488 were excited with the 488-nm line of the argon laser, whereas Alexa Fluor 647 was excited with a 633-nm HeNe laser. The pinhole diameter was kept at 1 μm. Quantification analysis shown in Fig. 1B was performed by measuring and comparing the average pixel intensity of the ECFP at the nuclear speckle site and adjacent nucleoplasm of a region of interest scan for each TCERG1 mutant. Thirty-six regions of interest from three different experiments with 12 cells per experiment (72 and 108 measurements for the speckles and nucleoplasm, respectively) were quantified for each TCERG1 mutant. Statistical analysis was performed using a standard Student's *t* test. Acquisition software was LAS AF v2.3.6 Build 5381sps. All images were digitally processed for presentation using Adobe Photoshop CS3 extended v10.0 software.

**NMR Spectroscopy**—All experiments were recorded on a Bruker Avance III 600-MHz spectrometer equipped with a z pulse field gradient unit and a triple (<sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N) resonance probe head. Double- and/or triple-labeled samples were pre-

pared to obtain sequence-specific (HNCACB/HN(CO)CACB or CBCA(CO)NH/CBCANH) experiments. All spectra were processed with the NMRPipe/NMRDraw (31) software and were analyzed with Computer Aided Resonance Assignment, CARA (32). The <sup>15</sup>N-relaxation experiments were acquired for <sup>15</sup>N-labeled samples of FF4+FF5 (0.5 mM) and FF4 (0.5 mM) essentially as described (33). Heteronuclear {<sup>1</sup>H}-<sup>15</sup>N NOE experiment was performed using standard 2D experiments, with the reference and proton saturated spectra collected in an interleaved fashion. The values of steady-state <sup>1</sup>H-<sup>15</sup>N NOEs were determined from the ratios of the peak intensities measured in spectra recorded either with (*I*<sub>s</sub>) or without (*I*<sub>o</sub>) presaturation during the relaxation delay as described (34). The standard deviation of the NOE was determined on the basis of measured background noise levels using the following relationship:

$$\sigma_{\text{NOE/NOE}} = ((\sigma_{I_{\text{sat}}/I_{\text{sat}}})^2 + (\sigma_{I_{\text{unsat}}/I_{\text{unsat}}})^2)^{1/2} \quad (\text{Eq. 1})$$

T1 and T2 experiments were acquired with 135 (t1) × 2048 (t2) total real points. T1 data points were obtained with 12 different relaxation periods: 20.8, 52, 104, 156, 265, 424, 520, 676, 832, 1040, 1352, and 1664 ms. Ten delay times were also sampled in the T2 experiments: 12, 24, 36, 48, 60, 72, 84, 120, 144, and 168 ms. T<sub>1</sub> and T<sub>2</sub> values were determined by fitting the measured peak heights to a two-parameter function of the form:

$$I(t) = I_0 \exp(-t/T_{1,2}) \quad (\text{Eq. 2})$$

where *I*(*t*) is the intensity after a delay of time *t* and *I*<sub>0</sub> is the intensity at time *t* = 0. Minimization performed using the Levenberg-Marquardt optimization algorithm was used to determine the optimum value of the *I*<sub>0</sub> and T<sub>1,2</sub> parameters by minimizing the χ<sup>2</sup> goodness of fit parameter:

$$\chi^2 = \sum (I_c(t) - I_e(t))^2 / s_i^2 \quad (\text{Eq. 3})$$

where *I*<sub>c</sub>(*t*) are the intensities calculated from the fitting parameters, *I*<sub>e</sub>(*t*) are the experimental intensities, *s*<sub>*i*</sub> is the standard deviation of the experimental intensity measurements, and summation is performed over the number of time points recorded in each experiment.

**Differential Scanning Calorimetry**—Thermal denaturation was studied using differential scanning calorimetry (DSC). The experiments were performed using a VP<sup>TM</sup> DSC MicroCalorimeter (in the Polymorphism and Calorimetry Platform of the Scientific and Technical Services (SCT), Universitat de Barcelona) in 20 mM sodium phosphate buffer, 130 mM NaCl, 0.5 mM NaN<sub>3</sub> (pH 5.8). The protein solution was heated up at a constant rate of 1 °C/min from 10 to 80° and a constant pressure. The temperature dependence of the excess heat capacity was analyzed and plotted with Origin 7.0 software (OriginLab Corp.).

**RT-PCR Analysis**—Approximately 2 × 10<sup>6</sup> HEK293T cells were seeded 24 h before transfection. Cells were transfected with 0.6 μg of Bcl-X2 minigene reporter (kindly provided by Benoit Chabot, University of Sherbrooke) and 1 μg of the

## Novel Speckle Periphery-targeting Signal

TCERG1 derivative plasmids as indicated in the legend of Fig. 6. Total RNA was isolated from transfected cells by using the TRIzol reagent (Invitrogen). Approximately 1  $\mu$ g of RNA was digested with 10 units of RNase-Free DNase (Roche). One-half of digested RNA was used for RT using the RT-Sveda primer and Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase (Invitrogen) according to the instructions of the manufacturer. 10% of the RT reaction was used as template together with X34 and X-Agel-R primers. Reaction products were analyzed on 2% agarose gel and quantified using Quantity One 4.5.0 software. The following primers were used: 5'-GGGAAGCTA-GAGTAAGTAG-3' (RT-Sveda1-Rev), 5'-AGGGAGGCAGG-CGACGAGTTT-3' (X34), and 5'-GTGGATCCCCGGGC-TGCAGGAATTCGAT-3' (X-Agel-R).

### RESULTS

*Identification of a Sequence Element Required for the Localization of TCERG1 to Nuclear Speckles*—We reported previously that TCERG1 is present along the periphery of the speckles (21) but that the sequences responsible for this location have not yet been defined. In our previous study, we found that a TCERG1 protein lacking the FF5 and FF6 repeat motifs is not able to localize to speckles (21). At least three interpretations are possible: 1) FF5 is essential for the subnuclear localization of TCERG1; 2) FF6 and FF5 are both important but neither is essential; or 3) a minimum number of FF repeats are required for the localization to nuclear speckles. To identify the elements within the FF repeat motifs of TCERG1 responsible for its accumulation in nuclear speckles, we constructed a series of mutants of human TCERG1 fused to ECFP and investigated their subnuclear localization in transfected HEK293T cells by confocal laser scanning microscopy. All mutants retain the putative nuclear localization signal found in the middle of the protein (22). The localization of full-length ECFP-tagged TCERG1 was similar to that of the endogenous TCERG1 (Fig. 1A, ECFP and TCERG1[1–1098]). A TCERG1 deletion that completely eliminated the FF domains was localized in a diffuse pattern throughout the nucleoplasm without any evident accumulation on the speckles (Fig. 1A, ECFP and TCERG1[1–662]), in agreement with our data published previously (21). A mutant containing the amino terminal region and the FF5 domain of TCERG1 showed a slight enrichment in speckles, indicating that this domain might be required for targeting the protein to nuclear speckles but may not be sufficient for efficient speckle localization (Fig. 1A, ECFP, compare the signal obtained with TCERG1[1–1098] to the one obtained with TCERG1[1–662]-FF5). Addition of the FF6 domain to this mutant created a fusion protein with a staining pattern similar to the previous mutant, and it did not accumulate in nuclear speckles in significant amounts (supplemental Fig. S1A, TCERG1[1–662]-FF5/FF6). To further analyze the speckle-targeting capacity of the FF domains, we generated a chimera with the amino-terminal region fused to the FF4 and FF5 domains. Remarkably, cells expressing this chimera showed strong nuclear speckle fluorescence that was indistinguishable from that of wild-type TCERG1 (Fig. 1A, ECFP and TCERG1[1–662]-FF4/FF5). Mutant ECFP-TCERG1[1–662]-FF4, lacking FF5, was dispersed throughout the nucleoplasm (supplemental Fig. S1A).

These results suggest that the combination of FF4 and FF5 is the sequence element necessary and sufficient for localization to speckles. This was not due to differential protein expression levels, as confirmed by Western blotting (supplemental Fig. S1B).

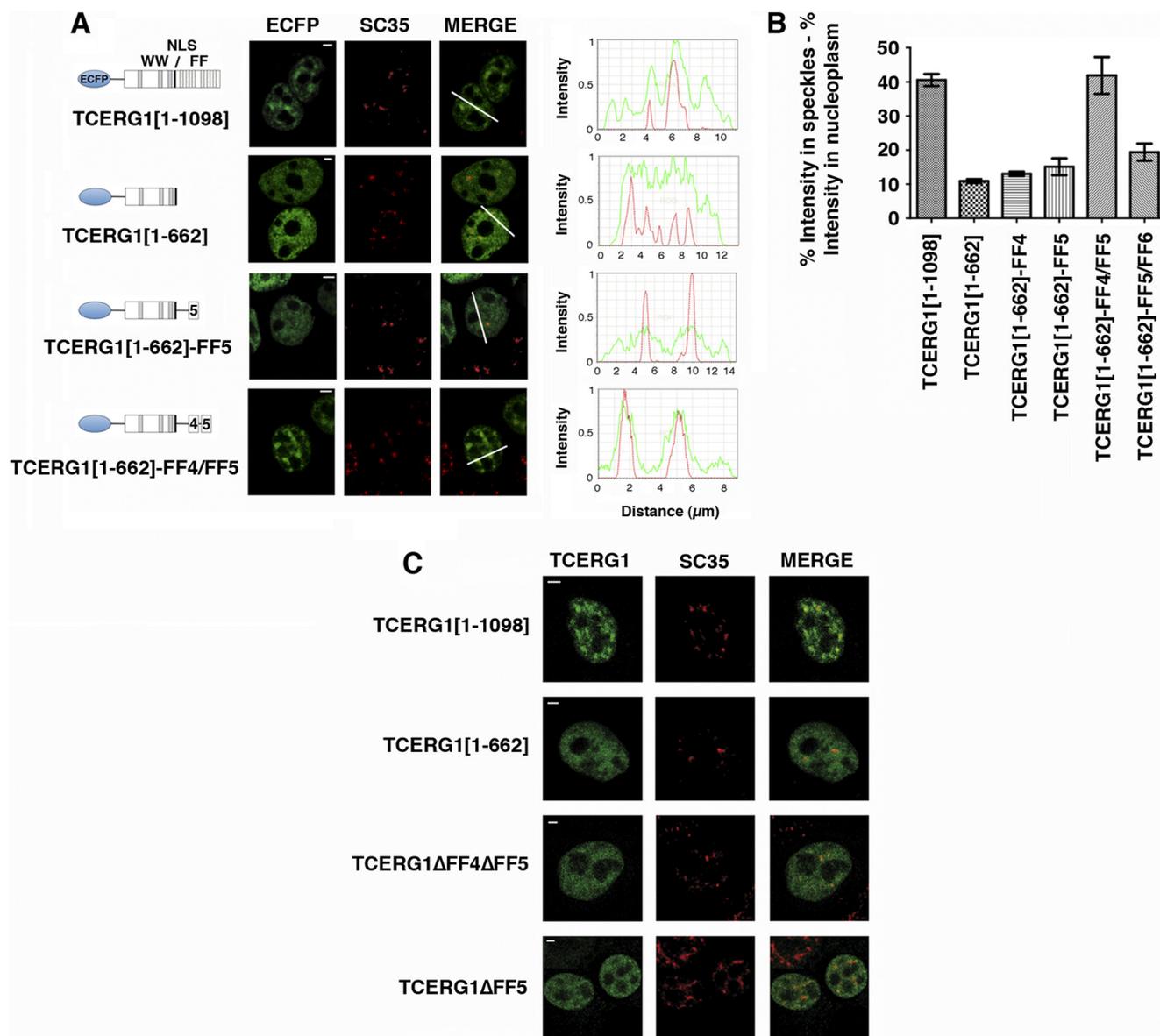
To confirm that the nuclear staining pattern of the transiently expressed proteins coincides with nuclear speckles, we carried out an immunofluorescence analysis with an antibody against the essential splicing factor SC35, which commonly serves to define nuclear speckles. Expression of wild-type and the FF4/FF5 mutant resulted in nuclear colocalization with speckles, whereas the FF5 mutant displayed a partially overlapping signal at the speckle region (Fig. 1A). The spatial relationship between wild-type and mutant TCERG1 variants relative to SC35 was identified by quantitatively scanning specific nuclear regions containing speckles and is shown on the right in Fig. 1A.

To further analyze the spatial distribution of TCERG1, we analyzed the spatial relationship between the relative spatial distributions of either wild-type or TCERG1[1–662]-FF4/FF5 and SC35 in individual nuclear speckles by confocal microscopy optical sectioning (0.3- $\mu$ m sections). We found that the speckles had peripheral and internally located the TCERG1 proteins that are excluded from the core region of the speckles (supplemental Fig. S2). The analysis of those images showed that the TCERG1 peaks partially overlap, but do not coincide with, the SC35 peaks (supplemental Fig. S2), lending additional support to the observation that wild-type (21) and TCERG1[1–662]-FF4/FF5 are enriched at the speckle periphery.

We performed quantitative and statistical analyses that involved computing the proportion of signal intensity contained in speckles (see “Experimental Procedures”). Intensity measurements agreed with the above data and confirmed that FF4/FF5 is required for the efficient targeting of TCERG1 to the speckle compartment (Fig. 1B).

To further corroborate that the FF4 and FF5 domains are the targeting signal for the localization of TCERG1 to nuclear speckles, we expressed wild-type TCERG1 and FF4/FF5-deleted protein tagged with ECFP at the amino terminus and examined their nuclear localization in HEK293T cells using immunofluorescence microscopy. As shown in Fig. 1C, the TCERG1 protein with a deletion encompassing the FF4 and FF5 domains (TCERG1 $\Delta$ FF4 $\Delta$ FF5) exhibited diffuse localization throughout the nucleoplasm without any evident accumulation on the speckles, thus supporting the conclusion that this region of TCERG1 contains the sequence required for proper localization to speckles.

Previously, we have shown that a deletion of the FF6 domain does not affect the speckle distribution of TCERG1, which is consistent with our current data, but that a further deletion that included the FF5 domain results in perturbation of TCERG1 localization to speckles (21). To test whether the FF5 domain is essential for the proper localization of TCERG1 to speckles, we transiently transfected cells with a plasmid encoding a TCERG1 protein in which the FF5 domain had been deleted (TCERG1 $\Delta$ FF5) and examined its nuclear distribution. The expressed protein was present diffusely throughout the nucleoplasm (Fig. 1C). These results indicate that a deletion of the



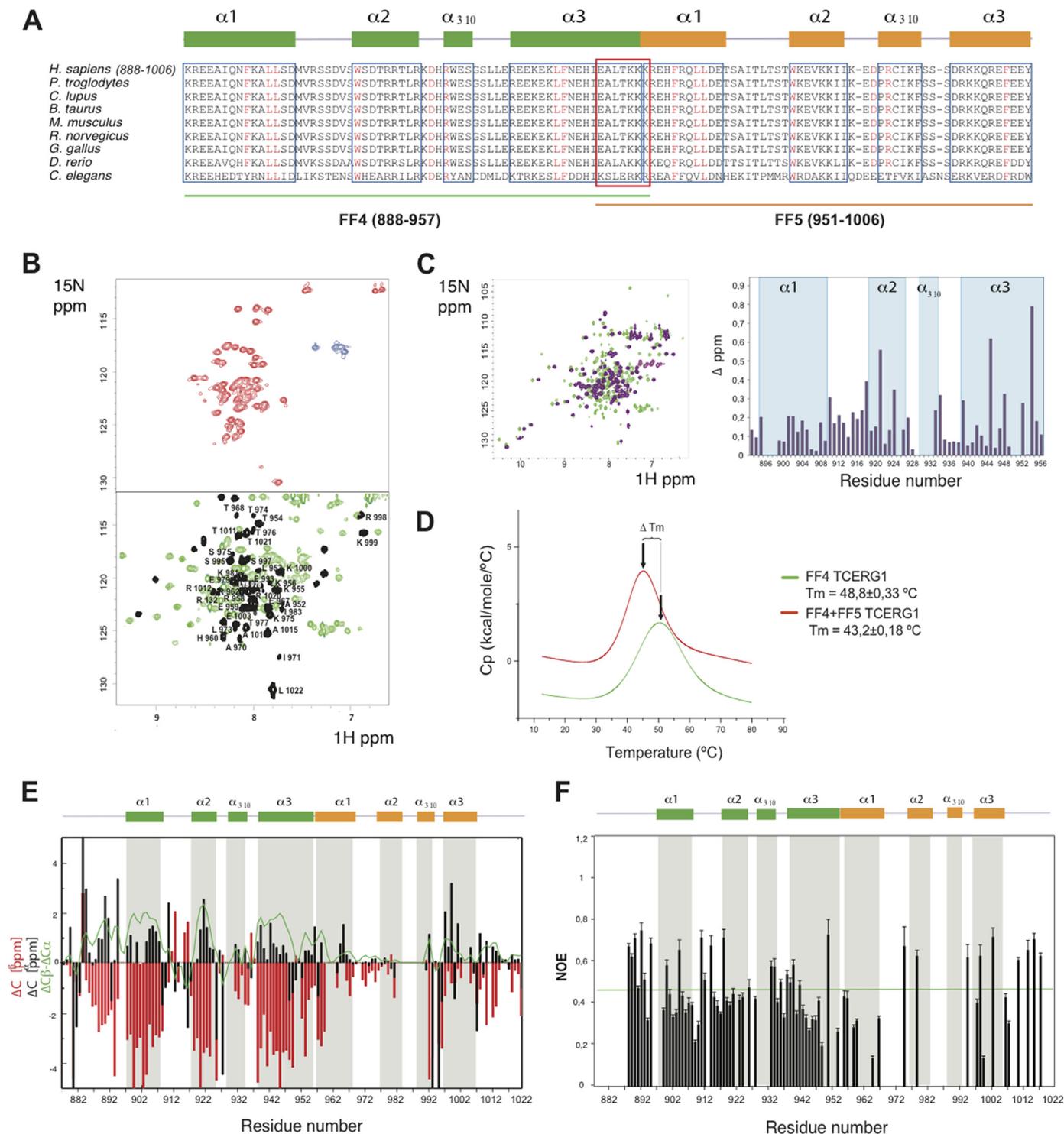
**FIGURE 1. The FF4/FF5 domains are required for efficient targeting of TCERG1 to the speckle compartment.** *A*, dual-labeling of cells transfected with ECFP-TCERG1[1-1098] or the indicated mutant constructs (ECFP, green) with the anti-SC35 antibody (SC35, red) was performed. A diagrammatic representation of the ECFP-TCERG1 fusion proteins is shown to the left of each panel. The numbers in parentheses represent the TCERG1 amino acids contained in the construct. Shown are the three WW domains, the putative nuclear localization signal (NLS), and the six FF domains. Line scans showing local intensity distributions of TCERG1 and SC35 are shown to the right of the panels. Bars in the merged panels indicate the positions of the line scans. Scale bars, 3  $\mu\text{m}$ . *B*, the nucleoplasmic fraction of TCERG1 proteins present in speckles was determined after measuring the fluorescence intensity in speckles relative to nucleoplasm as described under "Experimental Procedures." The data shown are from three independent experiments. Student's *t* test was performed, and differences were shown to be highly significant:  $p(\text{TCERG1}[1-1098] \text{ versus } \text{TCERG1}[1-662]) < 0.0001$ ;  $p(\text{TCERG1}[1-1098] \text{ versus } \text{TCERG1}[1-662]-\text{FF4}) = 0.0001$ ;  $p(\text{TCERG1}[1-1098] \text{ versus } \text{TCERG1}[1-662]-\text{FF5}) = 0.0011$ ;  $p(\text{TCERG1}[1-1098] \text{ versus } \text{TCERG1}[1-662]-\text{FF4}/\text{FF5}) = 0.8261$ ;  $p(\text{TCERG1}[1-1098] \text{ versus } \text{TCERG1}[1-662]-\text{FF5}/\text{FF6}) = 0.0023$ ;  $p(\text{TCERG1}[1-662]-\text{FF4}/\text{FF5} \text{ versus } \text{TCERG1}[1-662]-\text{FF4}) = 0.0060$ ;  $p(\text{TCERG1}[1-662]-\text{FF4}/\text{FF5} \text{ versus } \text{TCERG1}[1-662]-\text{FF5}) = 0.0106$ ; and  $p(\text{TCERG1}[1-662]-\text{FF4}/\text{FF5} \text{ versus } \text{TCERG1}[1-662]-\text{FF5}/\text{FF6}) = 0.0192$ . The bar graph represents mean  $\pm$  S.D. *C*, immunofluorescence analysis of cells transfected with the indicated plasmids. Images show a speckled pattern for full-length TCERG1[1-1098], but TCERG1[1-662], TCERG1[ $\Delta$ FF4/FF5], and TCERG1[ $\Delta$ FF5] are more diffusely dispersed throughout the nucleoplasm. Individual and merged images of the cell labeled with the indicated fluorescent proteins (FP, in green) and with the anti-SC35 antibody (SC35, red) are shown. Scale bars = 3  $\mu\text{m}$ .

FF5 domain disrupts the localization of TCERG1 to nuclear speckles and imply that FF5 is essential for this association.

**NMR Studies of the FF4 and FF5 Domains of TCERG1**—The sequence comparison of the FF4-FF5 pair of domains in different species revealed a high level of sequence conservation in this pair of domains (Fig. 2A). Remarkably, secondary structure predictions of the FF4 and FF5 domains suggested that seven residues (951–957) are shared by these two FF domains, with a fragment of the last helix of the FF4 overlapping with a part of

the first helix of the FF5 domain. This prediction could imply that the FF4 and FF5 domains may not fold independently and might require the presence of the pair to acquire a stable fold. To evaluate the presence or absence of tertiary structure in the pair and to compare it with that of the independent domains, we used recombinant fragments corresponding to the FF4 and FF5 domains as well as a construct containing the FF4/FF5 pair and acquired heteronuclear single-quantum coherence (HSQC) NMR spectra at 298 K. These experiments provide

## Novel Speckle Periphery-targeting Signal



**FIGURE 2.  $^{15}\text{N}$ -HSQC NMR spectra and thermal unfolding of FF4-FF5.** *A*, amino acid alignment of TCERG1 homologues. Secondary structure elements of FF4 and FF5 are drawn above the sequences in *green* and *orange*, respectively. Boxed sequences in *blue* are the residues that form the  $\alpha$ -helices. The overlapped residues between FF4 and FF5 are shown in a *red* box. Strictly conserved residues are highlighted in *red*. *B*,  $^{15}\text{N}$ -HSQC spectra at 298 K for FF5 and FF4-FF5 constructs. The NMR data show the improvement of FF5 backbone amino acid dispersion (shown in *black*) observed in the pair when compared with that of the isolated construct. *C*, overlapped  $^{15}\text{N}$ -HSQC spectra at 298 K for FF4 and FF4-FF5 constructs and chemical shift differences bar representation of FF4 residues. *D*, thermal unfolding of the FF4 and FF4-FF5 domains as monitored by DSC. The difference in  $T_m$  is shown as an *arrow* connecting both maxima. *E*, secondary structure elements of the FF4-FF5 pair. Chemical shift distribution of  $C_\alpha$  and  $C_\beta$ . The ratio of  $C_\alpha$  and  $C_\beta$  is shown as a *green* line. The majorities of the values are positive, indicating the presence of a helical structure. *F*, heteronuclear  $\{^1\text{H}\}$ - $^{15}\text{N}$  NOE. Unassigned residues, proline residues that lack a proton amide, and overlapped peaks were excluded from the analysis. Heteronuclear NOE values show that the pair of FF4-FF5 has secondary structure throughout the construct.

information about the chemical shift dispersion of every amide proton in the sequence. Well dispersed signals are interpreted as an indication of folded samples, whereas overlapped signals

usually correspond to partially folded samples. Under the experimental conditions used in this work, we observed that the independently expressed FF4 domain adopted a canonical FF

fold with an  $\alpha 1\text{-}\alpha 2\text{-}3_{10}\text{-}\alpha 3$  topology (35, 36), whereas the FF5 domain presented properties of a sample containing mixtures of partially folded and unfolded molecules. A superimposition of the HSQC spectra of the independent domains to that of the FF4/FF5 pair revealed chemical shift differences in several residues of the FF4 domain and also that the FF5 domain substantially improved its chemical shift dispersion when compared with that of the isolated FF5 domain (Fig. 2, B and C). These differences seem to indicate that in the FF4-FF5 pair, the domains may contact one another. To support this hypothesis further, we analyzed the thermodynamic behavior of the FF4/FF5 pair and compared it to that of the independent FF4 using DSC. DSC measures the enthalpy of unfolding processes because of thermal denaturation and also provides the thermal transition midpoint ( $T_m$ ), which correlates with protein stability. In agreement with the NMR data, the FF4/FF5 pair presented a unique unfolding curve, suggesting that the pair unfolds in a concerted manner (Fig. 2D). The observed decrease in the midpoint thermal transition of the pair with respect to that of the FF4 domain (5.6 °C) is attributed to an aggregation process occurring in the pair that is absent in the isolated FF4 domain. The aggregation process, which affects the overall stability of the pair, probably involves the FF5 domain because of its lower stability compared with that of FF4.

To identify the elements of secondary structure present in the FF4/FF5 pair, we used a ( $^2\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$ ) recombinant sample and acquired backbone triple-resonance experiments. The assignment was achieved up to a 70% of all possible residues. A comparison of the  $\alpha$  and  $\beta$  carbon chemical shifts of the assigned residues to that of random coil values revealed positive values that correlate with the presence of  $\alpha$  helices, characteristic of FF domains, shown in Fig. 2E as a *green line*. The high number of overlapped residues in the FF5 domain precluded a proper identification of all secondary structure elements (37). However, the positive  $^{13}\text{C}$  values obtained for the unambiguously assigned regions of the FF5 domain suggested that this domain also presents helical secondary structure. From the obtained data it seems that the first helix of FF5 domain is a continuation of the last helix of the FF4 domain, that the  $3_{10}$  and third helices are shorter than expected on the basis of a sequence comparison with other described FF domains, and finally, that the second helix is almost undetectable.

To further examine the structural and dynamical properties of the pair, we performed heteronuclear ( $^{34}$ )  $\{^1\text{H}\}\text{-}^{15}\text{N}$  NOE relaxation experiments. Internal dynamics are directly related to folding, with less ordered proteins displaying higher internal motions. The data obtained for the assigned residues (shown in Fig. 2F) revealed that the distribution of NOE values was, on average, close for both domains in the pair. To further explore the dynamic behavior of the FF4/FF5 pair, we also performed T1 and T2 experiments and compared the obtained values to the independent FF4 domain. The correlation times obtained for the assigned residues (supplemental Fig. S3, A and B), T1, T2, and the T1/T2 ratio are predominantly in the same range, with fewer uniform data for the pair with respect to that of FF4. We attributed the observed results to the slightly higher flexibility of FF5 when compared with FF4, probably because of the presence of long loops connecting the helices in FF5, whereas

the values corresponding to the secondary structured regions were more similar.

These results point out that the fourth and fifth FF domains of TCERG1 have similar properties when expressed together, suggesting that the FF5 domain is more stable in the pair. This organization is a feature specific to these two domains, as we did not observe any structural organization when FF5 was combined with FF6 (supplemental Fig. S3C).

*FF4 and FF5 of TCERG1 Comprise a Novel Nuclear Localization Signal Targeting Proteins to the Periphery of Speckles*—The SR protein SRSF1 (formerly SF2/ASF) is an essential splicing factor that regulates alternative splicing of many pre-mRNAs and is located at the speckle region (38). SRSF1 contains two functional modules: an RS (arginine/serine-rich) domain and two RRM. Previous work demonstrated that at least two of these domains are necessary for SRSF1 localization to nuclear speckles (14). The same work described mutant proteins carrying individual domains (RRM1 and RRM2) that localized throughout the cell without any evident accumulation on the nuclear speckles. We sought to test whether the FF4/FF5 sequence could target those SRSF1 mutants to nuclear speckles. To this end, we ligated the FF4/FF5 and FF5 sequences to the C terminus of the RRM1 and RRM2 gene coding sequences, and the fusion genes were transfected into HeLa cells. All constructs encoded proteins with a bacteriophage T7 epitope tag at their amino terminus, allowing detection of the exogenous proteins with antibodies that recognize this epitope. Immunofluorescence experiments using a T7 tag antibody showed that the transiently expressed wild-type SRSF1 protein localized exclusively in the nucleus with a typical speckled pattern (Fig. 3a). Double-labeling experiments with anti-SC35 antibodies confirmed SRSF1 localization at nuclear speckles (Fig. 3, b and c). In contrast, when individual domains were expressed (RRM1 or RRM2), the mutant protein localized throughout the cell, and colocalization with nuclear speckles was not detected (Fig. 3, d–f and m–o), which agrees with a previous report (14). Small changes in the staining pattern were observed upon addition of the FF5 region to the mutant proteins (Fig. 3, panels g–i and p–r). However, mutant proteins containing the FF4/FF5 region clearly led to speckle localization of the fusion protein (Fig. 3, j–l and s–u). Expression of the various SRSF1 constructs was nearly identical, as assessed by Western blotting (supplemental Fig. S4).

Next, we generated GFP fusion proteins containing the FF4/FF5 domains and the FF5 domain alone. The GFP-FF5 protein showed a diffuse staining pattern similar to that of the GFP alone with small punctuate areas in the nucleus (18) (supplemental Fig. S5). The GFP-FF4/FF5 protein clearly accumulated in the speckle compartment (supplemental Fig. S5), indicating that the presence of this sequence is sufficient for nuclear speckle localization of the GFP. These results, together with the analysis of the SRSF1 mutants, indicate that the FF4/FF5 region shown to be necessary for the localization of TCERG1 protein to speckles is sufficient to direct the localization of heterologous proteins to nuclear speckles.

To explore further the association of FF4/FF5-containing RRMs with speckles, we treated cells with the RNA polymerase II inhibitor  $\alpha$ -amanitin. Upon RNA polymerase II inhibition,

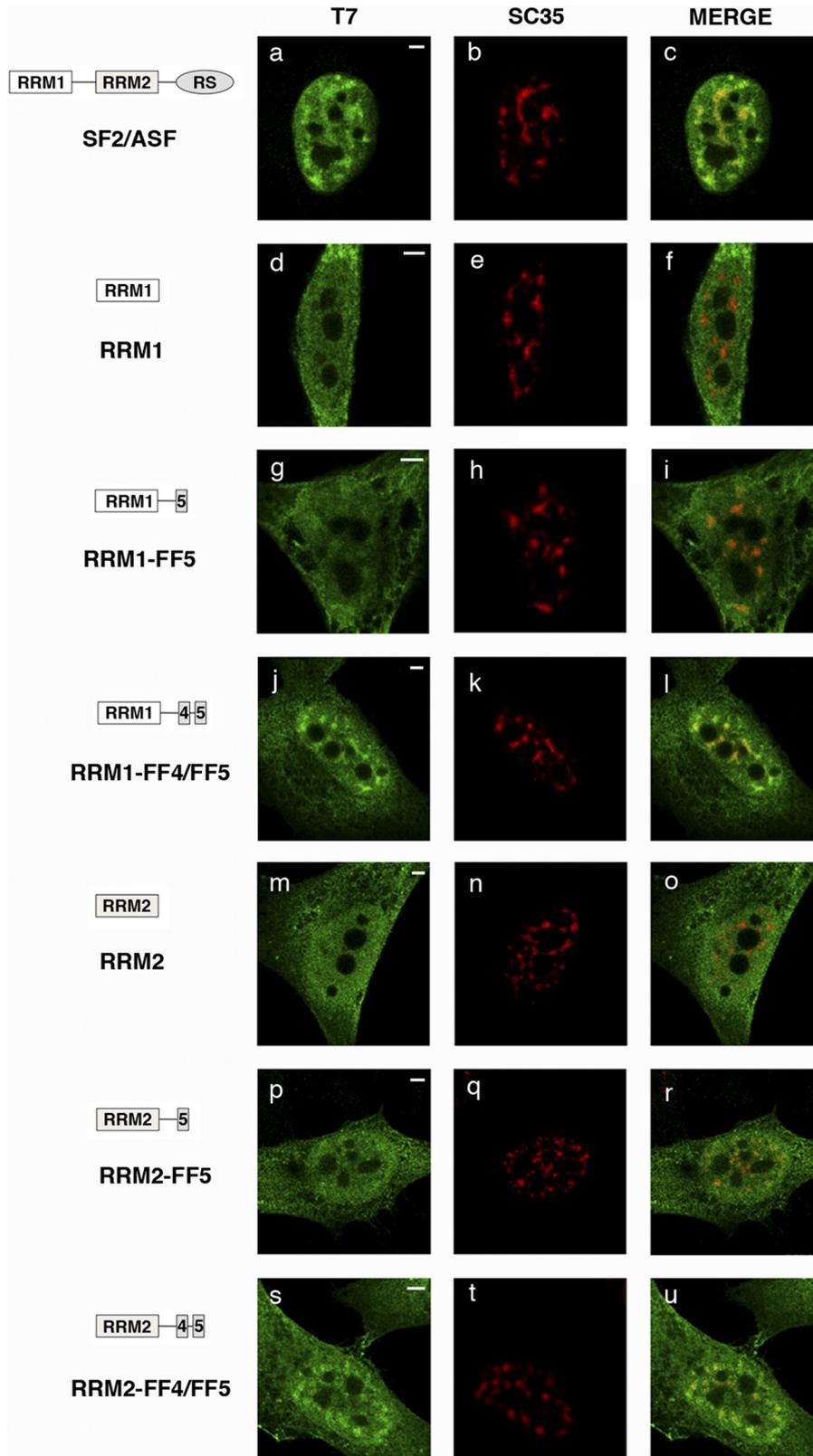


FIGURE 3. **FF4/FF5 directs SRSF1 domain-deletion mutants to nuclear speckles.** Cells were transfected with the indicated plasmids and dually labeled with antibodies directed against the expressed SRSF1 protein (*left column, green*) and SC35 (*center column, red*). The merged images are also shown (*right column*). In all cases, colocalization of expressed proteins with the endogenous marker was assessed by confocal imaging. A diagrammatic representation of the T7-tagged SRSF1 mutants used is shown at the *left* of the figure. The structure of the SRSF1 domain-deletion mutants was described previously (51). Scale bars = 3  $\mu$ m.

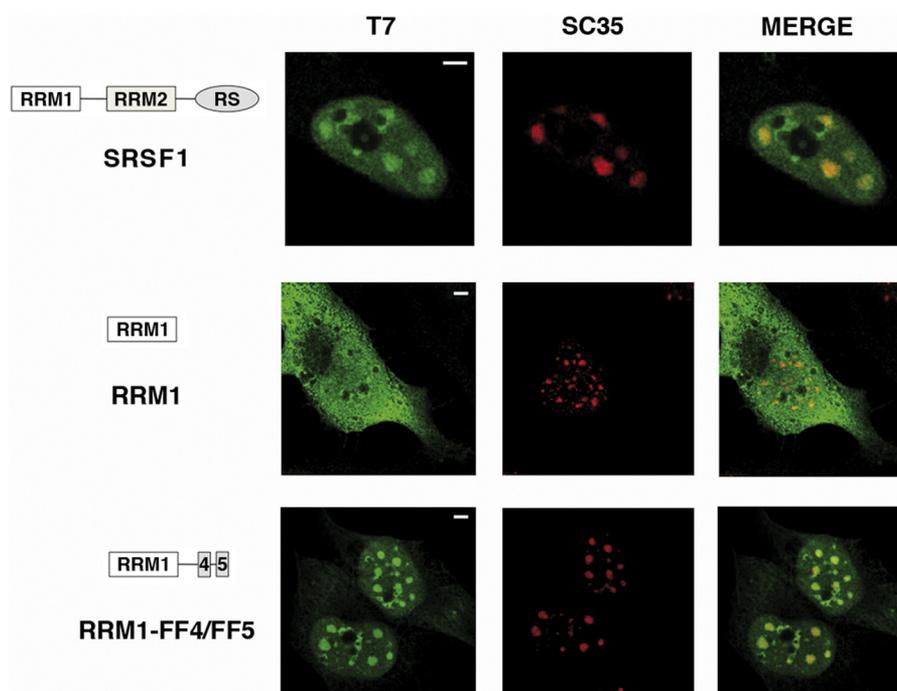


FIGURE 4. **Colocalization of the FF4/FF5-containing RRM1 protein with nuclear speckles is not perturbed following inhibition of transcription.** HeLa cells were transfected with the indicated expression plasmids and treated with 25  $\mu\text{g/ml}$  of  $\alpha$ -amanitin for 6 h at 37  $^{\circ}\text{C}$  and then processed for immunofluorescence analysis. Dual-labeling of cells with antibodies directed against SRSF1 (T7, green) and with the SC35 antibody (red) was performed. Individual staining and merge images of the cell stained with the indicated antibodies are shown. A diagrammatic representation of the T7-tagged SRSF1 mutants used is shown at the left of the figure. Scale bars = 3  $\mu\text{m}$ .

speckles decrease in number, enlarge, and become rounded because of the accumulation of the splicing machinery (39). We found similar changes in the immunofluorescence pattern using antibodies against SC-35 and RRM1-FF4/FF5 proteins upon transcriptional inhibition (Fig. 4). We conclude that the localization of FF4/FF5 fusion proteins to nuclear speckles is not dependent on active transcription and that the accumulation of these proteins shows behavior similar to that of splicing factors.

FF motifs are putative protein-protein interaction domains named for two conserved phenylalanine (F) residues (41). To study the requirement of particular residues within the FF4/FF5 domain for their nuclear targeting activity, we generated TCERG1[1–662]-FF4/FF5 constructs containing phenylalanine-to-alanine mutations at positions Phe-903, Phe-946, Phe-961, or the double mutants F903A,F946A, and F903A,F961A. We then expressed the wild-type FF4/FF5 domain and the phenylalanine-to-alanine mutants tagged with the ECFP at the amino terminus and examined their nuclear localization using immunofluorescence microscopy. As shown in Fig. 5, the TCERG1[1–662]-FF4/FF5 protein and the mutants within the FF4 domain (F903A, F946A, and the double mutant F903A,F946A) exhibited a similar nucleoplasm distribution with an increased signal in speckles. Mutation of the first phenylalanine of the FF5 domain at position 961 showed a decreased enrichment in speckles. The F903A,F961A double mutant localized in a diffuse pattern throughout the nucleoplasm without any evident accumulation on the speckles (Fig. 5). Expression of the various phenylalanine-to-alanine constructs was nearly identical, as assessed by Western blotting (supplemental Fig. S6). These data have been repeated with the

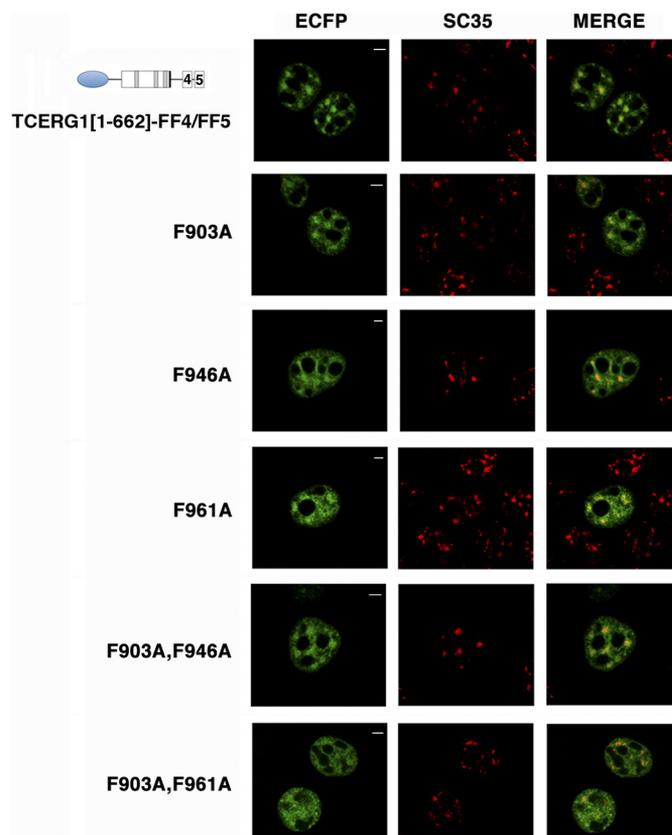
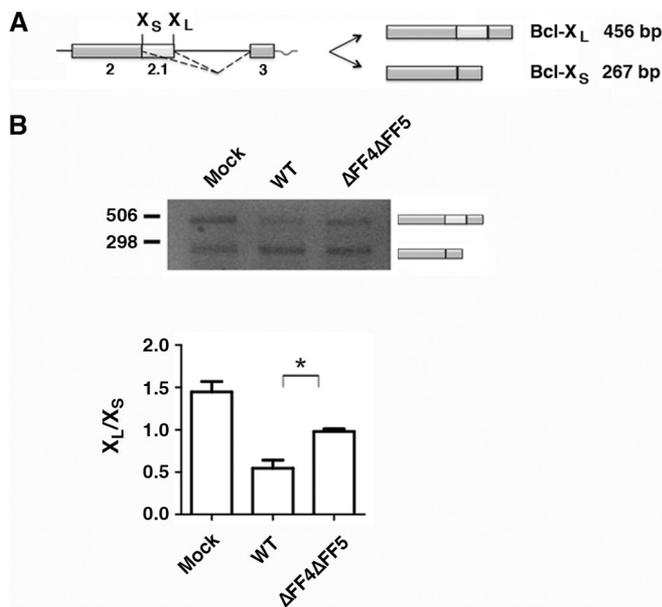


FIGURE 5. **Strictly conserved phenylalanine residues within the FF4/FF5 domains are required for the targeting to nuclear speckles.** Dual labeling of HEK293T cells transfected with TCERG1[1–662]-FF4/FF5 or the indicated phenylalanine-to-alanine mutant constructs (ECFP, green) with the anti-SC35 antibody (SC35, red) was performed. Individual and merged images of the cell are shown. Scale bars = 3  $\mu\text{m}$ .

## Novel Speckle Periphery-targeting Signal



**FIGURE 6. Deletion of the FF4/FF5 domains of TCERG1 affects *Bcl-x* alternative splicing.** *A*, schematic representation of the structure of the *Bcl-x* minigene is drawn with exons (boxes) and introns (lines). The *Bcl-x* pre-mRNA is alternatively spliced (dotted lines) to produce two major isoforms, *Bcl-x<sub>L</sub>* and *Bcl-x<sub>S</sub>*. *B*, splicing assay of the *Bcl-x* minigene in HEK293T cells transfected with the indicated constructs. Cells were harvested ~44 h after transfection and processed for RT-PCR. The data are presented as the ratio of *Bcl-x<sub>L</sub>* to *Bcl-x<sub>S</sub>* from three independent experiments (mean  $\pm$  S.D.). \*,  $p < 0.05$ .

same phenylalanine-to-alanine mutations in the context of a full-length TCERG1 protein (data not shown). These results demonstrate the involvement of the highly conserved phenylalanine residue within the FF5 domain in the targeting to the nuclear speckles. These results show that TCERG1 protein is able to localize to the periphery of nuclear speckles when the F-903 and F-946 residues within the FF4 domain have been mutated. This might indicate that the folding of this domain is less perturbed upon mutating these residues. This hypothesis is suggested by our NMR data (Fig. 2), which indicate a higher degree of stability for this domain.

**Role of the FF4/FF5 Domains in TCERG1 RNA Splicing Activity**—TCERG1 is involved in the process of pre-mRNA splicing and can affect the splicing of several minigene splicing reporters (see “Introduction”). We sought to investigate the role of the FF4/FF5 domains in the regulation of alternative splicing. Recently, we have found that TCERG1 can affect alternative splicing of the apoptosis gene *Bcl-x* (40). *Bcl-x* pre-mRNA uses an alternative 5' splice site in exon 2 to produce the antiapoptotic *Bcl-x<sub>L</sub>* or the proapoptotic *Bcl-x<sub>S</sub>* isoforms (Fig. 6A). A *Bcl-x* minigene was transfected into HEK293T cells in combination with vectors expressing wild-type or FF4/FF5-deleted TCERG1. The presence of different splice products was assessed by RT-PCR. As shown in Fig. 6B, cotransfection of wild-type TCERG1 with this reporter led to an increase in short-variant *Bcl-x<sub>S</sub>* with respect to transfection of the vector alone (Fig. 6B, WT). Furthermore, deletion of the FF4/FF5 domains reduced the ability of TCERG1 to affect 5' splice site selection (Fig. 6B,  $\Delta$ FF4 $\Delta$ FF5). These results demonstrate the involvement of the FF4/FF5 domains in the determination of 5' splice site

selection in the *Bcl-x* gene, which supports the assumption of the FF4/FF5 unit as a functional entity. Although more work is clearly needed to show how the FF4/FF5 mutants are affecting splicing of the *Bcl-x* reporter, our data suggest that TCERG1 might exert its activity through its localization to the nuclear speckle region.

## DISCUSSION

We have identified the sequences within the transcription and splicing-related factor TCERG1 that target it to the periphery of splicing factor-rich nuclear speckles. These consist of two contiguous FF domains, FF4 and FF5, of TCERG1: amino acids 878–1022. They are both necessary but not sufficient by themselves to achieve complete accumulation within the speckle region. We showed previously that the FF5 domain was implicated in the localization of TCERG1 to nuclear speckles (21). We show here that FF5 is necessary but not sufficient in itself to induce the accumulation of proteins to speckles. The adjacent FF4 domain is required for proper localization to the splicing speckled regions. The comparison of the TCERG1 speckle-targeting sequences with known speckle domain-targeting sequences (11–15, 17, 18) suggests that the TCERG1 speckle determinants are novel. This is consistent with the fact that the spatial distribution of TCERG1 is somewhat different from the distribution of many splicing factors that show more diffuse nucleoplasmic staining than TCERG1. The FF4 and FF5 domains of TCERG1 represent a novel nuclear speckle-targeting signal. They are necessary for TCERG1 localization and are sufficient for targeting a heterologous protein to speckles. To our knowledge, this targeting signal is the first sequence reported to direct proteins to the periphery of nuclear speckles at the interface between speckles and nearby transcription sites.

Compared with the better-characterized WW domain, the function of the FF domains is less well understood. The FF domain is an ~60-amino acid module that contains two strictly conserved phenylalanine residues near the N terminus and C terminus, respectively. FF domains often occur in repeated arrays of four to six domains separated by linker sequences of variable length (41), and this organization is likely to be important for their biological function. Structural studies of the FF domains of TCERG1 show a flexibility that is consistent with a domain organization model that visualizes the FF domains as multifunctional units acting as a scaffold to bind to a diverse repertoire of molecules (42, 43). The regulation of FF domain interactions is an intriguing issue. The TCERG1 FF domains could have individual specificities, or they could act in concert to achieve optimal recognition of binding partners. The interactions of TCERG1 with the phosphorylated carboxy-terminal domain (CTD) of RNAPII (44) and with the splicing factor Tat-SF1 (29) via the FF domains suggest that each FF domain within a tandem repeat possesses independent binding activity, conferring an ability to mediate many different protein-protein interactions through relatively weak binding affinity contacts. The possibility of each FF domain within a repeated structure mediating distinct ligand recognition was initially supported by our own data implicating the TCERG1 FF5 domain as the critical FF domain for TCERG1 localization to the periphery of speckles (21). Here, we have presented NMR data that chal-

lenges the current view of FF domains acting as separate entities and provided evidence that the FF domains can organize in pairs to optimally achieve a relevant biological function, such as targeting to speckles. Although future investigations are necessary, our results provide a starting point for understanding the structural basis of the targeting to speckles by FF domains.

Speckles are enriched in pre-mRNA splicing factors (1, 45). However, transcription and pre-mRNA splicing take place outside of the speckle compartment (7, 8, 46, 47). We described previously that the majority of speckle-associated TCERG1 is found along the periphery of nuclear speckles (21) where newly synthesized RNA is also found (7). The distribution of TCERG1 is consistent with a potential role of this factor in linking transcription with splicing machinery. Indeed, many of the factors that interact with TCERG1, such as SF1 and U2AF<sup>65</sup>, are also present in speckles. An interesting hypothesis is that TCERG1 could recruit processing components from speckles to the nearby transcription sites. If so, interfering with the localization of TCERG1 would probably disrupt some of these interactions. This is suggested by our experiments that showed that the signals determining TCERG1 nuclear speckle localization are important for its alternative splicing function (Fig. 6). Real-time mRNA biogenesis studies will address the question of whether TCERG1 shuttles between speckles and speckle-associated transcription sites.

TCERG1 undergoes posttranslational modification by SUMOylation (27). Computer algorithms aimed at locating putative SUMO acceptor residues predict two potential SUMOylation sites within the FF4/FF5 domains. Because SUMOylation influences the targeting of proteins to different cell compartments, we investigated whether putative SUMOylation of sequences within FF4/FF5 domains affects its targeting properties. Our results show that mutation of the putative SUMO acceptor lysine residues within FF4/FF5 generates a protein that localizes to the nucleus similarly to the wild-type protein, ruling out an effect of SUMO modification for the spatial distribution of this protein in the cell.<sup>4</sup> TCERG1 can also be modified by phosphorylation (48, 49). Computational predictions of phosphorylation sites using the NetPhos 2.0 server show that many serine and threonine residues within the FF4/FF5 domains are potential phosphorylation sites. Some of these sites are located in the loops of FF5. This is reminiscent of the phosphorylation status of SR proteins. The RS domain of SR proteins is extensively phosphorylated on serine residues, and this controls the subnuclear localization and activity of SR proteins (50). Similarly, phosphorylation events might regulate the nuclear localization and functionality of TCERG1. Moreover, TCERG1 interacts directly with phosphorylated RNAPII CTD via its FF repeats (44), and interactions between SR-related proteins and RNA polymerase II CTD have also been reported (52). Although no classical RRM has been identified, a previous study (53) described the presence of a putative RGG box in the N terminus of the *Chironomus tentans* homologue of TCERG1, which is also found in its human counterpart. The RGG box is a protein motif present in one class of RNA-binding proteins

involved in various aspects of RNA processing (54), is characterized by the presence of an RGG triplet, and is rich in arginines and glycines. In addition, TCERG1 has highly basic FF domains, raising the possibility that some of these, such as FF1–3, FF5, and FF6, may interact with RNA. In fact, we observed nucleic acid binding activity with recombinant FF1 and FF2 proteins.<sup>5</sup> On the basis of these observations, an exciting possibility remains that TCERG1 behaves as an SR protein in the coupling between transcription and splicing. Strikingly, the *Drosophila* homologue of human TCERG1 was identified as an RS domain-containing protein in a genome-wide survey of RS domain proteins (55).

*Acknowledgments*—We thank members of the laboratory for helpful suggestions, critical discussions, and comments. We also thank Javier Cáceres, Benoit Chabot, Mariano Garcia-Blanco, and Juan Valcárcel for providing reagents; José Luis Luque for confocal imaging assistance; Pau Martín-Malpartida for spectrum analysis; and Román Bonet for protein cloning.

## REFERENCES

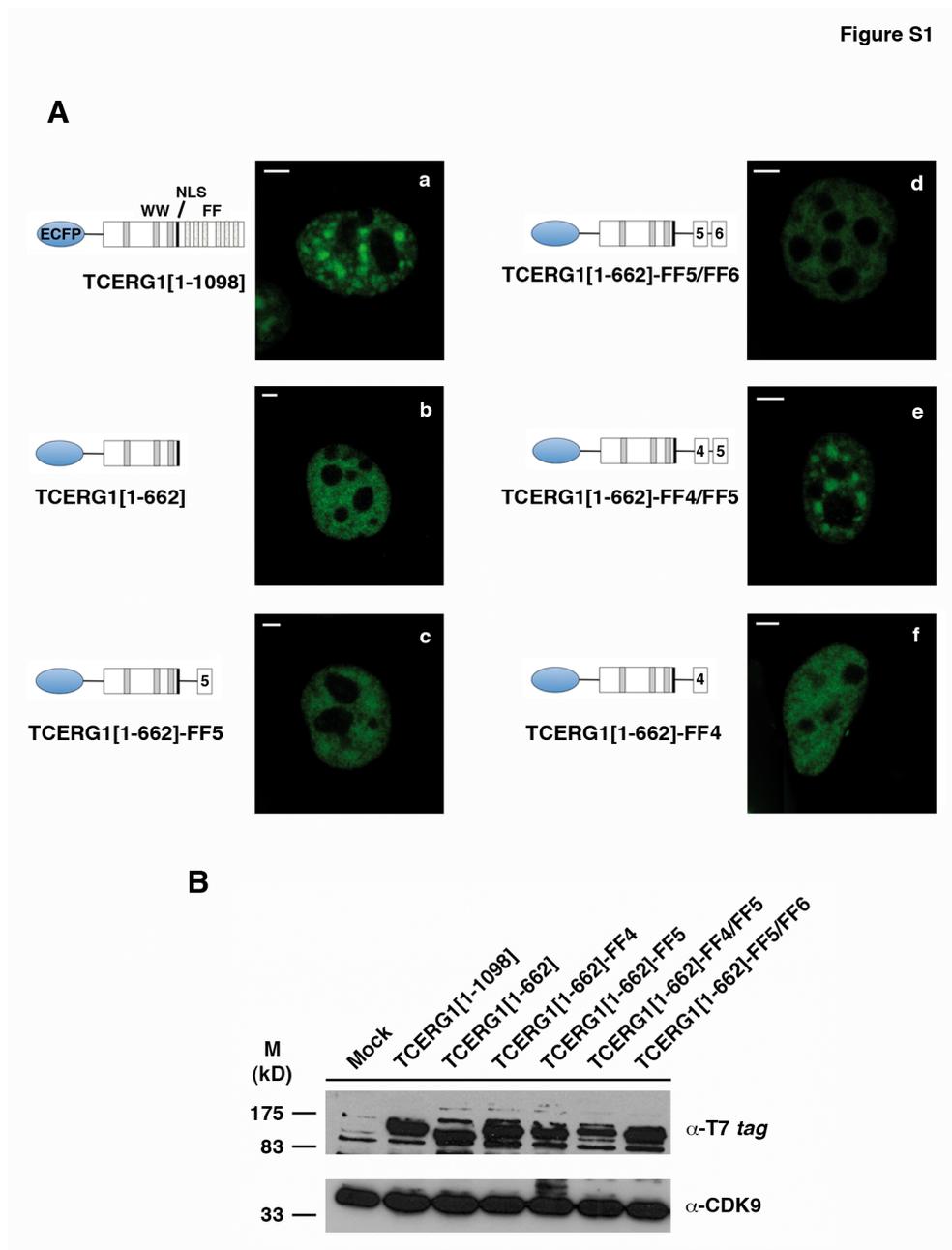
1. Lamond, A. I., and Spector, D. L. (2003) Nuclear speckles. A model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* **4**, 605–612
2. Spector, D. L. (1990) Higher order nuclear organization. Three-dimensional distribution of small nuclear ribonucleoprotein particles. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 147–151
3. Spector, D. L., Fu, X. D., and Maniatis, T. (1991) Associations between distinct pre-mRNA splicing components and the cell nucleus. *EMBO J.* **10**, 3467–3481
4. Huang, S., and Spector, D. L. (1992) U1 and U2 small nuclear RNAs are present in nuclear speckles. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 305–308
5. Fu, X. D., and Maniatis, T. (1990) Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **343**, 437–441
6. Spector, D. L., Schrier, W. H., and Busch, H. (1983) Immunoelectron microscopic localization of snRNPs. *Biol. Cell* **49**, 1–10
7. Wei, X., Somanathan, S., Samarabandu, J., and Berezney, R. (1999) Three-dimensional visualization of transcription sites and their association with splicing factor-rich nuclear speckles. *J. Cell Biol.* **146**, 543–558
8. Cmarko, D. (1999) Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection. *Mol. Biol. Cell* **10**, 211–223
9. Moen, P. T., Jr., Johnson, C. V., Byron, M., Shopland, L. S., de la Serna, I. L., Imbalzano, A. N., and Lawrence, J. B. (2004) Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis. *Mol. Biol. Cell* **15**, 197–206
10. Misteli, T., Cáceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F., and Spector, D. L. (1998) Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription *in vivo*. *J. Cell Biol.* **143**, 297–307
11. Li, H., and Bingham, P. M. (1991) Arginine/serine-rich domains of the su(wa) and tra RNA processing regulators target proteins to a subnuclear compartment implicated in splicing. *Cell* **67**, 335–342
12. Hedley, M. L., Amrein, H., and Maniatis, T. (1995) An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine-rich splicing factor. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11524–11528
13. Gama-Carvalho, M., Krauss, R. D., Chiang, L., Valcárcel, J., Green, M. R., and Carmo-Fonseca, M. (1997) Targeting of U2AF65 to sites of active splicing in the nucleus. *J. Cell Biol.* **137**, 975–987
14. Cáceres, J. F., Misteli, T., Sreaton, G. R., Spector, D. L., and Krainer, A. R. (1997) Role of the modular domains of SR proteins in subnuclear localiza-

<sup>4</sup> N. Sánchez-Hernández and C. Suñé, unpublished data.

<sup>5</sup> L. Ruiz and M. J. Macías, unpublished data.

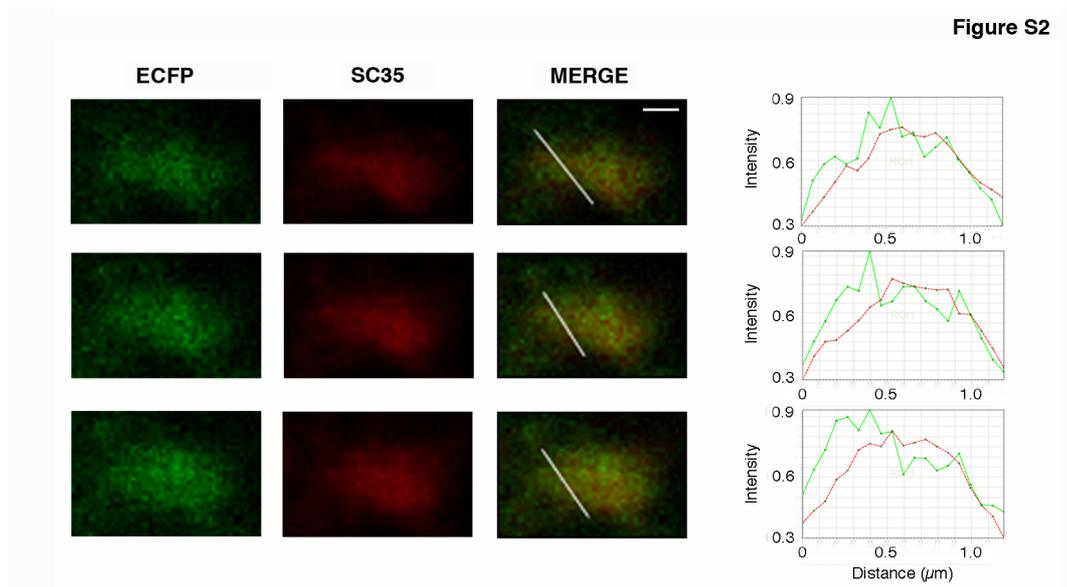
- tion and alternative splicing specificity. *J. Cell Biol.* **138**, 225–238
15. Eilbracht, J., and Schmidt-Zachmann, M. S. (2001) Identification of a sequence element directing a protein to nuclear speckles. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3849–3854
  16. Wagner, S., Chiosea, S., and Nickerson, J. A. (2003) The spatial targeting and nuclear matrix binding domains of SRm160. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3269–3274
  17. Ko, T. K., Kelly, E., and Pines, J. (2001) CrkRS. A novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. *J. Cell Sci.* **114**, 2591–2603
  18. Alvarez, M., Estivill, X., and de la Luna, S. (2003) DYRK1A accumulates in splicing speckles through a novel targeting signal and induces speckle disassembly. *J. Cell Sci.* **116**, 3099–3107
  19. Salichs, E., Ledda, A., Mularoni, L., Alba, M. M., and de la Luna, S. (2009) Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment. *PLoS Genet.* **5**, e1000397
  20. Goldstrohm, A. C., Greenleaf, A. L., and Garcia-Blanco, M. A. (2001) Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene* **277**, 31–47
  21. Sánchez-Alvarez, M., Goldstrohm, A. C., Garcia-Blanco, M. A., and Suñé, C. (2006) Human transcription elongation factor CA150 localizes to splicing factor-rich nuclear speckles and assembles transcription and splicing components into complexes through its amino and carboxyl regions. *Mol. Cell Biol.* **26**, 4998–5014
  22. Suñé, C., Hayashi, T., Liu, Y., Lane, W. S., Young, R. A., and Garcia-Blanco, M. A. (1997) CA150, a nuclear protein associated with the RNA polymerase II holoenzyme, is involved in Tat-activated human immunodeficiency virus type 1 transcription. *Mol. Cell Biol.* **17**, 6029–6039
  23. Suñé, C., and Garcia-Blanco, M. A. (1999) Transcriptional cofactor CA150 regulates RNA polymerase II elongation in a TATA-box-dependent manner. *Mol. Cell Biol.* **19**, 4719–4728
  24. Lin, K. T., Lu, R. M., and Tarn, W. Y. (2004) The WW domain-containing proteins interact with the early spliceosome and participate in pre-mRNA splicing *in vivo*. *Mol. Cell Biol.* **24**, 9176–9185
  25. Cheng, D., Côté, J., Shaaban, S., and Bedford, M. T. (2007) The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Mol. Cell* **25**, 71–83
  26. Pearson, J. L., Robinson, T. J., Muñoz, M. J., Kornblihtt, A. R., and Garcia-Blanco, M. A. (2008) Identification of the cellular targets of the transcription factor TCERG1 reveals a prevalent role in mRNA processing. *J. Biol. Chem.* **283**, 7949–7961
  27. Sánchez-Alvarez, M., Montes, M., Sánchez-Hernández, N., Hernández-Munain, C., and Suñé, C. (2010) Differential effects of sumoylation on transcription and alternative splicing by transcription elongation regulator 1 (TCERG1). *J. Biol. Chem.* **285**, 15220–15233
  28. Goldstrohm, A. C., Albrecht, T. R., Suñé, C., Bedford, M. T., and Garcia-Blanco, M. A. (2001) The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol. Cell Biol.* **21**, 7617–7628
  29. Smith, M. J., Kulkarni, S., and Pawson, T. (2004) FF domains of CA150 bind transcription and splicing factors through multiple weak interactions. *Mol. Cell Biol.* **24**, 9274–9285
  30. Marley, J., Lu, M., and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR* **20**, 71–75
  31. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe. A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293
  32. Bartels, C., Xia, T.-H., Billeter, M., Güntert, P., and Wüthrich, K. (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* **6**, 1–10
  33. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. (1992) Backbone dynamics of calmodulin studied by <sup>15</sup>N relaxation using inverse detected two-dimensional NMR spectroscopy. The central helix is flexible. *Biochemistry* **31**, 5269–5278
  34. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by <sup>15</sup>N NMR relaxation. *Biochemistry* **33**, 5984–6003
  35. Gasch, A., Wiesner, S., Martin-Malpartida, P., Ramirez-Espain, X., Ruiz, L., and Macias, M. J. (2006) The structure of Prp40 FF1 domain and its interaction with the crn-TPR1 motif of Clf1 gives a new insight into the binding mode of FF domains. *J. Biol. Chem.* **281**, 356–364
  36. Bonet, R., Ruiz, L., Morales, B., and Macias, M. J. (2009) Solution structure of the fourth FF domain of yeast Prp40 splicing factor. *Proteins* **77**, 1000–1003
  37. Bonet, R., Ramirez-Espain, X., and Macias, M. J. (2008) Solution structure of the yeast URN1 splicing factor FF domain: comparative analysis of charge distributions in FF domain structures-FFs and SURPs, two domains with a similar fold. *Proteins* **73**, 1001–1009
  38. Long, J. C., and Cáceres, J. F. (2009) The SR protein family of splicing factors. Master regulators of gene expression. *Biochem. J.* **417**, 15–27
  39. O'Keefe, R. T., Mayeda, A., Sadowski, C. L., Krainer, A. R., and Spector, D. L. (1994) Disruption of pre-mRNA splicing *in vivo* results in reorganization of splicing factors. *J. Cell Biol.* **124**, 249–260
  40. Montes, M., Cloutier, A., Sánchez-Hernández, N., Michelle, L., Lemieux, B., Blanchette, M., Hernández-Munain, C., Chabot, B., and Suñé, C. (2012) TCERG1 regulates alternative splicing of the Bcl-x gene by modulating the rate of RNA polymerase II transcription. *Mol. Cell Biol.* **32**, 751–762
  41. Bedford, M. T., and Leder, P. (1999) The FF domain: a novel motif that often accompanies WW domains. *Trends Biochem. Sci.* **24**, 264–265
  42. Murphy, J. M., Hansen, D. F., Wiesner, S., Muhandiram, D. R., Borg, M., Smith, M. J., Sicheri, F., Kay, L. E., Forman-Kay, J. D., and Pawson, T. (2009) Structural studies of FF domains of the transcription factor CA150 provide insights into the organization of FF domain tandem arrays. *J. Mol. Biol.* **393**, 409–424
  43. Lu, M., Yang, J., Ren, Z., Sabui, S., Espejo, A., Bedford, M. T., Jacobson, R. H., Jeruzalmi, D., McMurray, J. S., and Chen, X. (2009) Crystal structure of the three tandem FF domains of the transcription elongation regulator CA150. *J. Mol. Biol.* **393**, 397–408
  44. Carty, S. M., Goldstrohm, A. C., Suñé, C., Garcia-Blanco, M. A., and Greenleaf, A. L. (2000) Protein-interaction modules that organize nuclear function. FF domains of CA150 bind the phosphoCTD of RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9015–9020
  45. Spector, D. L. (2001) Nuclear domains. *J. Cell Sci.* **114**, 2891–2893
  46. Huang, S., and Spector, D. L. (1996) Dynamic organization of pre-mRNA splicing factors. *J. Cell Biochem.* **62**, 191–197
  47. Phair, R. D., and Misteli, T. (2000) High mobility of proteins in the mammalian cell nucleus. *Nature* **404**, 604–609
  48. Shimada, M., Saito, M., Katakai, T., Shimizu, A., Ichimura, T., Omata, S., and Horigome, T. (1999) Molecular cloning and splicing isoforms of mouse p144, a homologue of CA150. *J. Biochem.* **126**, 1033–1042
  49. Yu, L. R., Zhu, Z., Chan, K. C., Issaq, H. J., Dimitrov, D. S., and Veenstra, T. D. (2007) Improved titanium dioxide enrichment of phosphopeptides from HeLa cells and high confident phosphopeptide identification by cross-validation of MS/MS and MS/MS/MS spectra. *J. Proteome Res.* **6**, 4150–4162
  50. Misteli, T., and Spector, D. L. (1997) Protein phosphorylation and the nuclear organization of pre-mRNA splicing. *Trends Cell Biol.* **7**, 135–138
  51. Cáceres, J. F., and Krainer, A. R. (1993) Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *EMBO J.* **12**, 4715–4726
  52. Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R. V., Gentile, C., Gebara, M., and Corden, J. L. (1996) The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6975–6980
  53. Sun, X., Zhao, J., Kylberg, K., Soop, T., Palka, K., Sonnhammer, E., Visa, N., Alzhanova-Ericsson, A. T., and Daneholt, B. (2004) Conspicuous accumulation of transcription elongation repressor hrp130/CA150 on the intron-rich Balbiani ring 3 gene. *Chromosoma* **113**, 244–257
  54. Burd, C. G., and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615–621
  55. Boucher, L., Ouzounis, C. A., Enright, A. J., and Blencowe, B. J. (2001) A genome-wide survey of RS domain proteins. *RNA* **7**, 1693–1701

Supplemental data



**FIGURE S1. The FF4/FF5 domains are required for efficient targeting of TCERG1 to the speckle compartment.** A, The ECFP-TCERG1 fusion proteins were expressed by transient transfection in HEK293T cells and their nuclear localization was analyzed by confocal microscopy. A diagrammatic representation of the ECFP-

TCERG1 fusion proteins is shown at the left of each panel. The numbers in parentheses represent the TCERG1 amino acids contained in the construct. Shown are the three WW domains, the putative nuclear localization signal (NLS), and the six FF domains. Bars, 3  $\mu\text{m}$ . B, Cell lysates from transfected cells were analyzed by immunoblotting with the indicated antibodies to detect the TCERG1 and CDK9 proteins. Numbers to the left of the gels indicate molecular masses in kDa.



**FIGURE S2. The speckle-associated TCERG1[1-662]-FF4/FF5 localizes mainly to the periphery and outside the core region of speckles.** Seriated optical sections (z-plane step, 0.3  $\mu\text{m}$ ) through individual nuclear speckles were analyzed at low intensity. Individual staining (ECFP and SC35) and merge images are shown. Line scans showing local intensity distributions of TCERG1[1-662]-FF4/FF5 in green and SC35 in red are shown to the right of the panels. Bar, 0.3  $\mu\text{m}$ .

Figure S3

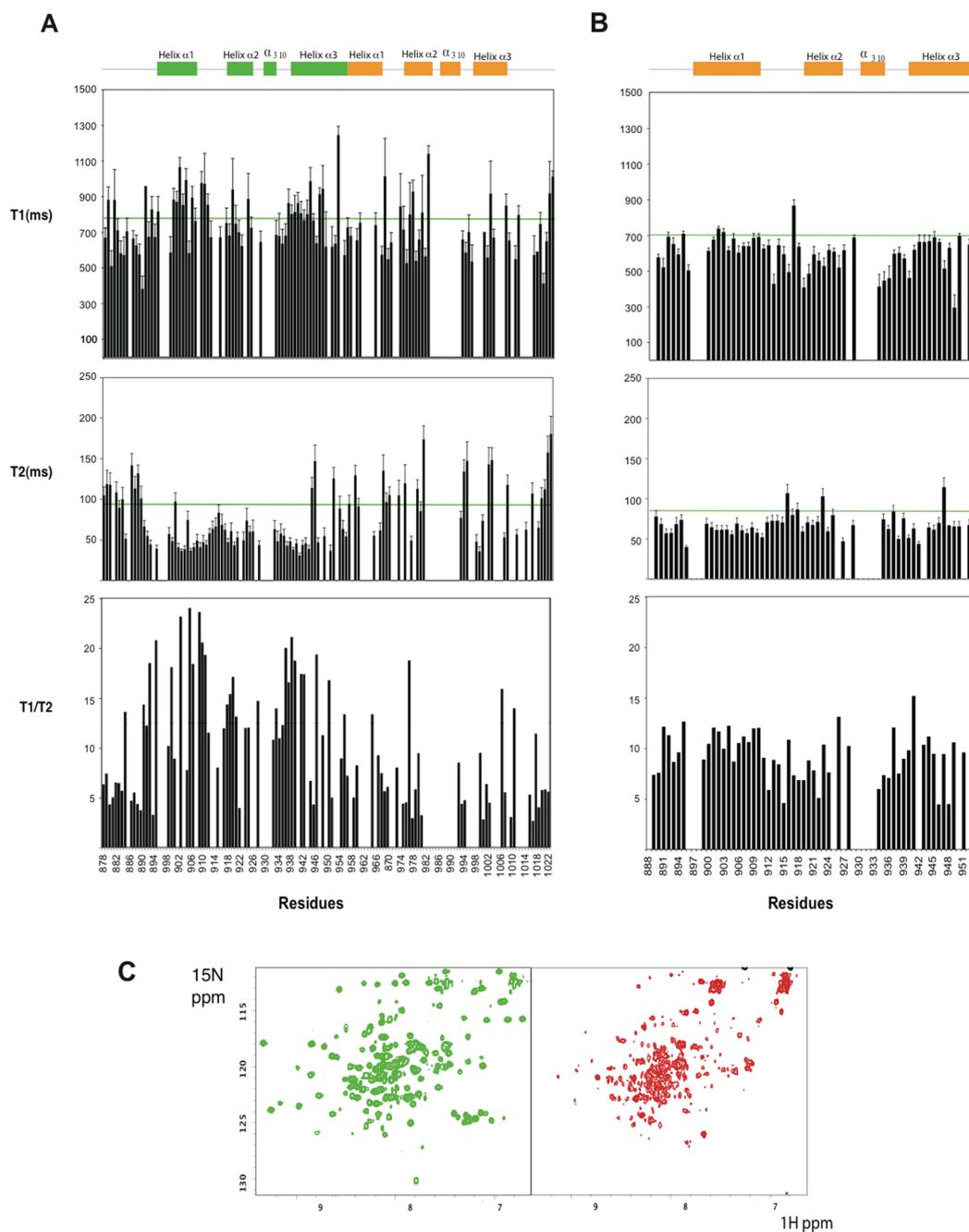


FIGURE S3. **Backbone amide  $^{15}\text{N}$  relaxation data.** A, Plots of  $T_1$ ,  $T_2$  and  $T_1/T_2$  ratio for the FF4-FF5 construct. Unassigned residues, proline residues that lack a proton amide and overlapped peaks were excluded from the analysis. B, Plots of  $T_1$ ,  $T_2$  and  $T_1/T_2$  ratio for FF4 construct. C,  $^{15}\text{N}$ -HSQC spectra at 298K for the FF4-FF5 (green)

and FF5-FF6 (red) pairs. The comparison of the NMR data shows a significant difference in peak dispersion. Data show no structural organization when FF5 is combined with FF6.

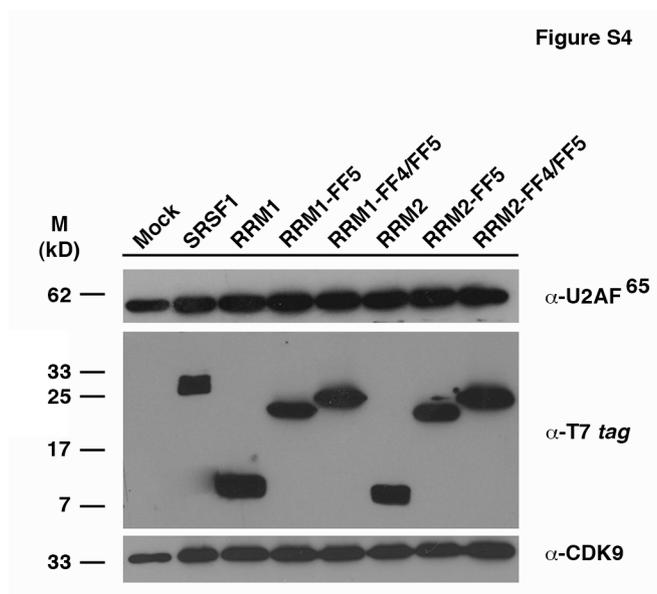


FIGURE S4. **Analysis of the expression levels by immunoblotting.** Cell lysates from transfected cells were analyzed by immunoblotting with the indicated antibodies to detect the TCERG1, U2AF<sup>65</sup>, and CDK9 proteins. Numbers to the left of the gels indicate molecular masses in kDa.

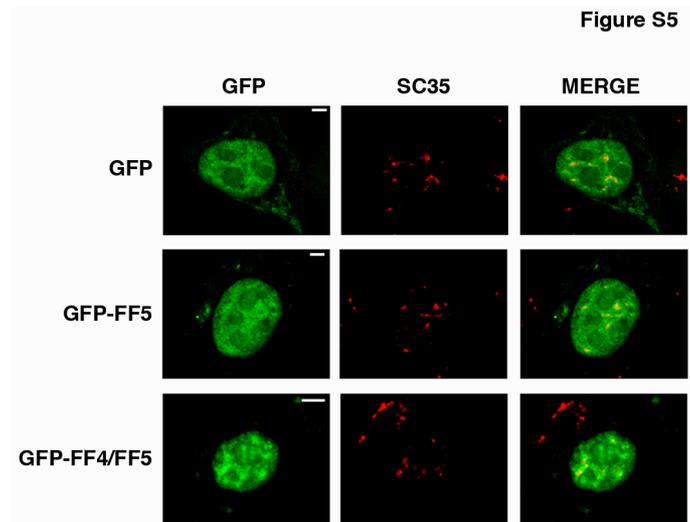


FIGURE S5. **The presence of the FF4/FF5 domains targets GFP to nuclear speckles.** GFP, GFP-FF5, and GFP-FF4/FF5 were expressed by transient transfection in HeLa cells and dually labeled with antibodies directed against the SC35 protein. In all cases, colocalization of expressed proteins with the endogenous marker was assessed by confocal microscopy. Bars, 3  $\mu$ m.

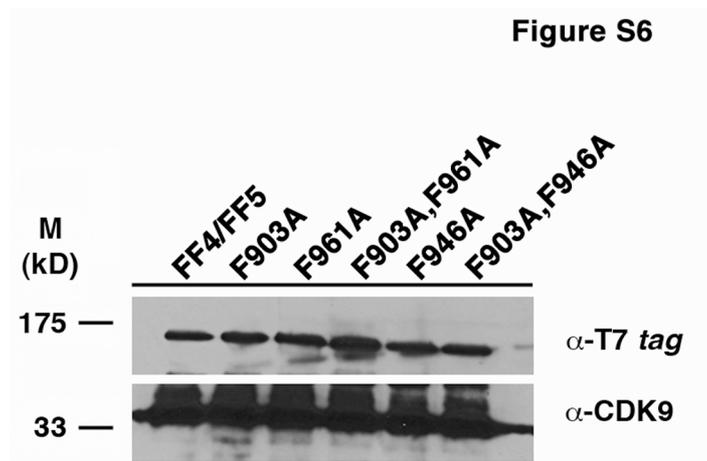


FIGURE S6. **Analysis of the expression levels by immunoblotting.** Cell lysates from transfected cells were analyzed by immunoblotting with antibodies against the T7-tag

and CDK9 to detect the TCERG1[1-662]-FF4/FF5 (FF4/FF5) and the indicated phenylalanine-to-alanine mutant constructs. Numbers to the left of the gels indicate molecular masses in kDa.

## Review Article

# Spatial Organization and Dynamics of Transcription Elongation and Pre-mRNA Processing in Live Cells

Miguel Sánchez-Álvarez,<sup>1</sup> Noemí Sánchez-Hernández,<sup>2</sup> and Carlos Suñé<sup>2</sup>

<sup>1</sup>*Dynamical Cell Systems Team, Section of Cellular and Molecular Biology, The Institute of Cancer Research, London SW3 6JB, UK*

<sup>2</sup>*Department of Molecular Biology, Instituto de Parasitología y Biomedicina “López Neyra” (IPBLN-CSIC), 18100 Armilla, Spain*

Correspondence should be addressed to Carlos Suñé, csune@ipb.csic.es

Received 22 July 2011; Accepted 5 September 2011

Academic Editor: Sebastián Chávez

Copyright © 2011 Miguel Sánchez-Álvarez et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

During the last 30 years, systematic biochemical and functional studies have significantly expanded our knowledge of the transcriptional molecular components and the pre-mRNA processing machinery of the cell. However, our current understanding of how these functions take place spatiotemporally within the highly compartmentalized eukaryotic nucleus remains limited. Moreover, it is increasingly clear that “the whole is more than the sum of its parts” and that an understanding of the dynamic coregulation of genes is essential for fully characterizing complex biological phenomena and underlying diseases. Recent technological advances in light microscopy in addition to novel cell and molecular biology approaches have led to the development of new tools, which are being used to address these questions and may contribute to achieving an integrated and global understanding of how the genome works at a cellular level. Here, we review major hallmarks and novel insights in RNA polymerase II activity and pre-mRNA processing in the context of nuclear organization, as well as new concepts and challenges arising from our ability to gather extensive dynamic information at the single-cell resolution.

## 1. Introduction

In eukaryotic cells, the regulation, expression, and subsequent processing steps of genomic sequences tend to be localized to defined spaces within the nucleus [1]. In the interphase nucleus, uncondensed chromosomes do not expand randomly but occupy defined volumes termed “chromosome territories,” whose relative positioning has recently been suggested to be determined by, or at least correlated with, differentiation stages and specific contexts [2–4]. This architecture facilitates the intermingling of specific subsets and combinations of genes that need to be coregulated in a given situation [4–7]. Indeed, active genes are most often positioned in the periphery of chromosome territories, while inactive genes remain located within more inaccessible areas of these regions. Although the molecular basis for this dynamic behavior of chromatin is not yet well understood, there are a significant number of studies supporting this concept, thus suggesting a novel layer of complexity in the regulation of gene expression. Genes are not inert entities waiting for the adequate subset of transcription factors to initiate the assem-

bly of a processive RNA polymerase II (RNAPII) complex; instead, the dynamic positioning contributes to their activation state and correlation with other gene units and regulatory elements, such as enhancers and insulators [8–11].

Chromosome territories delimitate a region of the nucleus (usually termed the interchromatin space) that is relatively empty of dense chromatin and is hypothesized to be highly interconnected across the nucleus with a higher-order organization [12–15]. The delimited interchromatin volume contains not only most of the transcriptional activity at its boundaries but also contains several nonmembrane-bound dynamic structures—nuclear bodies—highly enriched with specific subsets of nuclear factors [16–18]. These nuclear bodies include Sam68 bodies, PML bodies, paraspeckles, Cajal bodies, and nuclear speckles.

## 2. The Transcription Factory: A Spatial and Functional Unit for RNAPII Transcription

In the early 1990s, two groups reported the use of novel techniques that allowed the visualization of transcriptionally

active sites within the nucleus through the incorporation of bromo-UTP in nascent transcripts [19, 20]. In these experiments, nascent transcripts remained immobilized at the site of their chromatin template, and they were studied in great detail using confocal and electron microscopy. Notably, the number of observed active sites appeared to be considerably lower than the estimated number of active molecules of RNAPII [21–23]. These discrete structures colocalize with hyperphosphorylated forms of RNAPII and are resistant to DNA digestion and extraction of soluble fractions [20, 21, 24–26]. These results suggest the existence of an immobile pool of hyperphosphorylated RNAPII within the eukaryotic nucleus. Subsequent fluorescence recovery after photobleaching- (FRAP-) based experiments performed on cells expressing a GFP-tagged construct of RNAPII support this interpretation [27]. Given that the number of observed transcription sites was significantly lower than the number of elongating RNAPII molecules as assessed by *in vitro* run on assays [22, 23], a model was proposed in which several active (mostly elongation-competent) RNAPII units assemble into higher-order structures termed “transcription factories” [21]. According to this model, chromatin loops are tethered to the factories through RNAPII or/and transcription factors for transcription to occur (recently reviewed in (Cook, [28])).

This model is consistent with the looped conformation model that several other independent approaches have suggested exists for an active eukaryotic gene [29–31]. In this model, upon recognition and activation by specific factors, the promoter sequence of the gene unit is tethered to the RNAPII subunit of the factories, and this attachment would be maintained during the transcription of the whole gene sequence, which is “reeled” on the RNAPII. This arrangement provides an additional layer of control and coordination over the different stages of transcription and positions the RNAPII units for subsequent rounds of transcription.

The existence of factories provides us with an elegant conceptual framework to explain the coregulation of functionally related groups of genes in specific contexts [30]. It has been observed that some of these active genes tend to be found in close proximity at a much higher frequency than would be expected by chance [32, 33]. Indeed, these genes tend to share a factory when they are positioned in close proximity, as observed by immunolabeling elongating RNAPII [33]. Although the structural resolution of transcribed genes is still technically limited, nevertheless, this reflects the potential crosstalk that can exist between the transcription factors recruited to each coregulated promoter. Some of the examples consistent with this model are the NF- $\kappa$ B/TNF $\alpha$  activation axis [34, 35] and the ER $\alpha$  module system [34]. Moreover, genes of different sizes and elongation timing coimmunoprecipitate with the elongating form of the RNAPII in a fashion consistent with the model in which they share the same factory and slide along the “polymerase reading heads” in sequential rounds of transcription, rather than just recruiting mobile polymerase complexes from proximal storage sites and undergoing independent read-throughs [35]. Transcription factories are also consistent with data suggesting that genes with shared features, such as promoter composition and the presence or absence of introns, tend

to associate among each other [36]. Finally, transcription factories also provide an explanation for observations that indicate that promoter composition and associated events can influence subsequent stages of transcription elongation [37, 38].

Recent studies have reported on the stability of RNAPII foci upon disrupting transcription [39]. Interestingly, treatment of cells with 5,6-dichloro-1- $\beta$ -D-ribofuranosylbenzimidazole (DRB; a highly specific inhibitor of the positive elongation transcription factor, P-TEFb, and thus an inhibitor of elongating polymerases) does not abolish the association of previously engaged genes with the RNAPII foci, at least for erythroid lineage-specific genes [39]. Observations in agreement with this model include independent genome-wide chromatin immunoprecipitation- (ChIP-) based studies that demonstrate that a significant number of genes is “primed” for transcription. These inactive genes have paused RNAPII complexes at their promoter regions and, upon gene activation, are released from the paused state, allowing elongation to proceed [40].

Initial studies using *in situ* spectroscopy have recently been carried out to determine the composition of the transcription factories [41]. In these studies, the authors demonstrated the existence of clearly defined ribonucleoprotein structures that coincide with sites of active transcription (the perichromatin fibrils), as assessed by BrU pulse incorporation and immunogold labeling. The size and estimated composition of carbon and nitrogen in these structures support the existence of the proposed model of assembled transcription factories, creating a more refined structural model in which the effector subunits of the RNAPII face outwards [41, 42].

Another important feature of transcription factories is the enhancement of the physical and functional coupling of transcription and downstream RNA processing steps. This is facilitated by the regulated recruitment of neighboring machinery for cotranscriptional mRNA maturation in an appropriate fashion and timing. This notion would expand the category of these structures to integrated “mRNA factories,” providing an intuitive physical framework for the numerous observed interactions among transcriptional and mRNA processing factors [43–46]. Other essential processes in the regulation of pre-mRNA synthesis, such as chromatin remodeling and histone modification, would similarly benefit from such a design [30].

Although this model of transcription factories (Figure 1) explains many observations of gene expression and nuclear organization, there are many intriguing questions that need to be resolved. What are the molecular mechanisms underlying the appropriate targeting of activated gene sequences to a factory, and how are they integrated in a given regulatory context? What is the inner structure of factories in the cell at a resolution beyond the conventional light diffraction limits? How are these structures assembled and organized during the cell cycle according to the requirements of the cell? Does it require the existence of an underlying structural scaffold or “nucleoskeleton?” How are different regulatory hallmarks, such as the phosphorylation cycle proposed for the CTD of the RNAPII during its progression through the transcription

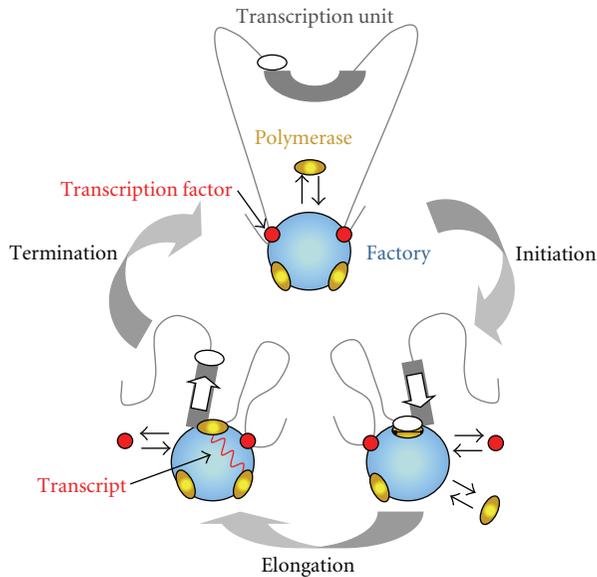


FIGURE 1: Model of the basic structure and function of RNAPII transcription factories. (a) A “gene-loop” is recruited upon activation to the transcription factory, which contains immobilized subunits of RNAPII. (b) The gene-loop is then “reeled” onto RNAPII for transcriptional elongation. (c) Upon termination, the anchoring of the locus allows for subsequent rounds of transcription. Adapted, with permission, from *The Journal of Cell Science* [47].

of a gene unit, integrated into the context of these structures? Finally, what is the functional relationship between transcription factories and other nuclear compartments related to the biogenesis of the mRNA?

### 3. Nuclear Speckles and the Regulation of Transcription and Pre-mRNA Processing

Many independent studies in the last two decades have led to a model in which the maturation of nascent transcripts take place simultaneously to their synthesis, that is, cotranscriptionally [48, 49]. This may be specific to a subset of genes or even to specific introns of a gene and is therefore considered not to be strictly required for the completion of pre-mRNA processing itself [50]. However, cotranscriptional processing allows for the functional coupling of the different steps of RNA biogenesis. The bidirectional interdependence among chromatin conformation and posttranslational modifications, in both transcription and different steps of pre-mRNA processing, constitutes an additional layer in gene expression regulation [51–56]. Moreover, it may play a pivotal role in complex processes, such as neuronal differentiation and activity, global integration of RNA processing signatures and DNA damage, and developmental programs [57–60].

If pre-mRNA processing is performed largely in a regulated cotranscriptional fashion, the dynamic distribution of pre-mRNA processing factors should be correlated with the organization of transcriptionally active sites in the nucleus. The distribution of pre-mRNA processing factors in the eukaryotic nucleus, as observed using immunofluorescence

staining and light microscopy, is not homogeneous and shows a dynamic pattern of localized accumulation in 10–30 irregular domains termed speckles, “SC35 domains” or “splicing factor compartments” (SFCs). At the level of electron microscopy, they correspond to two distinct structures: (i) interchromatin granules clusters (IGCs), composed of particles measuring 20–25 nm in diameter and (ii) perichromatin fibrils, 3–5 nm fibrillar structures localized both at the periphery of IGCs and in other nucleoplasmic regions, which are the sites of nascent pre-mRNAs (for extensive reviews, see [61, 62]) These structures were first identified using immunostaining with specific antibodies against different small nuclear ribonucleoproteins (snRNPs) [63, 64]. This and other observations that show the presence of poly(A) RNA colocalizing with snRNPs and SC35-rich domains [65, 66] further support a link between nuclear speckles and pre-mRNA metabolism. Pioneering mass spectrometry studies [67, 68] and a still-growing list of publications using immunofluorescence or live-cell imaging labeled with tagged constructs corroborate a marked enrichment of these compartments with factors involved in pre-mRNA transcription and processing, especially pre-mRNA splicing.

Several models, which are not mutually exclusive, have been proposed to explain the role of these nuclear bodies in the regulation of gene expression: (1) they function as storage/assembly/modification compartments that can supply processing factors to the surrounding active transcription sites; (2) they function as sequestering sites for the dynamic control of transcription and processing factors; (3) they serve as functional “hubs” for coregulated genes and their products; (4) they play an active role in posttranscriptional pre-mRNA processing and surveillance and/or in the coupling of early steps of mRNA biogenesis (Figure 2).

The concept that nuclear speckles are transcriptionally inactive compartments that serve as storage or recycling sites of pre-mRNA metabolism complexes from which these complexes are recruited to nearby sites of active transcription according to demand is a widely held view supported by many experimental results [69–77]. Importantly, recruitment to active sites of transcription requires the integrity of the carboxyl terminal domain of the RNAPII [78], which indicates that transcriptional elongation plays a critical role in the recruitment of pre-mRNA processing factors.

This proposal is compatible with the view that speckles act as inhibitory sites where specific factors are actively sequestered when their functional repression is required. The essential splicing factor SRSF1 is sequestered into these regions upon the induction of stress through a mechanism dependent on the dynamic interaction of SRPK with stress chaperone complexes, including Hsp70 and Hsp90 [79]. Similarly, the *MALAT1* large noncoding (nc) RNA has been proposed to regulate the phosphorylation-dependent dynamics of splicing factors and their equilibrium between nucleoplasm availability and nuclear speckle sequestration [80]. Linking transcriptional elongation control to this model, the 7SK small ncRNA, which is a scaffold component of transcription elongation CDK9-CycT1 inactive complexes together with HEXIM proteins (see below), has been proposed to function, at least partially, by sequestering

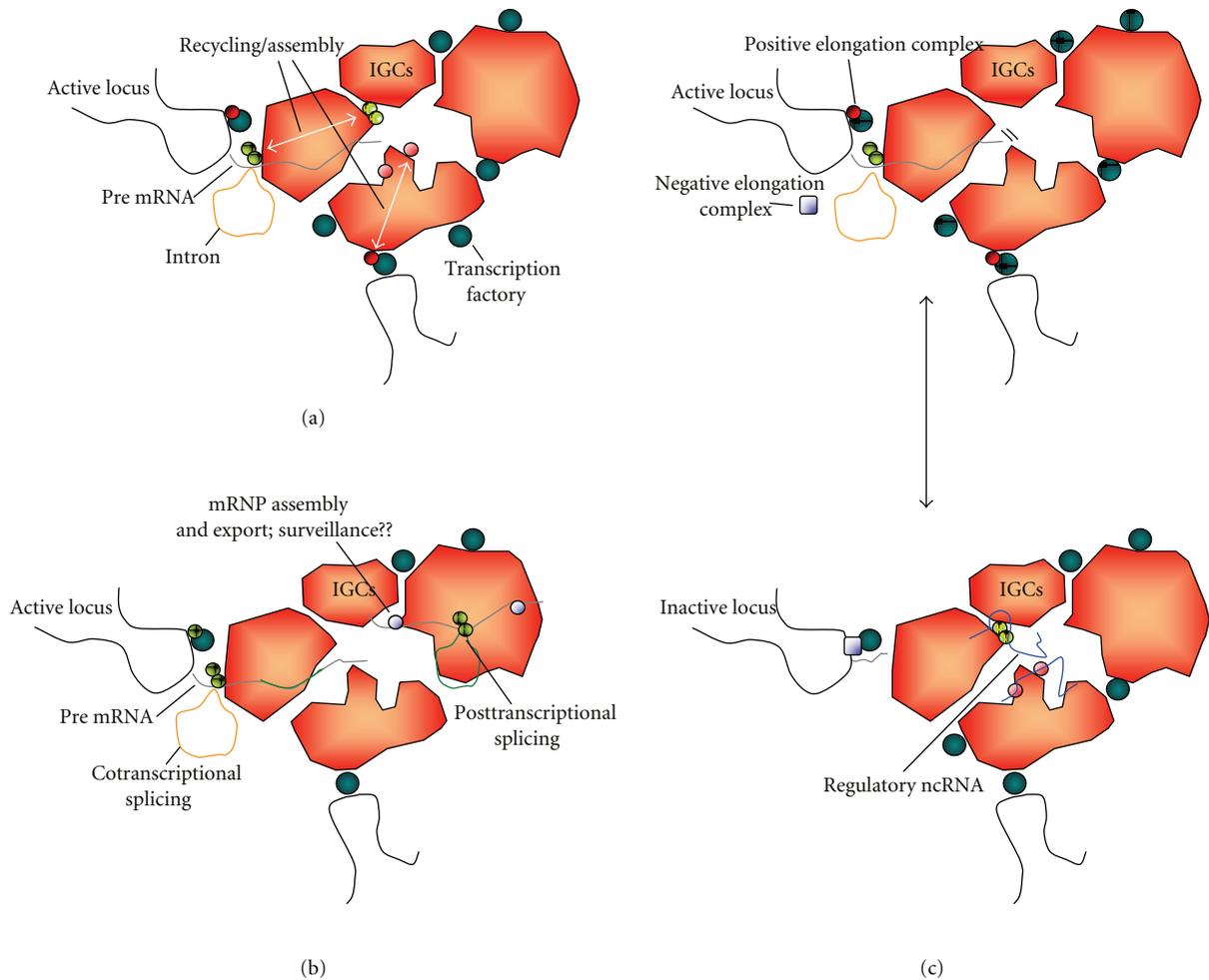


FIGURE 2: Three potential, nonexclusive models for the role of the interchromatin granule clusters (IGCs) in the regulation of transcription and pre-mRNA processing machinery. IGCs are depicted surrounded by transcriptionally competent sites or “factories” (grey beads). (a) IGCs may be specialized sites for the recycling and assembly of transcription (dark and light red beads) and pre-mRNA processing complexes (dark and light green beads) through regulated posttranslational modification cycles. Dark and light hue code denotes active and inactive pools of factors, respectively. (b) Posttranscriptional processing steps and potential surveillance of mRNA quality may be integrated in these structures, constituting a “checkpoint link” between mRNA transcription and mRNP assembly and export. A given transcript may include both introns that are spliced cotranscriptionally outside of the IGCs (orange lariat) and intron sequences that are processed posttranscriptionally (dark brown stretch and lariat). The later event may be also coupled in the IGCs to surveillance mechanisms, mRNP assembly (blue beads), and export. (c) Specific subsets of nuclear factors, such as ncRNAs (MALAT1, 7SK; see main text; depicted in the lower panel as thin blue threads), can function as active quenchers or sequesters of transcription and pre-mRNA processing factors (red and green beads, resp.), blocking the recruitment of these complexes from the IGCs to nearby active sites of transcription.

these inactive P-TEFb complexes at nuclear speckles [81]. However, there is currently little information known about the actual relevance of these dynamic interactions regarding the response of the cell to specific stimuli or its correlation with changes in transcriptomic profiles. For example, P-TEFb components colocalize at nuclear speckles with these negative regulators and also with the transcription activator adaptor Brd4 [82]. Rigorous quantitative approaches, as opposed to qualitative descriptions, especially those addressing the kinetics of these interactions, such as FRET-based studies, may complement these lines of research.

Nuclear speckles are most often adjacent to a relatively high density of transcriptionally active regions [21, 25, 46, 71,

83–85], and these active sites mostly represent elongation-competent complexes. Many of these active units correspond to specific, functionally interrelated protein-coding genes, and their juxtaposition to speckles may constitute an important part of their functional program, as has been suggested for genes involved in muscle differentiation [86]. A proposed role for nuclear speckles in these associations is the recruitment of splicing factors at specific active genes in early G1 phase, signaling for the subsequent recruitment of other functionally related gene units later on in the cell cycle [87, 88]. However, this model may be incomplete for explaining the highly dynamic behavior of these structures as observed by live-cell imaging. The inducible recruitment

of active genes compared to the dynamics of nuclear speckles has indeed been observed in live cells recently [89]. The authors proposed the following three different, nonmutually exclusive interpretations for this dynamic association such that taking the induced locus as a reference, the speckles could either be (i) assembling *de novo*, (ii) gathering by “coalescence” of smaller speckles, or (iii) recruiting the active gene to their surface. Interestingly, this inducible spatial correlation was dependent on the integrity of the inducible promoter driving the construct, *Hsp70*, which is known to be regulated by the activity of this model gene at the stage of elongation. Again, these observations might indirectly support a functional coordination between transcriptional elongation and the recruitment of pre-mRNA processing machinery. It remains to be fully resolved whether the nuclear speckles have an active role in the higher-order organization and functional coordination of the expression of specific genes or whether they rather arise as a consequence of the spatial concentration of required factors in areas proximal to active, coregulated genes. Importantly, it also remains to be elucidated if there is any posttranscriptional advantage for coregulated genes to converge at the same speckle. It would be interesting to unravel, for example, if subsets of genes that preferentially localize to the nuclear speckle periphery are enriched in genes that are mostly regulated at the level of elongation and whether both their synthesis and processing are enhanced upon appropriate recruitment to these compartments.

What is the behavior of the synthesized pre-mRNAs and mRNAs as related to nuclear speckles? A majority of introns are spliced, presumably in a cotranscriptional fashion, outside of the nuclear speckles [90–92]. This peripheral region of nuclear speckles can be therefore considered a potential interphase for cotranscriptional pre-mRNA processing. However, some introns undergo posttranscriptional processing, and their relative accumulation can be observed in these compartments, as in the case of intron 26 of the *COL1A1* gene. Mutations that alter the splicing of this intron provoke increased accumulation of the transcript in nuclear speckles [91]. These data raise the possibility that nuclear speckles have a role in posttranscriptional or even postmaturation steps linked to mRNA surveillance and/or nuclear export pathways [90, 93]. Recently, a role for these structures has been proposed in the regulation of posttranscriptional “quality assessment” and the export of Herpesvirus mRNAs [94]. In this case, inducible recruitment of “viral transcription factories” to the peripheral areas of these structures was observed. Bimolecular fluorescence complementation (BiFC) experiments, showing the interaction between the exon-junction complex component Y14 and nuclear export factor 1 (NXF1) and their significant accumulation in nuclear speckles and peripheral areas, further indicate that export-competent spliced mRNAs localize at speckles and that this domain might play an active role in mRNA processing, including maturation and/or transport [95, 96]. This model is also compatible with the apparent subcompartmentalized organization of nuclear speckles [97, 98], which renders a “sponge-like” or “porous” structure that would easily allow for the transit of macromolecular

complexes through its interior [99]. Coupling transcription elongation and mRNP assembly with export has been described in yeast and humans, and its potential relationship with the dynamics of mRNP transit through the speckles has been suggested recently [100–104]. Nuclear speckles may represent specialized compartments for the appropriate regulation and coordination of these functions.

Intriguingly, although transcription does not take place within nuclear speckles, a large subset of transcription factors accumulates in these regions, and transcription elongation factors are specifically enriched in these regions. RNAPII is also found to be associated with these compartments [44, 68, 105]. Quantitative laser confocal analysis of ultrathin cryosections has shown that nuclear speckles do not act as major storage sites for inactive complexes, but they instead contain a minor, stable pool of RNAPII molecules phosphorylated at the serine 2 residue of the carboxyl-terminal domain (CTD). Importantly, this subpopulation of RNAPII-2pSer is insensitive to DRB treatment [85, 105]. This fact and the absence of *de novo* synthesis of transcripts at nuclear speckles, as determined by UTP analogue incorporation, suggest that these hyperphosphorylated forms of RNAPII are not engaged in active transcription. An intriguing interpretation might be that these RNAPII molecules serve as a platform for posttranscriptional splicing of transcripts that are trafficking through the speckles.

#### **4. Gene Expression Regulation at the Single-Cell Resolution: Studying the Kinetics of RNA Biogenesis**

The study of spatial and dynamic properties is essential for understanding gene expression regulation. Techniques, such as FRAP and fluorescence loss in photobleaching (FLIP), are used to obtain very detailed information about diffusion rates, residency times, or proportion of immobile or stably tethered subpopulations of a given molecule in a delimited volume in the cell at a very high temporal resolution. Fluorescence Resonance Energy Transfer (FRET) and fluorescence lifetime imaging microscopy (FLIM), BiFluorescent complementation (BiFC), and specific applications based on fluorescence correlation spectroscopy (FCS) allow the monitoring and semiquantitation of close (mostly direct) interactions between molecules and the mapping of such interactions in relation to different structures of the cell. The use of these techniques in the study of transcription and pre-mRNA processing has led to interesting new concepts regarding their regulation and nuclear organization [106–112].

Not all nuclear factors freely diffuse through the nuclear space. The movement of some of these are compatible with a model in which factors “scan” unspecific genomic sequences or/and bind components of the RNA machinery through a weak and transient binding until they engage in a favorable, specific assembly on their target sites [113]. Nuclear functions and organization likely arise not from the static state of their components but from an extremely dynamic equilibrium between multiple functional interactions

[10, 114]. Rino and coworkers [115] found that nuclear speckles acquired a rounder and more quiescent morphology, as expected, upon transcription elongation inhibition using the P-TEFb inhibitor DRB. However, when studying the interchange rate of molecules bound to nuclear speckles within the nucleoplasm pool using FLIP, they found that the fluorescence was lost at a higher speed than it was in cells that had not been treated with DRB. Similar studies have been conducted for the transcription elongation complex P-TEFb in the context of Tat-mediated transactivation of the HIV-1 [53, 116]. Interestingly, these studies suggest that a potential mechanism by which Tat might contribute to P-TEFb-mediated transactivation is through the stabilization of CDK9 binding to the transcription site, increasing its residency time by almost tenfold. These observations suggest that cellular factors are constantly engaged in dynamic and highly transient interactions even within some apparently static structures.

Specific interactions among different spliceosomal components have also been studied in the context of nuclear organization and live-cell behavior using FRET and FLIM techniques. The dynamic model mentioned above is also compatible with the presence of preassembled subcomplexes, such as the spliceosomal components, which can be dynamically recruited to form even higher-order functional complexes [114]. Of note, these subcomplexes exist in even the absence of ongoing transcription and in nuclear compartments in which they do not actively function [75, 115, 117].

Recent adaptations of these techniques have been used to study fluorescent proteins that interact with high affinity and specificity to DNA and RNA sequences, such as the LacZ bacterial repressor and the MS2 and PP7 nucleocapsid coating proteins, respectively. The use of engineered constructs containing several tandem repeats of these target elements allows for the efficient recruitment of fluorescent molecules to the desired chromatin template or nascent transcript, amplification of the signal, and reliable detection by conventional light microscopy in living cells. FRAP measurements can then be used to estimate the rates of transcript elongation and release because these stages of mRNA biogenesis correlate with distinct kinetic steps, which can be mathematically dissected with appropriate modeling using the gathered data. Additionally, by combining this transcript-tagging system with the tagging of RNAPII with a distinct fluorophore, the dynamics of promoter binding and transcription initiation can also be inferred using the aforementioned data of transcript synthesis kinetics [118]. Interestingly, transcription appears to be a rather inefficient process, as only approximately 1% of recruited RNAPII molecules are engaged in processive transcription; however, these figures may vary considerably depending on the genetic model studied [119, 120]. These types of studies have led to an estimation of the rate of elongation for the human RNAPII of  $\sim 4.3$  kb/min, although these figures can also vary widely depending on the experimental setting used [118, 121]. For example, a study based on the HIV-1 gene yielded unexpectedly long pausing times for RNAPII units located at the proximal promoter and 3' terminal regions with an estimated elongation rate of  $\sim 1.7$  kb/min [53, 119]. An important, unresolved issue

is the unification of the mathematical models used to infer the kinetic properties of transcription elongation because considering or excluding certain phenomena in the reference model can drastically affect the entire interpretation [121].

These tools have been used not only for characterizing transcription dynamics, but also for facilitating the study of the *in vivo* functional coupling between transcription elongation and pre-mRNA processing, allowing us to obtain novel insights into the basis of pre-mRNA processing in the living cell. For example, it has been determined that early-spliceosomal components are actively recruited to transcribing genes lacking intronic sequences [122]. These observations are in accordance with previous biochemical and functional studies that describe a stable interaction between the initiating RNAPII complexes and the U1 snRNP [123]. In fact, stepwise cotranscriptional recruitment of the spliceosome has been reported recently [124]. Importantly, global splicing inhibition did not prevent recruitment of spliceosomal components to the active transcription site, further supporting that nuclear organization and coordination of pre-mRNA metabolism are significantly determined by transcription.

Finally, innovative microscopy and spectroscopy tools, coupled with powerful statistical analysis and modeling, have led to the first studies in estimating the dynamics of transcription at the single-molecule resolution [121, 125]. These novel approaches will allow us to gain further insights regarding single-cell behavior, and the aspects of noise, robustness, and cell-to-cell variability in pre-mRNA formation and processing, which may be important to globally understand the regulation of gene expression.

## 5. Concluding Remarks

The quantitative study of the spatial and dynamic aspects of transcription and pre-mRNA processing is revealing itself as an essential complement to well-established, classical biochemistry-based approaches to fully understand how the regulation of gene expression is exerted in the cell. As stated in this review, many recent insights that help to explain long-standing questions regarding mRNA biogenesis could not have been achieved otherwise. However, these studies also give rise to important new questions. What is the functional relevance of spatial organization and regulation of dynamics in the different stages of mRNA biogenesis for the cell in a given context? Can we obtain a genome-wide picture of these parameters for all protein-coding genes in a systematic manner? How is cell-to-cell variability regulated within a cellular population to be advantageous for the cellular population as a whole? Is the dynamic regulation of the spatial distribution of the involved factors an essential component for the fine tuning of functional coupling of transcription elongation and pre-mRNA processing? It is expected that there will be a remarkable increase in the usage and optimization of these approaches, combined with more conventional biochemical and functional approaches, in the study of all aspects of mRNA formation and function.

## Acknowledgments

This work was supported by the Spanish Ministry of Science and Innovation Grant no. (BFU2008-01599), by Fundación para la Investigación y Prevención del SIDA en España Grant no. (36768), and by Junta de Andalucía (Proyecto de Excelencia 2009/CVI-4626). Support from the Fondo Europeo de Desarrollo Regional (FEDER) is also acknowledged.

## References

- [1] M. Carmo-Fonseca and C. Carvalho, "Nuclear organization and splicing control," *Advances in Experimental Medicine and Biology*, vol. 623, pp. 1–13, 2007.
- [2] T. Cremer and C. Cremer, "Chromosome territories, nuclear architecture and gene regulation in mammalian cells," *Nature Reviews Genetics*, vol. 2, no. 4, pp. 292–301, 2001.
- [3] L. Parada and T. Misteli, "Chromosome positioning in the interphase nucleus," *Trends in Cell Biology*, vol. 12, no. 9, pp. 425–432, 2002.
- [4] R. Zhao, M. S. Bodnar, and D. L. Spector, "Nuclear neighborhoods and gene expression," *Current Opinion in Genetics and Development*, vol. 19, no. 2, pp. 172–179, 2009.
- [5] M. R. Branco and A. Pombo, "Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations," *Plos Biology*, vol. 4, no. 5, article e138, 2006.
- [6] G. Cavalli, "Chromosome kissing," *Current Opinion in Genetics and Development*, vol. 17, no. 5, pp. 443–450, 2007.
- [7] J. Q. Ling and A. R. Hoffman, "Epigenetics of long-range chromatin interactions," *Pediatric Research*, vol. 61, no. 5, pp. 11R–16R, 2007.
- [8] M. Carmo-Fonseca, "How genes find their way inside the cell nucleus," *Journal of Cell Biology*, vol. 179, no. 6, pp. 1093–1094, 2007.
- [9] M. Dundr and T. Misteli, "Functional architecture in the cell nucleus," *Biochemical Journal*, vol. 356, no. 2, pp. 297–310, 2001.
- [10] T. Misteli, "Concepts in nuclear architecture," *BioEssays*, vol. 27, no. 5, pp. 477–487, 2005.
- [11] E. Soutoglou and T. Misteli, "Mobility and immobility of chromatin in transcription and genome stability," *Current Opinion in Genetics and Development*, vol. 17, no. 5, pp. 435–442, 2007.
- [12] K. Ahmed, R. Li, and D. P. Bazett-Jones, "Electron spectroscopic imaging of the nuclear landscape," *Methods in Molecular Biology*, vol. 464, pp. 415–423, 2009.
- [13] D. P. Bazett-Jones, R. Li, E. Fussner, R. Nisman, and H. Dehghani, "Elucidating chromatin and nuclear domain architecture with electron spectroscopic imaging," *Chromosome Research*, vol. 16, no. 3, pp. 397–412, 2008.
- [14] H. Dehghani, G. Dellaire, and D. P. Bazett-Jones, "Organization of chromatin in the interphase mammalian cell," *Micron*, vol. 36, no. 2, pp. 95–108, 2005.
- [15] G. Dellaire, R. Nisman, and D. P. Bazett-Jones, "Correlative light and electron spectroscopic imaging of chromatin in situ," *Methods in Enzymology*, vol. 375, pp. 456–478, 2004.
- [16] L. de Jong, M. A. Grande, K. A. Mattern, W. Schul, and R. van Driel, "Nuclear domains involved in RNA synthesis, RNA processing, and replication," *Critical Reviews in Eukaryotic Gene Expression*, vol. 6, no. 2-3, pp. 215–246, 1996.
- [17] A. I. Lamond and W. C. Earnshaw, "Structure and function in the nucleus," *Science*, vol. 280, pp. 547–553, 1998.
- [18] Y. S. Mao, B. Zhang, and D. L. Spector, "Biogenesis and function of nuclear bodies," *Trends in Genetics*, vol. 27, no. 8, pp. 295–306, 2011.
- [19] D. A. Jackson, A. B. Hassan, R. J. Errington, and P. R. Cook, "Visualization of focal sites of transcription within human nuclei," *The Embo Journal*, vol. 12, no. 3, pp. 1059–1065, 1993.
- [20] D. G. Wansink, W. Schul, I. van der Kraan, B. van Steensel, R. van Driel, and L. de Jong, "Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus," *Journal of Cell Biology*, vol. 122, no. 2, pp. 283–293, 1993.
- [21] F. J. Iborra, A. Pombo, D. A. Jackson, and P. R. Cook, "Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei," *Journal of Cell Science*, vol. 109, part 6, pp. 1427–1436, 1996.
- [22] F. J. Iborra, A. Pombo, J. McManus, D. A. Jackson, and P. R. Cook, "The topology of transcription by immobilized polymerases," *Experimental Cell Research*, vol. 229, no. 2, pp. 167–173, 1996.
- [23] D. A. Jackson, F. J. Iborra, E. M. Manders, and P. R. Cook, "Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei," *Molecular Biology of the Cell*, vol. 9, no. 9, pp. 1523–1536, 1998.
- [24] M. J. Mortillaro, B. J. Blencowe, X. Wei et al., "A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 16, pp. 8253–8257, 1996.
- [25] X. Wei, S. Somanathan, J. Samarabandu, and R. Berezney, "Three-dimensional visualization of transcription sites and their association with splicing factor-rich nuclear speckles," *Journal of Cell Biology*, vol. 146, no. 3, pp. 543–558, 1999.
- [26] C. Zeng, E. Kim, S. L. Warren, and S. M. Berget, "Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity," *The Embo Journal*, vol. 16, no. 6, pp. 1401–1412, 1997.
- [27] C. Lux, H. Albiez, R. D. Chapman et al., "Transition from initiation to promoter proximal pausing requires the CTD of RNA polymerase II," *Nucleic Acids Research*, vol. 33, no. 16, pp. 5139–5144, 2005.
- [28] P. R. Cook, "A model for all genomes: the role of transcription factories," *Journal of Molecular Biology*, vol. 395, no. 1, pp. 1–10, 2010.
- [29] P. Fraser, "Transcriptional control thrown for a loop," *Current Opinion in Genetics and Development*, vol. 16, no. 5, pp. 490–495, 2006.
- [30] P. Fraser and W. Bickmore, "Nuclear organization of the genome and the potential for gene regulation," *Nature*, vol. 447, no. 7143, pp. 413–417, 2007.
- [31] B. van Steensel and J. Dekker, "Genomics tools for unraveling chromosome architecture," *Nature Biotechnology*, vol. 28, no. 10, pp. 1089–1095, 2009.
- [32] J. M. Brown, J. Leach, J. E. Reittie et al., "Coregulated human globin genes are frequently in spatial proximity when active," *Journal of Cell Biology*, vol. 172, no. 2, pp. 177–187, 2006.
- [33] C. S. Osborne, L. Chakalova, K. E. Brown et al., "Active genes dynamically colocalize to shared sites of ongoing transcription," *Nature Genetics*, vol. 36, no. 10, pp. 1065–1071, 2004.

- [34] Q. Hu, Y. S. Kwon, E. Nunez et al., "Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 49, pp. 19199–19204, 2008.
- [35] A. Papantonis and P. R. Cook, "Fixing the model for transcription: the DNA moves, not the polymerase," *Transcription*, vol. 2, no. 1, pp. 41–44, 2011.
- [36] M. Xu and P. R. Cook, "Similar active genes cluster in specialized transcription factories," *Journal of Cell Biology*, vol. 181, no. 4, pp. 615–623, 2008.
- [37] I. Montanuy, R. Torremocha, C. Hernandez-Munain, and C. Sune, "Promoter influences transcription elongation: TATA-box element mediates the assembly of processive transcription complexes responsive to cyclin-dependent kinase 9," *Journal of Biological Chemistry*, vol. 283, no. 12, pp. 7368–7378, 2008.
- [38] H. Takahashi, T. J. Parmely, S. Sato et al., "Human mediator subunit MED26 functions as a docking site for transcription elongation factors," *Cell*, vol. 146, no. 1, pp. 92–104, 2011.
- [39] J. A. Mitchell and P. Fraser, "Transcription factories are nuclear subcompartments that remain in the absence of transcription," *Genes and Development*, vol. 22, no. 1, pp. 20–25, 2008.
- [40] J. Li and D. S. Gilmour, "Promoter proximal pausing and the control of gene expression," *Current Opinion in Genetics & Development*, vol. 21, pp. 231–235, 2011.
- [41] C. H. Eskiw, A. Rapp, D. R. Carter, and P. R. Cook, "RNA polymerase II activity is located on the surface of protein-rich transcription factories," *Journal of Cell Science*, vol. 121, no. 12, pp. 1999–2007, 2008.
- [42] D. R. Carter, C. Eskiw, and P. R. Cook, "Transcription factories," *Biochemical Society Transactions*, vol. 36, no. 4, pp. 585–589, 2008.
- [43] A. C. Goldstrohm, A. L. Greenleaf, and M. A. Garcia-Blanco, "Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing," *Gene*, vol. 277, no. 1-2, pp. 31–47, 2001.
- [44] E. Kim, L. Du, D. B. Bregman, and S. L. Warren, "Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA," *Journal of Cell Biology*, vol. 136, no. 1, pp. 19–28, 1997.
- [45] S. Pandit, D. Wang, and X. D. Fu, "Functional integration of transcriptional and RNA processing machineries," *Current Opinion in Cell Biology*, vol. 20, no. 3, pp. 260–265, 2008.
- [46] A. Pombo and P. R. Cook, "The localization of sites containing nascent RNA and splicing factors," *Experimental Cell Research*, vol. 229, no. 2, pp. 201–203, 1996.
- [47] P. R. Cook, "Nongenic transcription, gene regulation and action at a distance," *Journal of Cell Science*, vol. 116, no. 22, pp. 4483–4491, 2003.
- [48] T. Maniatis and R. Reed, "An extensive network of coupling among gene expression machines," *Nature*, vol. 416, pp. 499–506, 2002.
- [49] G. Orphanides and D. Reinberg, "A unified theory of gene expression," *Cell*, vol. 108, no. 4, pp. 439–451, 2002.
- [50] K. M. Neugebauer, "On the importance of being co-transcriptional," *Journal of Cell Science*, vol. 115, no. 20, pp. 3865–3871, 2002.
- [51] P. Bjork and L. Wieslander, "Nucleocytoplasmic mRNP export is an integral part of mRNP biogenesis," *Chromosoma*, vol. 120, no. 1, pp. 23–38, 2011.
- [52] R. F. Luco, M. Allo, I. E. Schor, A. R. Kornblihtt, and T. Misteli, "Epigenetics in alternative pre-mRNA splicing," *Cell*, vol. 144, no. 1, pp. 16–26, 2011.
- [53] P. Maiuri, A. Knezevich, E. Bertrand, and A. Marcello, "Real-time imaging of the HIV-1 transcription cycle in single living cells," *Methods*, vol. 53, pp. 62–67, 2010.
- [54] M. J. Moore and N. J. Proudfoot, "Pre-mRNA processing reaches back to transcription and ahead to translation," *Cell*, vol. 136, no. 4, pp. 688–700, 2009.
- [55] M. J. Munoz, M. de la Mata, and A. R. Kornblihtt, "The carboxy terminal domain of RNA polymerase II and alternative splicing," *Trends in Biochemical Sciences*, vol. 35, no. 9, pp. 497–504, 2010.
- [56] T. Trcek and R. H. Singer, "The cytoplasmic fate of an mRNP is determined cotranscriptionally: exception or rule?" *Genes and Development*, vol. 24, no. 17, pp. 1827–1831, 2010.
- [57] M. Barboric, T. Lenasi, H. Chen, E. B. Johansen, S. Guo, and B. M. Peterlin, "7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 19, pp. 7798–7803, 2009.
- [58] M. J. Munoz, M. S. P. Santangelo, M. P. Paronetto et al., "DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation," *Cell*, vol. 137, no. 4, pp. 708–720, 2009.
- [59] S. E. Sanchez, E. Petrillo, E. J. Beckwith et al., "A methyl transferase links the circadian clock to the regulation of alternative splicing," *Nature*, vol. 468, no. 7320, pp. 112–116, 2010.
- [60] I. E. Schor, N. Rascovan, F. Pelisch, M. Alio, and A. R. Kornblihtt, "Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 11, pp. 4325–4330, 2009.
- [61] A. I. Lamond and D. L. Spector, "Nuclear speckles: a model for nuclear organelles," *Nature Reviews Molecular Cell Biology*, vol. 4, no. 8, pp. 605–612, 2003.
- [62] D. L. Spector and A. I. Lamond, "Nuclear speckles," *Cold Spring Harbor Perspectives in Biology*, vol. 3, no. 2, 2011.
- [63] M. R. Lerner, J. A. Boyle, J. A. Hardin, and J. A. Steitz, "Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus," *Science*, vol. 211, pp. 400–402, 1981.
- [64] D. L. Spector, W. H. Schrier, and H. Busch, "Immunoelectron microscopic localization of SnRNPs," *Biology of the Cell*, vol. 49, no. 1, pp. 1–10, 1983.
- [65] K. C. Carter, D. Bowman, W. Carrington et al., "A three-dimensional view of precursor messenger RNA metabolism within the mammalian nucleus," *Science*, vol. 259, pp. 1330–1335, 1993.
- [66] K. C. Carter, K. L. Taneja, and J. B. Lawrence, "Discrete nuclear domains of poly(A) RNA and their relationship to the functional organization of the nucleus," *Journal of Cell Biology*, vol. 115, no. 5, pp. 1191–1202, 1991.
- [67] P. J. Mintz, S. D. Patterson, A. F. Neuwald, C. S. Spahr, and D. L. Spector, "Purification and biochemical characterization of interchromatin granule clusters," *The Embo Journal*, vol. 18, no. 15, pp. 4308–4320, 1999.
- [68] N. Saitoh, C. S. Spahr, S. D. Patterson, P. Bubulya, A. F. Neuwald, and D. L. Spector, "Proteomic analysis of interchromatin granule clusters," *Molecular Biology of the Cell*, vol. 15, no. 8, pp. 3876–3890, 2004.

- [69] S. Huang and D. L. Spector, "Dynamic organization of pre-mRNA splicing factors," *Journal of Cellular Biochemistry*, vol. 62, no. 2, pp. 191–197, 1996.
- [70] S. Huang and D. L. Spector, "Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription," *Journal of Cell Biology*, vol. 133, no. 4, pp. 719–732, 1996.
- [71] D. Cmarko, P. J. Verschure, T. E. Martin et al., "Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection," *Molecular Biology of the Cell*, vol. 10, no. 1, pp. 211–223, 1999.
- [72] S. Fakan, "Perichromatin fibrils are in situ forms of nascent transcripts," *Trends in Cell Biology*, vol. 4, no. 3, pp. 86–90, 1994.
- [73] T. Misteli, J. F. Caceres, and D. L. Spector, "The dynamics of a pre-mRNA splicing factor in living cells," *Nature*, vol. 387, pp. 523–527, 1997.
- [74] Y. Xing, C. V. Johnson, P. T. Moen Jr., J. A. McNeil, and J. B. Lawrence, "Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains," *Journal of Cell Biology*, vol. 131, no. 6, pp. 1635–1647, 1995.
- [75] J. D. Ellis, D. Lleres, M. Denegri, A. I. Lamond, and J. F. Caceres, "Spatial mapping of splicing factor complexes involved in exon and intron definition," *Journal of Cell Biology*, vol. 181, no. 6, pp. 921–934, 2008.
- [76] S. A. Mabon and T. Misteli, "Differential recruitment of pre-mRNA splicing factors to alternatively spliced transcripts in vivo," *Plos Biology*, vol. 3, no. 11, article e374, 1901 pages, 2005.
- [77] A. Pombo, J. Ferreira, E. Bridge, and M. Carmo-Fonseca, "Adenovirus replication and transcription sites are spatially separated in the nucleus of infected cells," *The Embo Journal*, vol. 13, no. 21, pp. 5075–5085, 1994.
- [78] T. Misteli and D. L. Spector, "RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo," *Molecular Cell*, vol. 3, no. 6, pp. 697–705, 1999.
- [79] X. Y. Zhong, J. H. Ding, J. A. Adams, G. Ghosh, and X. D. Fu, "Regulation of SR protein phosphorylation and alternative splicing by modulating kinetic interactions of SRPK1 with molecular chaperones," *Genes and Development*, vol. 23, no. 4, pp. 482–495, 2009.
- [80] V. Tripathi, J. D. Ellis, Z. Shen et al., "The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation," *Molecular Cell*, vol. 39, no. 6, pp. 925–938, 2010.
- [81] K. V. Prasanth, M. Camiolo, G. Chan et al., "Nuclear organization and dynamics of 7SK RNA in regulating gene expression," *Molecular Biology of the Cell*, vol. 21, no. 23, pp. 4184–4196, 2010.
- [82] K. J. Moon, K. Mochizuki, M. Zhou, H. S. Jeong, J. N. Brady, and K. Ozato, "The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription," *Molecular Cell*, vol. 19, no. 4, pp. 523–534, 2005.
- [83] S. Fakan and R. van Driel, "The perichromatin region: a functional compartment in the nucleus that determines large-scale chromatin folding," *Seminars in Cell and Developmental Biology*, vol. 18, no. 5, pp. 676–681, 2007.
- [84] J. Niedojadlo, C. Perret-Vivancos, K. H. Kalland et al., "Transcribed DNA is preferentially located in the perichromatin region of mammalian cell nuclei," *Experimental Cell Research*, vol. 317, pp. 433–444, 2010.
- [85] S. Q. Xie, S. Martin, P. V. Guillot, D. L. Bentley, and A. Pombo, "Splicing speckles are not reservoirs of RNA polymerase II, but contain an inactive form, phosphorylated on serine2 residues of the C-terminal domain," *Molecular Biology of the Cell*, vol. 17, no. 4, pp. 1723–1733, 2006.
- [86] P. T. Moen Jr., C. V. Johnson, M. Byron et al., "Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis," *Molecular Biology of the Cell*, vol. 15, no. 1, pp. 197–206, 2004.
- [87] J. M. Brown, J. Green, R. P. des Neves et al., "Association between active genes occurs at nuclear speckles and is modulated by chromatin environment," *Journal of Cell Biology*, vol. 182, no. 6, pp. 1083–1097, 2008.
- [88] L. S. Shopland, C. V. Johnson, M. Byron, J. McNeil, and J. B. Lawrence, "Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods," *Journal of Cell Biology*, vol. 162, no. 6, pp. 981–990, 2003.
- [89] Y. Hu, I. Kireev, M. Plutz, N. Ashourian, and A. S. Belmont, "Large-scale chromatin structure of inducible genes: transcription on a condensed, linear template," *Journal of Cell Biology*, vol. 185, no. 1, pp. 87–100, 2009.
- [90] L. L. Hall, K. P. Smith, M. Byron, and J. B. Lawrence, "Molecular anatomy of a speckle," *Anatomical Record A Discoveries in Molecular, Cellular, and Evolutionary Biology*, vol. 288, no. 7, pp. 664–675, 2006.
- [91] C. Johnson, D. Primorac, M. McKinstry, J. McNeil, D. Rowe, and J. B. Lawrence, "Tracking COL1A1 RNA in osteogenesis imperfecta: splice-defective transcripts initiate transport from the gene but are retained within the SC35 domain," *Journal of Cell Biology*, vol. 150, no. 3, pp. 417–431, 2000.
- [92] L. S. Shopland, C. V. Johnson, and J. B. Lawrence, "Evidence that all SC-35 domains contain mRNAs and that transcripts can be structurally constrained within these domains," *Journal of Structural Biology*, vol. 140, no. 1–3, pp. 131–139, 2002.
- [93] J. B. Lawrence and C. M. Clemson, "Gene associations: true romance or chance meeting in a nuclear neighborhood?" *Journal of Cell Biology*, vol. 182, no. 6, pp. 1035–1038, 2008.
- [94] L. Chang, W. J. Godinez, I. H. Kim et al., "PNAS plus: herpesviral replication compartments move and coalesce at nuclear speckles to enhance export of viral late mRNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. E136–E144, 2011.
- [95] U. Schmidt, K. B. IM, C. Benzing et al., "Assembly and mobility of exon-exon junction complexes in living cells," *The Royal Naval Association*, vol. 15, no. 5, pp. 862–876, 2009.
- [96] U. Schmidt, K. Richter, A. B. Berger, and P. Lichter, "In vivo BiFC analysis of Y14 and NXF1 mRNA export complexes: preferential localization within and around SC35 domains," *Journal of Cell Biology*, vol. 172, no. 3, pp. 373–381, 2006.
- [97] G. Donnert, J. Keller, R. Medda et al., "Macromolecular-scale resolution in biological fluorescence microscopy," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 31, pp. 11440–11445, 2006.
- [98] P. J. Mintz and D. L. Spector, "Compartmentalization of RNA processing factors within nuclear speckles," *Journal of Structural Biology*, vol. 129, no. 2–3, pp. 241–251, 2000.
- [99] K. E. Handwerger, J. A. Cordero, and J. G. Gall, "Cajal bodies, nucleoli, and speckles in the Xenopus oocyte nucleus have a low-density, sponge-like structure," *Molecular Biology of the Cell*, vol. 16, no. 1, pp. 202–211, 2005.
- [100] H. Cheng, K. Dufu, C. S. Lee, J. L. Hsu, A. Dias, and R. Reed, "Human mRNA export machinery recruited to the 5' end of mRNA," *Cell*, vol. 127, no. 7, pp. 1389–1400, 2006.

- [101] A. P. Dias, K. Dufu, H. Lei, and R. Reed, "A role for TREX components in the release of spliced mRNA from nuclear speckle domains," *Nature Communications*, vol. 1, no. 7, 2010.
- [102] R. Luna, H. Gaillard, C. Gonzalez-Aguilera, and A. Aguilera, "Biogenesis of mRNPs: integrating different processes in the eukaryotic nucleus," *Chromosoma*, vol. 117, no. 4, pp. 319–331, 2008.
- [103] A. G. Rondon, S. Jimeno, M. Garcia-Rubio, and A. Aguilera, "Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation," *Journal of Biological Chemistry*, vol. 278, no. 40, pp. 39037–39043, 2003.
- [104] P. Valencia, A. P. Dias, and R. Reed, "Splicing promotes rapid and efficient mRNA export in mammalian cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 9, pp. 3386–3391, 2008.
- [105] D. B. Bregman, L. Du, S. Van der Zee, and S. L. Warren, "Transcription-dependent redistribution of the large subunit of RNA polymerase II to discrete nuclear domains," *Journal of Cell Biology*, vol. 129, no. 2, pp. 287–298, 1995.
- [106] A. B. Houtsmuller and W. Vermeulen, "Macromolecular dynamics in living cell nuclei revealed by fluorescence redistribution after photobleaching," *Histochemistry and Cell Biology*, vol. 115, no. 1, pp. 13–21, 2001.
- [107] T. K. Kerppola, "Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells," *Annual Review of Biophysics*, vol. 37, pp. 465–487, 2008.
- [108] M. Koster, T. Frahm, and H. Hauser, "Nucleocytoplasmic shuttling revealed by FRAP and FLIP technologies," *Current Opinion in Biotechnology*, vol. 16, no. 1, pp. 28–34, 2005.
- [109] F. Mueller, D. Mazza, T. J. Stasevich, and J. G. McNally, "FRAP and kinetic modeling in the analysis of nuclear protein dynamics: what do we really know?" *Current Opinion in Cell Biology*, vol. 22, no. 3, pp. 403–411, 2010.
- [110] J. Rino, J. Braga, R. Henriques, and M. Carmo-Fonseca, "Frontiers in fluorescence microscopy," *International Journal of Developmental Biology*, vol. 53, no. 8-10, pp. 1569–1579, 2009.
- [111] T. C. Voss, I. A. Demarco, and R. N. Day, "Quantitative imaging of protein interactions in the cell nucleus," *BioTechniques*, vol. 38, no. 3, pp. 413–424, 2005.
- [112] D. B. Zorov, E. Kobrin, M. Juhaszova, and S. J. Sollott, "Examining intracellular organelle function using fluorescent probes: from animalcules to quantum dots," *Circulation Research*, vol. 95, no. 3, pp. 239–252, 2004.
- [113] R. D. Phair and T. Misteli, "High mobility of proteins in the mammalian cell nucleus," *Nature*, vol. 404, no. 6778, pp. 604–609, 2000.
- [114] J. Rino and M. Carmo-Fonseca, "The spliceosome: a self-organized macromolecular machine in the nucleus?" *Trends in Cell Biology*, vol. 19, no. 8, pp. 375–384, 2009.
- [115] J. Rino, J. M. Desterro, T. R. Pacheco, T. W. Gadella Jr., and M. Carmo-Fonseca, "Splicing factors SF1 and U2AF associate in extraspliceosomal complexes," *Molecular and Cellular Biology*, vol. 28, no. 9, pp. 3045–3057, 2008.
- [116] D. Molle, P. Maiuri, S. Boireau et al., "A real-time view of the TAR:Tat:P-TEFb complex at HIV-1 transcription sites," *Retrovirology*, vol. 4, article 36, 2007.
- [117] J. Chusainow, P. M. Ajuh, L. Trinkle-Mulcahy, J. E. Sleeman, J. Ellenberg, and A. I. Lamond, "FRET analyses of the U2AF complex localize the U2AF35/U2AF65 interaction in vivo and reveal a novel self-interaction of U2AF35," *The Royal Naval Association*, vol. 11, no. 8, pp. 1201–1214, 2005.
- [118] X. Darzacq, Y. Shav-Tal, V. de Turris et al., "In vivo dynamics of RNA polymerase II transcription," *Nature Structural and Molecular Biology*, vol. 14, no. 9, pp. 796–806, 2007.
- [119] S. Boireau, P. Maiuri, E. Basyuk et al., "The transcriptional cycle of HIV-1 in real-time and live cells," *Journal of Cell Biology*, vol. 179, no. 2, pp. 291–304, 2007.
- [120] J. Yao, M. B. Ardehali, C. J. Fecko, W. W. Webb, and J. T. Lis, "Intranuclear distribution and local dynamics of RNA polymerase II during transcription activation," *Molecular Cell*, vol. 28, no. 6, pp. 978–990, 2007.
- [121] T. Lionnet, B. Wu, D. Grunwald, R. H. Singer, and D. R. Larson, "Nuclear physics: quantitative single-cell approaches to nuclear organization and gene expression," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 75, pp. 113–126, 2010.
- [122] B. Spiluttini, B. Gu, P. Belagal et al., "Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells," *Journal of Cell Science*, vol. 123, no. 12, pp. 2085–2093, 2010.
- [123] R. Das, J. Yu, Z. Zhang et al., "SR proteins function in coupling RNAP II transcription to pre-mRNA splicing," *Molecular Cell*, vol. 26, no. 6, pp. 867–881, 2007.
- [124] U. Schmidt, E. Basyuk, M. C. Robert et al., "Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation," *The Journal of Cell Biology*, vol. 193, pp. 819–829, 2011.
- [125] D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer, "Real-time observation of transcription initiation and elongation on an endogenous yeast gene," *Science*, vol. 332, pp. 475–478, 2011.